

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

НАВЧАЛЬНО-МЕТОДИЧНИЙ КОМПЛЕКС ДИСЦИПЛІНИ
Теорія ймовірностей, ймовірності процеси та математична статистика

Галузь знань: **12 «Інформаційні технології»**
(шифр і назва галузі)

Спеціальність: **122 «Комп'ютерні науки»**
(код та найменування спеціальності)

Освітня програма: **«Інформаційні технології проектування»**
(найменування освітньої програми)

Рівень вищої освіти: **перший (бакалаврський)**

Харків 2020 рік

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Керівник проектної групи/
Голова НМК

(підпис)

(ініціали та прізвище)

«_____» _____ 2019 р.

**РОБОЧА ПРОГРАМА ОBOB'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Теорія ймовірностей, ймовірнісні процеси
та математична статистика**

Галузь знань:

12 «Інформаційні технології»
(шифр і назва галузі)

Спеціальність:

122 «Комп'ютерні науки»
(код та найменування спеціальності)

Освітня програма:

«Інформаційні технології проектування»
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2019 рік

Робоча програма навчальної дисципліни «Теорія ймовірностей, ймовірнісні процеси та математична статистика»

(назва дисципліни)

для студентів за спеціальністю 122 «Комп'ютерні науки»
освітньою програмою «Інформаційні технології проектування»

«7» червня 2019 р., – 10 с.

Розробник: ст. викладач кафедри вищої математики та системного аналізу

О.А. Мураховська

(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій проектування (№ 105)

Протокол № _____ від « _____ » _____ 2019р.

Завідувач кафедри інформаційних технологій проектування

д.т.н., професор _____
(підпис)

(Є. А. Дружинін)
(прізвище та ініціали)

"

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів - 5	<u>Галузь знань:</u> 12 «Інформаційні технології» <u>Спеціальність:</u> 122 «Комп'ютерні науки» <u>Освітня програма:</u> «Інформаційні технології проектування» <u>Рівень вищої освіти: перший (бакалаврський)</u>	Навчальний рік
Модулів - 2		2019/2020
Змістових модулів - 4		Семестр
		2-й
		Лекції¹⁾
Розрахункова робота «Випадкові величини та функції випадкових величин»		32 год.
Загальна кількість годин – 64/150		Практичні — год.
Тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 5,4		Лабораторні роботи¹⁾ 32 год.
	Самостійна робота	
	86 год.	
	Вид контролю: іспит	

Примітка

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: – 64/150.

¹⁾ Аудиторне навантаження може бути збільшене або зменшене на одну годину в залежності від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета дисципліни – розвинення ймовірнісного мислення, засвоєння термінології та понять теорії випадкових процесів, статистичного опису процесів і систем, лінійних перетворень випадкових функцій, їх канонічних представлень, що широко використовуються на практиці моделей випадкових явищ.

Завдання дисциплін – формування теоретичних знань та практичних навичок використання теорії ймовірностей, математичної статистики та випадкових процесів для формалізації та математичного моделювання прикладних завдань для їх ефективної програмно-апаратної реалізації.

Міждисциплінарні зв'язки: вища математика.

Як **результат вивчення** дисципліни студент повинен **знати**:

- поняття випадкової події та випадкової величини, їх основні види;
- означення ймовірності та її властивості;
- основні формули комбінаторики; елементи алгебри подій;
- теореми додавання і множення ймовірностей;
- форми представлення випадкових величин та їх основні характеристики;
- найбільш вживані закони розподілу випадкових величин;
- поняття вибірки та методи її формування;
- методи отримання точкових та інтервальних оцінок числових характеристик сукупностей;
- правила побудови критеріїв для перевірки статистичних гіпотез;
- методи кореляційного та регресійного аналізу;
- основні поняття теорії випадкових процесів: визначення і класифікація випадкових процесів;
- закони розподілу та основні характеристики випадкових процесів;
- перетворення випадкових процесів та операції над ними;
- стаціонарні та ергодичні випадкові процеси;
- спектральний опис випадкових процесів;
- потоки подій, їх властивості та класифікація;
- ланцюги Маркова та марковські процеси з дискретним і неперервним часом.

вміти:

- обчислювати ймовірність випадкових подій з використанням основних означень та теорем;
- будувати закони розподілу та обчислювати характеристики випадкових величин;
- проводити первинну обробку статистичних даних;
- визначати точкові та інтервальні оцінки числових характеристик;
- вибирати слушні критерії для перевірки статистичних гіпотез та користуватися таблицями відповідних розподілів;
- обґрунтовано вибирати класи регресійних моделей при виявленні залежностей між величинами; обчислювати коефіцієнти рівнянь лінійних та нелінійних залежностей;
- розраховувати основні характеристики випадкових процесів за заданими законами розподілу ймовірності;
- аналізувати процеси на стаціонарність та ергодичність;
- описувати процеси в спектральній області;
- застосовувати ланцюги Маркова та марковські процеси з дискретним і неперервним часом для розв'язання практичних задач;
- застосовувати поняття та співвідношення теорії масового обслуговування для аналізу задач.

мати навички:

- використання стандартного програмного забезпечення для аналізу ймовірнісних процесів та статистичної обробки даних;
- побудови математичних моделей реальних випадкових процесів;
- самостійного розширення своїх математичних знань за допомогою літера-тури або навчаючих програм.

3. Програма навчальної дисципліни

Модуль 1

Змістовий модуль 1. Випадкові події та величини

Тема 1. Випадкові події

Предмет теорії ймовірностей. Події та їх класифікація на достовірні, неможливі та випадкові. Сумісні та несумісні події. Протилежні події. Повна група подій. Означення ймовірності (класичне, геометричне, статистичне) та її властивості. Формули комбінаторики, їх застосування при обчисленні ймовірностей. Алгебра подій. Поняття сумісних і несумісних випадкових подій. Формули додавання несумісних випадкових подій. Поняття залежності і незалежності випадкових подій. Умовна ймовірність та її властивості. Формули множення ймовірностей для залежних та незалежних випадкових подій. Формули додавання сумісних випадкових подій. Використання формул множення ймовірностей для оцінки надійності деяких систем. Формула повної ймовірності та формула Байеса.

Тема 2. Моделі повторних випробувань

Схема повторних незалежних випробувань. Формула Бернуллі для обчислення ймовірності та найімовірнішого числа подій. Асимптотичні форми формули Бернуллі (локальна та інтегральна теореми Муавра-Лапласа). Використання інтегральної теореми. Формула Пуассона для малої ймовірності випадкових подій.

Тема 3. Випадкові величини та їх числові характеристики

Визначення та види випадкових величин. Закон розподілу випадкових величин. Функція розподілу та її властивості. Властивості неперервної випадкової величини. Щільність розподілу та її властивості. Числові характеристики випадкових величин. Мода і медіана. Математичне сподівання та його властивості. Характеристики розсіяння випадкової величини. Моменти випадкових величин.

Змістовий модуль 2. Закони розподілу. Системи випадкових величин

Тема 4. Найбільш поширені закони розподілу випадкових величин

Найпростіші закони розподілу дискретних випадкових величин. Біномний розподіл. Розподіл Пуассона. Геометричний розподіл. Розподіл Паскаля. Гіпергеометричний розподіл. Показниковий розподіл. Рівномірний розподіл неперервної величини. Нормальний закон розподілу. Нормальна крива: вплив параметрів розподілу на її форму. Ймовірність попадання випадкової величини з нормальним законом розподілу у заданий інтервал. Ймовірність заданого відхилення. Правило трьох сигм.

Тема 5. Системи випадкових величин

Поняття про багатовимірні випадкові величини і системи випадкових величин. Функція розподілу ймовірностей системи та її властивості. Функція щільності розподілу та її властивості. Числові характеристики системи випадкових величин. Умовні закони розподілу та їх характеристики. Регресія однієї випадкової величини на іншу. Властивості регресії. Коефіцієнт кореляції і його властивості. Приклади аналізу систем випадкових величин.

Тема 6. Граничні теореми теорії ймовірностей

Поняття про закон великих чисел. Нерівність Чебишева та її значення. теорема Чебишева і теорема Бернуллі. Центральна гранична теорема (теорема Ляпунова) та її практичне застосування.

Модульний контроль

Модуль 2

Змістовий модуль 3. Основи математичної статистики. Точкові та інтервальні оцінки.

Тема 7. Основні поняття математичної статистики

Завдання математичної статистики. Вибірковий метод і його основні поняття. Варіаційний ряд. Емпіричний закон розподілу, емпірична функція розподілу. Полігон та гістограма. Оцінки характеристик положення. Оцінки характеристик розсіяння. Оцінка коефіцієнта кореляції системи випадкових величин. Інтервальна оцінка математичного сподівання. Інтервальна оцінка середнього квадратичного відхилення. Інтервальна оцінка коефіцієнта кореляції.

Тема 8. Статистична перевірка статистичних гіпотез

Загальне поняття про перевірку гіпотез. Основна (нульова) та альтернативна гіпотези, помилки першого та другого роду. Статистичні критерії, розрахункові та критичні значення критеріїв, потужність критерію. Загальна схема перевірки статистичних гіпотез.

Порівняння математичних сподівань. Порівняння математичного сподівання і сталої. Порівняння двох дисперсій. Порівняння відносної частоти і ймовірності. Перевірка гіпотези про вид закону розподілу генеральної сукупності за допомогою критерію χ^2 .

Тема 9. Елементи кореляційного та регресійного аналізу

Функціональна, статистична та кореляційна залежність. Вибірковий коефіцієнт кореляції та його інтервальна оцінка. Перевірка гіпотези про значущість коефіцієнта кореляції. Поняття про регресію. Лінійна регресія. Прямі регресії та знаходження їх параметрів методом найменших квадратів. Найпростіші випадки нелінійної регресії. Поняття про множинну регресію.

Змістовий модуль 4. Випадкові процеси.

Тема 10. Випадкові процеси

Основні означення. Класифікація випадкових процесів. Елементарні випадкові функції. Закони розподілу випадкових процесів.

Тема 11. Характеристики випадкових процесів

Математичне сподівання випадкового процесу. Дисперсія випадкового процесу. Нормована кореляційна функція випадкового процесу. Взаємна кореляційна функція двох стахостичних процесів. Нормована взаємна кореляційна функція двох стахостичних процесів.

Тема 12. Однорідні ланцюги Маркова. Стаціонарні випадкові процеси

Означення ланцюга Маркова. Опис і зображення ланцюгів Маркова. Властивості ланцюгів Маркова. Матриця переходу. Класифікація станів. Поняття марковського процесу. Марковські випадкові процеси з дискретним часом. Потоки подій. Найпростіший потік подій. Марковські дискретні процеси, характерні неперервним часом. Марковські процеси типу «загибель-розмноження». Рівняння Колмогорова. Граничні ймовірності станів. Поняття про стаціонарні випадкові процеси. Спектральні зображення стаціонарних процесів. Ергодична властивість стаціонарних випадкових процесів.

Модульний контроль

4. Структура навчальної дисципліни

Назва змістових модулів і тем	Кількість годин (денна форма)					
	усьо- го	у тому числі				
		лк	пр	лаб	інд	с.р.
1	2	3	4	5	6	7
Модуль 1						
Змістовий модуль 1. Випадкові події та величини						
Тема 1. Випадкові події	10	2	—	2	—	6
Тема 2. Моделі повторних випробувань	12	2	—	2	—	8
Тема 3. Випадкові величини та їх числові характеристики	16	4	—	4	—	8
Разом за змістовим модулем 1	38	8	—	8	—	22
Змістовий модуль 2. Закони розподілу. Системи випадкових величин.						

Тема 4. Найбільш поширені закони розподілу випадкових величин	10	2	—	2	—	6
Тема 5. Системи випадкових величин	18	4	—	4	—	10
Тема 6. Граничні теореми теорії ймовірностей	10	2	—	2	—	6
Разом за змістовим модулем 2	38	8	—	8	—	22
Усього годин за модулем 1	76	16	—	16	—	44
Модуль 2						
Змістовий модуль 3. Основи математичної статистики. Точкові та інтервальні оцінки.						
Тема 7. Основні поняття математичної статистики	10	2	—	2	—	6
Тема 8. Статистична перевірка статистичних гіпотез	16	4	—	4	—	8
Тема 9. Елементи кореляційного та регресійного аналізу	16	4	—	4	—	8
Разом за змістовим модулем 3	42	10	—	10	—	22
Змістовий модуль 4. Випадкові процеси.						
Тема 10. Випадкові процеси	10	2	—	2	—	6
Тема 11. Характеристики випадкових процесів	12	2	—	2	—	8
Тема 12. Однорідні ланцюги Маркова. Стаціонарні випадкові процеси	10	2	—	2	—	6
Разом за змістовим модулем 4	32	6	—	6	—	20
Усього годин за модулем 2	74	16	—	16	—	42
Усього годин	150	32	—	32	—	86

5. Теми семінарських занять

Семінарські заняття навчальним планом не передбачені

6. Теми практичних занять

Практичні заняття навчальним планом не передбачені

7. Теми лабораторних занять

№ п/п	Назва теми	Кількість годин
1	Комбінаторика. Ймовірність. Теореми додавання і множення ймовірностей	2
2	Формула повної ймовірності та формула Байєса. Повторні незалежні випробування	2
3	Числові характеристики дискретних випадкових величин	2
4	Числові характеристики неперервних випадкових величин	2
5	Закони розподілу дискретних та неперервних випадкових величин	2
6	Аналіз систем випадкових величин	4
7	Граничні теореми теорії ймовірностей	2
8	Статистичний аналіз вибірових сукупностей: визначення емпіричного розподілу, полігона, гістограми, емпіричної функції розподілу та побудова її графіку.	2
9	Точкові оцінки параметрів генеральної сукупності. Інтервальні оцінки числових характеристик	2
10	Статистична перевірка статистичних гіпотез	2

11	Елементи кореляційного та регресійного аналізу	4
12	Випадкові процеси	2
13	Характеристики випадкових процесів	2
14	Марковські випадкові процеси з дискретним часом. Марковські випадкові процеси з неперервним часом	2
Разом		32

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1.	Випадкові події. Алгебра подій. Поняття сумісних і несумісних випадкових подій. Формули додавання несумісних випадкових подій. Поняття залежності і незалежності випадкових подій. Умовна ймовірність та її властивості. Формули множення ймовірностей для залежних та незалежних випадкових подій.	6
2.	Моделі повторних випробувань. Асимптотичні форми формули Бернуллі (локальна та інтегральна теорема Муавра-Лапласа).	8
3.	Випадкові величини та їх числові характеристики.	8
4.	Найбільш поширені закони розподілу випадкових величин	6
5.	Умовні закони розподілу та їх характеристики. Регресія однієї випадкової величини на іншу. Властивості регресії. Коефіцієнт кореляції і його властивості. Приклади аналізу систем випадкових величин.	10
6.	Поняття про закон великих чисел. Нерівність Чебишева та її значення. теорема Чебишева і теорема Бернуллі. Центральна гранична теорема (теорема Ляпунова) та її практичне застосування.	6
7.	Оцінка коефіцієнта кореляції системи випадкових величин. Інтервальна оцінка математичного сподівання. Інтервальна оцінка середнього квадратичного відхилення. Інтервальна оцінка коефіцієнта кореляції.	6
8.	Порівняння математичних сподівань. Порівняння математичного сподівання і сталої. Порівняння двох дисперсій. Порівняння відносної частоти і ймовірності. Перевірка гіпотези про вид закону розподілу генеральної сукупності за допомогою критерію χ^2 .	8
9	Поняття про регресію. Лінійна регресія. Прямі регресії та знаходження їх параметрів методом найменших квадратів. Найпростіші випадки нелінійної регресії. Поняття про множинну регресію.	8
10.	Випадкові процеси	6
11.	Характеристики випадкових процесів	8
12.	Поняття марковського процесу. Марковські випадкові процеси з дискретним часом. Потіки подій. Найпростіший потік подій. Марковські дискретні процеси, характерні неперервним часом. Марковські процеси типу «загибель-розмноження».	6
Разом		86

9. Індивідуальні завдання

Виконання розрахункової роботи «Випадкові величини та функції випадкових величин»

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, лабораторних робіт, індивідуальні

консультації (при необхідності), самостійна робота студентів за матеріалами, опублікованими кафедрою (методичні посібники).

11. Методи контролю

Проведення поточного контролю у вигляді тестів, усної здачі індивідуальних робіт, письмового модульного контролю, фінальний контроль у вигляді семестрового контролю: звлік (проводиться у разі відмови студента від балів поточного тестування та за наявності допуску).

12. Критерії оцінювання та розподіл балів, які отримують студенти

12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Модуль 1			
Виконання лабораторних робіт	0...3	7	0...21
Модульний контроль	0...23	1	0...23
Модуль 2			
Виконання лабораторних робіт	0...3	7	0...21
Модульний контроль	0...20	1	0...20
Виконання та захист розрахункової роботи	0...15	1	0...15
Всього за семестр			0...100

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування та за наявності допуску. При складанні іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з двох теоретичних та трьох практичних завдань. За кожне теоретичне питання та практичне завдання студент може отримати до 20 балів. Максимальна сума всіх балів – 100.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

знати основні поняття теорії ймовірностей та математичної статистики такі як стохастичний експеримент, випадкові події, означення ймовірності, дискретні та неперервні випадкові величин, закон розподілу та числові характеристики, моменти випадкової величини. Основні закони розподілу випадкових величин: біноміальний, Пуассона, геометричний, гіпергеометричний, рівномірний, показників, нормальний, t-розподіл Стьюдента. χ^2 -розподіл, F- розподіл Фішера-Снедекора. Закон великих чисел, граничні теореми та їх застосування. Генеральна та вибіркові сукупності, статистичні ряди розподілу вибірки, полігон та гістограмам частот, вибірковий метод. оцінки параметрів (точкові і інтервальні) оцінки параметрів вибірки, статистичні гіпотези, критерії узгодження Пірсона, Фішера, Стьюдента, Колмогорова, критерій χ^2 про вигляд розподілу.

Необхідний обсяг умінь для одержання позитивної оцінки:

уміти самостійно вивчати математичну літературу; розвинути логічне, абстрактне і алгоритмічне мислення; підбирати математичні методи для створення та аналізу математичної моделі чи задачі; строго викладати свої думки; оволодіти методами математичної статистики дослідження прикладних задач; застосовувати методи розв'язання ймовірносних задач; набути навичок самостійного пошуку математичних методів та підбору моделі в літературі з математичних дисциплін.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити всі лабораторні роботи. Застосовувати основні основні поняття теорії ймовірностей та математичної статистики для розв'язання задач. Відповіді студента розкривають суть питань без достатньої повноти і обґрунтування, або у

відповідях є неправильне тлумачення окремих понять та неточність у формулюванні відповідних термінів.

Добре (75-89). Твердо знати матеріал, захистити всі лабораторні роботи. Розв'язувати більш складні задачі. Вміти оцінювати область використання ймовірнісних методів. У відповідях студента можуть допускатися окремі помилки непринципового характеру, які не впливають на розкриття суті теоретичних питань. Завдання в цілому виконуються без помилок, але в обґрунтуванні розв'язання є певні недоліки.

Відмінно (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та уміти застосовувати їх.

Курсову роботу не передбачено навчальним планом.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

14. Рекомендована література

14.1. Базова

1. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. Пособие для студентов вузов. - 6-е изд., доп. – М.: Высшая школа, 2002. – 405 с.

2. Авраменко В. І., Карімов І. К. Теорія ймовірностей і математична статистика: Навчальний посібник. – Дніпродзержинськ: ДДТУ, 2009. – 254 с.

3. Волощенко А.Б. Теорія ймовірностей та математична статистика: навч.- метод. Посібник / А.Б. Волощенко, І.А. Джалладов. - К.: КНЕУ, 2005. – 256 с. – ISBN966–574–459–3.

4. Сеньо П.С. Випадкові процеси: підручник / П.С.Сеньо. Міністерство освіти і науки України, ЛНУ. – Львів: Компакт-ЛВ, 2006. – 288 с.– ISBN966–96414–7–0.

14.2. Допоміжна

5. Практикум з теорії ймовірностей та математичної статистики: Навч. посібник / Р.К. Чорней, О.Ю. Дюженкова, О.Б. Жильцов та ін.. За редакцією Р. К. Чорней. - К.: МАУП, 2003.- 328 с.

6. Грищенко В. О. Теорія ймовірностей і математична статистика. Практикум: Навч. посіб. – К.: Київ. нац. торг.-екон. ун-т. 2002. – 164 с.

7. Турчин В. М. Теорія ймовірностей і математична статистика: Основні поняття, приклади, задачі. — Дніпропетровськ: ДНУ, 2006. — 476 с.

8. Толбатов Ю. А. Математична статистика і задачі оптимізації в алгоритмах і програмах – К.: Вища школа, 1994. – 399 с.

9. Новицький І.В. Випадкові процеси. Навчальний посібник / І.В.Новицький, С.А. Ус. – Дніпропетровськ: Національний гірничий університет, 2011. – 125 с.

10. Вентцель Е.С. Теория случайных процессов и её инженерные приложения: учеб. пособие / Е.С. Вентцель, Л.А. Овчаров. – 4-е изд. – М. Высш. шк., 2007. – 479 с. – ISBN 978–5–06–005820–8.

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Керівник проектної групи/
Голова НМК

(підпис)

(ініціали та прізвище)

«_____» _____ 2019 р.

**РОБОЧА ПРОГРАМА ОBOB'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Теорія ймовірностей, ймовірнісні процеси
та математична статистика**

Галузь знань:

12 «Інформаційні технології»
(шифр і назва галузі)

Спеціальність:

122 «Комп'ютерні науки»
(код та найменування спеціальності)

Освітня програма:

«Інформаційні технології проектування»
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2019 рік

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Керівник проектної групи/
Голова НМК

«_____» _____ 2019 р.

**РОБОЧА ПРОГРАМА ОБОВ'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Теорія ймовірностей, ймовірнісні процеси
та математична статистика**

Галузь знань: 12 «Інформаційні технології»
(шифр і назва галузі)

Спеціальність: 122 «Комп'ютерні науки»
(код та найменування спеціальності)

Освітня програма: «Системне проектування»
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2019 рік

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Керівник проектної групи/
Голова НМК

(підпис)

(ініціали та прізвище)

« ____ » _____ 2019 р.

**РОБОЧА ПРОГРАМА ОБОВ'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Теорія ймовірностей, ймовірнісні процеси
та математична статистика**

Галузь знань: **12 «Інформаційні технології»**
(шифр і назва галузі)

Спеціальність: **126 «Інформаційні системи та технології»**
(код та найменування спеціальності)

Освітня програма: **«Інформаційні системи та технології
підтримки віртуальних середовищ»**
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2019 рік

Робоча програма навчальної дисципліни «Теорія ймовірностей, ймовірнісні процеси та математична статистика»

(назва дисципліни)

для студентів за спеціальністю 122 «Комп'ютерні науки»
освітньою програмою «Системне проектування»

«7» червня 2019 р., – 10 с.

Розробник: ст. викладач кафедри вищої математики та системного аналізу

О.А. Мураховська

(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій проектування (№ 105)

Протокол № _____ від « _____ » _____ 2019р.

Завідувач кафедри інформаційних технологій проектування

д.т.н., професор _____
(підпис)

(Є. А. Дружинін)
(прізвище та ініціали)

Робоча програма навчальної дисципліни «Теорія ймовірностей, ймовірнісні процеси та математична статистика»

(назва дисципліни)

для студентів за спеціальністю 122 «Комп'ютерні науки»
освітньою програмою «Інформаційні технології проектування»

«7» червня 2019 р., – 10 с.

Розробник: ст. викладач кафедри вищої математики та системного аналізу

О.А. Мураховська

(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій проектування (№ 105)

Протокол № _____ від « _____ » _____ 2019р.

Завідувач кафедри інформаційних технологій проектування

д.т.н., професор _____
(підпис)

(Є. А. Дружинін)
(прізвище та ініціали)

Робоча програма навчальної дисципліни «Теорія ймовірностей, ймовірнісні процеси та математична статистика»

(назва дисципліни)

для студентів за спеціальністю 126 «Інформаційні системи та технології»
освітньою програмою «Інформаційні системи та технології підтримки віртуальних середовищ»

«7» червня 2019 р., – 10 с.

Розробник: ст. викладач кафедри вищої математики та системного аналізу

О.А. Мураховська

(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій проектування (№ 105)

Протокол № _____ від « _____ » _____ 2019р.

Завідувач кафедри інформаційних технологій проектування

д.т.н., професор _____
(підпис)

(Є. А. Дружинін)
(прізвище та ініціали)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів - 3	<p style="text-align: center;"><u>Галузь знань:</u> 12 «Інформаційні технології»</p> <p style="text-align: center;"><u>Спеціальність:</u> 122 «Комп'ютерні науки»</p> <p style="text-align: center;"><u>Освітня програма:</u> «Інформаційні технології проектування»</p> <p style="text-align: center;"><u>Рівень вищої освіти: перший (бакалаврський)</u></p>	Навчальний рік
Модулів - 2		2019/2020
		Семестр
Змістових модулів - 4		2-й
		Лекції¹⁾
Розрахункова робота «Випадкові величини та функції випадкових величин»		24 год.
Загальна кількість годин – 48/90		Практичні — год.
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 2,6		Лабораторні роботи¹⁾ 24 год.
		Самостійна робота
		42 год.
	Вид контролю: іспит	

Примітка

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: – 48/90.

¹⁾ Аудиторне навантаження може бути збільшене або зменшене на одну годину в залежності від розкладу занять.

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів - 3	<p style="text-align: center;"><u>Галузь знань:</u> 12 «Інформаційні технології»</p> <p style="text-align: center;"><u>Спеціальність:</u> 126 «Інформаційні системи та технології»</p> <p style="text-align: center;"><u>Освітня програма</u> «Інформаційні системи та технології підтримки віртуальних середовищ»</p> <p style="text-align: center;"><u>Рівень вищої освіти: перший (бакалаврський)</u></p>	Навчальний рік
Модулів - 2		2019/2020
Змістових модулів - 4		Семестр
		2-й
		Лекції¹⁾
		24 год.
Розрахункова робота «Випадкові величини та функції випадкових величин»		Практичні — год.
Загальна кількість годин – 48/90		Лабораторні роботи¹⁾ 24 год.
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 2,6		Самостійна робота
		42 год.
	Вид контролю: іспит	

Примітка

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: – 48/90.

¹⁾ Аудиторне навантаження може бути збільшене або зменшене на одну годину в залежності від розкладу занять.

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів - 3	<u>Галузь знань:</u> 12 «Інформаційні технології» <u>Спеціальність:</u> 122 «Комп'ютерні науки» <u>Освітня програма:</u> «Системне проектування» <u>Рівень вищої освіти: перший</u> (бакалаврський)	Навчальний рік
Модулів - 2		2019/2020
Змістових модулів - 4		Семестр
		2-й
		Лекції¹⁾
Розрахункова робота «Випадкові величини та функції випадкових величин»		24 год.
		Практичні — год.
Загальна кількість годин – 48/90		Лабораторні роботи¹⁾ 24 год.
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 2,6		Самостійна робота
		42 год.
	Вид контролю: іспит	

Примітка

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: – 48/90.

¹⁾ Аудиторне навантаження може бути збільшене або зменшене на одну годину в залежності від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета дисципліни – розвинення ймовірнісного мислення, засвоєння термінології та понять теорії випадкових процесів, статистичного опису процесів і систем, лінійних перетворень випадкових функцій, їх канонічних представлень, що широко використовуються на практиці моделей випадкових явищ.

Завдання дисциплін – формування теоретичних знань та практичних навичок використання теорії ймовірностей, математичної статистики та випадкових процесів для формалізації та математичного моделювання прикладних завдань для їх ефективної програмно-апаратної реалізації.

Міждисциплінарні зв'язки: вища математика.

Як **результат вивчення** дисципліни студент повинен **знати**:

- поняття випадкової події та випадкової величини, їх основні види;
- означення ймовірності та її властивості;
- основні формули комбінаторики; елементи алгебри подій;
- теореми додавання і множення ймовірностей;
- форми представлення випадкових величин та їх основні характеристики;
- найбільш вживані закони розподілу випадкових величин;
- поняття вибірки та методи її формування;
- методи отримання точкових та інтервальних оцінок числових характеристик сукупностей;
- правила побудови критеріїв для перевірки статистичних гіпотез;
- методи кореляційного та регресійного аналізу;
- основні поняття теорії випадкових процесів: визначення і класифікація випадкових процесів;
- закони розподілу та основні характеристики випадкових процесів;
- перетворення випадкових процесів та операції над ними;
- стаціонарні та ергодичні випадкові процеси;
- спектральний опис випадкових процесів;
- потоки подій, їх властивості та класифікація;
- ланцюги Маркова та марковські процеси з дискретним і неперервним часом.

вміти:

- обчислювати ймовірність випадкових подій з використанням основних означень та теорем;
- будувати закони розподілу та обчислювати характеристики випадкових величин;
- проводити первинну обробку статистичних даних;
- визначати точкові та інтервальні оцінки числових характеристик;
- вибирати слушні критерії для перевірки статистичних гіпотез та користуватися таблицями відповідних розподілів;
- обґрунтовано вибирати класи регресійних моделей при виявленні залежностей між величинами; обчислювати коефіцієнти рівнянь лінійних та нелінійних залежностей;
- розраховувати основні характеристики випадкових процесів за заданими законами розподілу ймовірності;
- аналізувати процеси на стаціонарність та ергодичність;
- описувати процеси в спектральній області;
- застосовувати ланцюги Маркова та марковські процеси з дискретним і неперервним часом для розв'язання практичних задач;
- застосовувати поняття та співвідношення теорії масового обслуговування для аналізу задач.

мати навички:

- використання стандартного програмного забезпечення для аналізу ймовірнісних процесів та статистичної обробки даних;
- побудови математичних моделей реальних випадкових процесів;
- самостійного розширення своїх математичних знань за допомогою літератури або навчаючих програм.

3. Програма навчальної дисципліни

Модуль 1

Змістовий модуль 1. Випадкові події та величини

Тема 1. Випадкові події

Предмет теорії ймовірностей. Події та їх класифікація на достовірні, неможливі та випадкові. Сумісні та несумісні події. Протилежні події. Повна група подій. Означення ймовірності (класичне, геометричне, статистичне) та її властивості. Формули комбінаторики, їх застосування при обчисленні ймовірностей. Алгебра подій. Поняття сумісних і несумісних випадкових подій. Формули додавання несумісних випадкових подій. Поняття залежності і незалежності випадкових подій. Умовна ймовірність та її властивості. Формули множення ймовірностей для залежних та незалежних випадкових подій. Формули додавання сумісних випадкових подій. Використання формул множення ймовірностей для оцінки надійності деяких систем. Формула повної ймовірності та формула Байєса.

Тема 2. Моделі повторних випробувань

Схема повторних незалежних випробувань. Формула Бернуллі для обчислення ймовірності та найімовірнішого числа подій. Асимптотичні форми формули Бернуллі (локальна та інтегральна теорема Муавра-Лапласа). Використання інтегральної теореми. Формула Пуассона для малої ймовірності випадкових подій.

Тема 3. Випадкові величини та їх числові характеристики

Визначення та види випадкових величин. Закон розподілу випадкових величин. Функція розподілу та її властивості. Властивості неперервної випадкової величини. Щільність розподілу та її властивості. Числові характеристики випадкових величин. Мода і медіана. Математичне сподівання та його властивості. Характеристики розсіяння випадкової величини. Моменти випадкових величин.

Змістовий модуль 2. Закони розподілу. Системи випадкових величин

Тема 4. Найбільш поширені закони розподілу випадкових величин

Найпростіші закони розподілу дискретних випадкових величин. Біномний розподіл. Розподіл Пуассона. Геометричний розподіл. Розподіл Паскаля. Гіпергеометричний розподіл. Показниковий розподіл. Рівномірний розподіл неперервної величини. Нормальний закон розподілу. Нормальна крива: вплив параметрів розподілу на її форму. Ймовірність попадання випадкової величини з нормальним законом розподілу у заданий інтервал. Ймовірність заданого відхилення. Правило трьох сигм.

Тема 5. Системи випадкових величин

Поняття про багатовимірні випадкові величини і системи випадкових величин. Функція розподілу ймовірностей системи та її властивості. Функція щільності розподілу та її властивості. Числові характеристики системи випадкових величин. Умовні закони розподілу та їх характеристики. Регресія однієї випадкової величини на іншу. Властивості регресії. Коефіцієнт кореляції і його властивості. Приклади аналізу систем випадкових величин.

Тема 6. Граничні теореми теорії ймовірностей

Поняття про закон великих чисел. Нерівність Чебишева та її значення. теорема Чебишева і теорема Бернуллі. Центральна гранична теорема (теорема Ляпунова) та її практичне застосування.

Модульний контроль

Модуль 2

Змістовий модуль 3. Основи математичної статистики. Точкові та інтервальні оцінки.

Тема 7. Основні поняття математичної статистики

Завдання математичної статистики. Вибірковий метод і його основні поняття. Варіаційний ряд. Емпіричний закон розподілу, емпірична функція розподілу. Полігон та гістограма. Оцінки характеристик положення. Оцінки характеристик розсіяння. Оцінка коефіцієнта кореляції системи випадкових величин. Інтервальна оцінка математичного сподівання. Інтервальна оцінка середнього квадратичного відхилення. Інтервальна оцінка коефіцієнта кореляції.

Тема 8. Статистична перевірка статистичних гіпотез

Загальне поняття про перевірку гіпотез. Основна (нульова) та альтернативна гіпотези, помилки першого та другого роду. Статистичні критерії, розрахункові та критичні значення критеріїв, потужність критерію. Загальна схема перевірки статистичних гіпотез.

Порівняння математичних сподівань. Порівняння математичного сподівання і сталої. Порівняння двох дисперсій. Порівняння відносної частоти і ймовірності. Перевірка гіпотези про вид закону розподілу генеральної сукупності за допомогою критерію χ^2 .

Тема 9. Елементи кореляційного та регресійного аналізу

Функціональна, статистична та кореляційна залежність. Вибірковий коефіцієнт кореляції та його інтервальна оцінка. Перевірка гіпотези про значущість коефіцієнта кореляції. Поняття про регресію. Лінійна регресія. Прямі регресії та знаходження їх параметрів методом найменших квадратів. Найпростіші випадки нелінійної регресії. Поняття про множинну регресію.

Змістовий модуль 4. Випадкові процеси.

Тема 10. Випадкові процеси

Основні означення. Класифікація випадкових процесів. Елементарні випадкові функції. Закони розподілу випадкових процесів.

Тема 11. Характеристики випадкових процесів

Математичне сподівання випадкового процесу. Дисперсія випадкового процесу. Нормована кореляційна функція випадкового процесу. Взаємна кореляційна функція двох стахостичних процесів. Нормована взаємна кореляційна функція двох стахостичних процесів.

Тема 12. Однорідні ланцюги Маркова. Стаціонарні випадкові процеси

Означення ланцюга Маркова. Опис і зображення ланцюгів Маркова. Властивості ланцюгів Маркова. Матриця переходу. Класифікація станів. Поняття марковського процесу. Марковські випадкові процеси з дискретним часом. Потоки подій. Найпростіший потік подій. Марковські дискретні процеси, характерні неперервним часом. Марковські процеси типу «загибель-розмноження». Рівняння Колмогорова. Граничні ймовірності станів. Поняття про стаціонарні випадкові процеси. Спектральні зображення стаціонарних процесів. Ергодична властивість стаціонарних випадкових процесів.

Модульний контроль

4. Структура навчальної дисципліни

Назва змістових модулів і тем	Кількість годин (денна форма)					
	усьо- го	у тому числі				
		лк	пр	лаб	інд	с.р.
1	2	3	4	5	6	7
Модуль 1						
Змістовий модуль 1. Випадкові події та величини						
Тема 1. Випадкові події	6	2	—	2	—	2
Тема 2. Моделі повторних випробувань	8	2	—	2	—	4
Тема 3. Випадкові величини та їх числові характеристики	10	4	—	2	—	4
Разом за змістовим модулем 1	24	8	—	6	—	10
Змістовий модуль 2. Закони розподілу. Системи випадкових величин.						

Тема 4. Найбільш поширені закони розподілу випадкових величин	8	2	—	2	—	4
Тема 5. Системи випадкових величин	12	4	—	2	—	6
Тема 6. Граничні теореми теорії ймовірностей	6	2	—	2	—	2
Разом за змістовим модулем 2	26	8	—	6	—	12
Усього годин за модулем 1	50	16	—	12	—	22
Модуль 2						
Змістовий модуль 3. Основи математичної статистики. Точкові та інтервальні оцінки.						
Тема 7. Основні поняття математичної статистики	5	1	—	2	—	2
Тема 8. Статистична перевірка статистичних гіпотез	10	2	—	4	—	4
Тема 9. Елементи кореляційного та регресійного аналізу	10	2	—	2	—	6
Разом за змістовим модулем 3	25	5	—	8	—	12
Змістовий модуль 4. Випадкові процеси.						
Тема 10. Випадкові процеси	5	1	—	2	—	2
Тема 11. Характеристики випадкових процесів	6	1	—	2	—	3
Тема 12. Однорідні ланцюги Маркова. Стаціонарні випадкові процеси	4	1	—	—	—	3
Разом за змістовим модулем 4	15	3	—	4	—	8
Усього годин за модулем 2	40	8	—	12	—	20
Усього годин	90	24	—	24	—	42

5. Теми семінарських занять

Семінарські заняття навчальним планом не передбачені

6. Теми практичних занять

Практичні заняття навчальним планом не передбачені

7. Теми лабораторних занять

№ п/п	Назва теми	Кількість годин
1	Комбінаторика. Ймовірність. Теореми додавання і множення ймовірностей	2
2	Формула повної ймовірності та формула Байєса. Повторні незалежні випробування	2
3	Числові характеристики дискретних та неперервних випадкових величин	2
4	Закони розподілу дискретних та неперервних випадкових величин	2
5	Аналіз систем випадкових величин	2
6	Граничні теореми теорії ймовірностей	2
7	Точкові оцінки параметрів генеральної сукупності. Інтервальні оцінки числових характеристик	2
8	Статистична перевірка статистичних гіпотез	4
9	Елементи кореляційного та регресійного аналізу	2
10	Випадкові процеси	2
11	Характеристики випадкових процесів	2

Разом	24
-------	----

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1.	Випадкові події. Алгебра подій. Поняття сумісних і несумісних випадкових подій. Формули додавання несумісних випадкових подій. Поняття залежності і незалежності випадкових подій. Умовна ймовірність та її властивості. Формули множення ймовірностей для залежних та незалежних випадкових подій.	2
2.	Моделі повторних випробувань. Асимптотичні форми формули Бернуллі (локальна та інтегральна теорема Муавра-Лапласа).	4
3.	Випадкові величини та їх числові характеристики.	4
4.	Найбільш поширені закони розподілу випадкових величин	4
5.	Умовні закони розподілу та їх характеристики. Регресія однієї випадкової величини на іншу. Властивості регресії. Коефіцієнт кореляції і його властивості. Приклади аналізу систем випадкових величин.	6
6.	Поняття про закон великих чисел. Нерівність Чебишева та її значення. теорема Чебишева і теорема Бернуллі. Центральна гранична теорема (теорема Ляпунова) та її практичне застосування.	2
7.	Оцінка коефіцієнта кореляції системи випадкових величин. Інтервальна оцінка математичного сподівання. Інтервальна оцінка середнього квадратичного відхилення. Інтервальна оцінка коефіцієнта кореляції.	2
8.	Порівняння математичних сподівань. Порівняння математичного сподівання і сталої. Порівняння двох дисперсій. Порівняння відносної частоти і ймовірності. Перевірка гіпотези про вид закону розподілу генеральної сукупності за допомогою критерію χ^2 .	4
9.	Поняття про регресію. Лінійна регресія. Прямі регресії та знаходження їх параметрів методом найменших квадратів. Найпростіші випадки нелінійної регресії. Поняття про множинну регресію.	6
10.	Випадкові процеси	2
11.	Характеристики випадкових процесів	3
12.	Поняття марковського процесу. Марковські випадкові процеси з дискретним часом. Потіки подій. Найпростіший потік подій. Марковські дискретні процеси, характерні неперервним часом. Марковські процеси типу «загибель-розмноження».	3
	Разом	42

9. Індивідуальні завдання

Виконання розрахункової роботи «Випадкові величини та функції випадкових величин»

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, лабораторних робіт, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами, опублікованими кафедрою (методичні посібники).

11. Методи контролю

Проведення поточного контролю у вигляді тестів, усної задачі індивідуальних робіт,

письмового модульного контролю, фінальний контроль у вигляді семестрового контролю: звлік (проводиться у разі відмови студента від балів поточного тестування та за наявності допуску).

12. Критерії оцінювання та розподіл балів, які отримують студенти

12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Модуль 1			
Виконання лабораторних робіт	0...3	7	0...21
Модульний контроль	0...23	1	0...23
Модуль 2			
Виконання лабораторних робіт	0...3	7	0...21
Модульний контроль	0...20	1	0...20
Виконання та захист розрахункової роботи	0...15	1	0...15
Всього за семестр			0...100

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування та за наявності допуску. При складанні іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з двох теоретичних та трьох практичних завдань. За кожне теоретичне питання та практичне завдання студент може отримати до 20 балів. Максимальна сума всіх балів – 100.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

знати основні поняття теорії ймовірностей та математичної статистики такі як стохастичний експеримент, випадкові події, означення ймовірності, дискретні та неперервні випадкові величин, закон розподілу та числові характеристики, моменти випадкової величини. Основні закони розподілу випадкових величин: біноміальний, Пуассона, геометричний, гіпергеометричний, рівномірний, показників, нормальний, t-розподіл Стьюдента. χ^2 -розподіл, F- розподіл Фішера-Снедекора. Закон великих чисел, граничні теореми та їх застосування. Генеральна та вибіркові сукупності, статистичні ряди розподілу вибірки, полігон та гістограмам частот, вибірковий метод. оцінки параметрів (точкові і інтервальні) оцінки параметрів вибірки, статистичні гіпотези, критерії узгодження Пірсона, Фішера, Стьюдента, Колмогорова, критерій χ^2 про вигляд розподілу.

Необхідний обсяг умінь для одержання позитивної оцінки:

уміти самостійно вивчати математичну літературу; розвинути логічне, абстрактне і алгоритмічне мислення; підбирати математичні методи для створення та аналізу математичної моделі чи задачі; строго викладати свої думки; оволодіти методами математичної статистики дослідження прикладних задач; застосовувати методи розв'язання ймовірносних задач; набути навичок самостійного пошуку математичних методів та підбору моделі в літературі з математичних дисциплін.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити всі лабораторні роботи. Застосовувати основні основні поняття теорії ймовірностей та математичної статистики для розв'язання задач. Відповіді студента розкривають суть питань без достатньої повноти і обґрунтування, або у відповідях є неправильне тлумачення окремих понять та неточність у формулюванні відповідних термінів.

Добре (75-89). Твердо знати матеріал, захистити всі лабораторні роботи. Розв'язувати більш складні задачі. Вміти оцінювати область використання ймовірнісних методів. У відповідях студента можуть допускатися окремі помилки непринципового характеру, які не впливають на розкриття суті теоретичних питань. Завдання в цілому виконуються без помилок, але в обґрунтуванні розв'язання є певні недоліки.

Відмінно (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та

уміти застосовувати їх.

Курсову роботу не передбачено навчальним планом.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

14. Рекомендована література

14.1. Базова

1. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. Пособие для студентов вузов. - 6-е изд., доп. – М.: Высшая школа, 2002. – 405 с.
2. Авраменко В. І., Карімов І. К. Теорія ймовірностей і математична статистика: Навчальний посібник. – Дніпродзержинськ: ДДТУ, 2009. – 254 с.
3. Волощенко А.Б. Теорія ймовірностей та математична статистика: навч.- метод. Посібник / А.Б. Волощенко, І.А. Джалладов. - К.: КНЕУ, 2005. – 256 с. – ISBN966–574–459–3.
4. Сеньо П.С. Випадкові процеси: підручник / П.С.Сеньо. Міністерство освіти і науки України, ЛНУ. – Львів: Компакт-ЛВ, 2006. – 288 с.– ISBN966–96414–7–0.

14.2. Допоміжна

5. Практикум з теорії ймовірностей та математичної статистики: Навч. посібник / Р.К. Чорней, О.Ю. Дюженкова, О.Б. Жильцов та ін.. За редакцією Р. К. Чорней. - К.: МАУП, 2003.- 328 с.
6. Грищенко В. О. Теорія ймовірностей і математична статистика. Практикум: Навч. посіб. – К.: Київ. нац. торг.-екон. ун-т. 2002. – 164 с.
7. Турчин В. М. Теорія ймовірностей і математична статистика: Основні поняття, приклади, задачі. — Дніпропетровськ: ДНУ, 2006. — 476 с.
8. Толбатов Ю. А. Математична статистика і задачі оптимізації в алгоритмах і програмах – К.: Вища школа, 1994. – 399 с.
9. Новицький І.В. Випадкові процеси. Навчальний посібник / І.В.Новицький, С.А. Ус. – Дніпропетровськ: Національний гірничий університет, 2011. – 125 с.
10. Вентцель Е.С. Теория случайных процессов и её инженерные приложения: учеб. пособие / Е.С. Вентцель, Л.А. Овчаров. – 4-е изд. – М. Высш. шк., 2007. – 479 с. – ISBN 978–5–06–005820–8.

Вступ

Теорія ймовірності, ймовірнісні процеси і математична статистика є самостійними дисциплінами, апарат яких широко використовується в теорії надійності, дослідженні операцій, системному аналізі тощо.

Народження теорії ймовірності відноситься до середини XVII сторіччя й зв'язане з вивченням азартних ігор, хоча роздуми про випадкове були вже у античні часи, але математичні обчислення ймовірності з'являються у письмових джерелах лише з XV сторіччя. Першою книгою з теорії ймовірності є "Книга про гру у кості" Джероламо Кардано (1501-1576), яка була опублікована лише в 1663 році. Побудовою принципово нових для свого часу ймовірнісних моделей займалися такі видатні вчені, як П.Ферма (1601-1665), Б.Паскаль (1623-1662), Х.Гюйгенс (1629-1695), Я.Бернуллі (1654-1705), П.Лаплас (1749-1827), К.Гаус (177-1855). Майже два сторіччя основним полем застосування теорії ймовірності були азартні ігри, страхування та демографія.

На початку XIX сторіччя в Росії була утворена петербурзька школа, яка поставила теорію ймовірностей на потужну математичну основу, що сприяло її виділенню у самостійну дисципліну. До цієї школи належать В.Я.Буняковський (1804-1889), П.Л.Чебишов (1821-1894), А.М.Ляпунов (1857-1918), А.Я.Хінчин (1894-1958). А.А.Марков (1856-1922) заклав основи теорії ймовірнісних процесів, а А.Н.Колмогоров (1903-1958) є одним з центральних творців теорії ймовірнісних процесів.

1. Алгебра подій. Класичне та геометричне означення ймовірності

Основні поняття

Під *подією* будемо розуміти будь-яке явище, що відбувається або не відбувається. Різні події позначаються великими літерами A, B, \dots

Подія Ω називається *вірогідною*, якщо вона напевно відбудеться (наприклад, випадання не більше шести очок при одному киданні стандартного грального кубика, поява білої кульки при її витягуванні з урни, що містить тільки білі кульки тощо).

Подія \emptyset називається *неможливою*, якщо вона не може відбутися за жодних обставин (наприклад, випадання більше шести очок при одному киданні стандартного грального кубика, поява білої кульки при її витягуванні з урни, що містить тільки чорні кульки тощо).

Подія \bar{A} називається *протилежною* до події A , якщо вона відбувається тоді і тільки тоді, коли подія A не відбувається (наприклад, A - випадання парної кількості очок при киданні стандартного грального кубика, \bar{A} - випадання непарної кількості очок при киданні стандартного грального кубика тощо). В теорії множин $\bar{A} = \Omega \setminus A$ - доповнення множини A до Ω .

Події A та B називаються *несумісними*, якщо поява однієї з них виключає появу іншої (наприклад, випадання шести очок та випадання двох очок при одному киданні стандартного грального кубика, навпаки, випадання кількості очок, що є кратною трьом, та випадання непарної кількості очок при одному киданні стандартного грального кубика не є несумісними). В теорії множин $A \cap B = \emptyset$. Очевидно, що A та \bar{A} завжди будуть несумісними, але не навпаки.

Сумою подій A та B називається подія, яка відбувається тоді і тільки тоді, коли відбувається принаймні одна з цих подій. В теорії множин $A \cup B$.

Добутком подій A та B називається подія, яка відбувається тоді і тільки тоді, коли відбувається і подія A , і подія B одночасно. В теорії множин $A \cap B$.

Розглянемо деяку сукупність подій A, B, \dots, K . Ці події прийнято вважати *єдино можливими*, якщо в результаті кожного експерименту хоча б одна з них обов'язково відбудеться. Таку множину попарно несумісних подій називають *повною групою подій*. Наприклад, випадання одного, двох, трьох, чотирьох, п'яти, шести очок при одному киданні стандартного грального кубика. Протилежні події утворюють повну групу подій.

Класичне та геометричне означення ймовірності

Нехай події A_1, A_2, \dots, A_n мають наступні властивості:

- 1) попарно несумісні ($A_i \cap A_j = \emptyset, i, j = \overline{1, n}, i \neq j$);
- 2) єдино можливі;
- 3) рівно можливі (не існує жодних об'єктивних причин вважати, що одна з подій могла б з'явитися частіше ніж інша).

Події A_1, A_2, \dots, A_n можуть називатися елементарними подіями, варіантами. Припустимо, що з загальної кількості n подій появи події A сприяють m подій. Тоді *класичне означення* ймовірність події виражається формулою

$$P(A) = m/n,$$

де m - кількість елементарних подій, що сприяють появі події A ; n - кількість можливих елементарних подій.

Статистичне означення ймовірності виражається через відносну частоту

події A за формулою

$$W(A) = m/n,$$

де m – кількість випробувань, в яких подія A з'явилася; n – загальна кількість випробувань.

Геометричне означення ймовірності виражається формулою

$$P(A) = \text{міра } m / \text{міра } n,$$

де "міра m " - кількість варіантів, що сприяють появі події A ; "міра n " - загальна кількість можливих варіантів; в ролі "міри" виступають геометричні міри довжин, площ, об'ємів, величин кутів тощо.

Елементи комбінаторики

Основний принцип комбінаторики. Якщо дію A_1 можна виконати n_1 способами, дію A_2 – n_2 способами і т.д., а дію A_k – n_k способами, то ці дії одночасно можна виконати $n = n_1 \cdot n_2 \cdot \dots \cdot n_k$ способами.

Задача про комбінації. Скількома способами можна серед n різних елементів вибрати m елементів:

$$C_n^m = \frac{n!}{m!(n-m)!}.$$

Задача про розміщення. Скількома способами можна серед n різних елементів вибрати m елементів та розмістити по m різних місцях:

$$A_n^m = \frac{n!}{(n-m)!}.$$

Задача про перестановки. Скількома способами можна переставити n різних елементів по n різних місцях: $P_n = n!$

Приклад 1.1. В бригаді, що складається з 6 робітників, працюють 3 робітники 6 розряду. За табельними номерами на виконання роботи направляються 3 робітники. Яка ймовірність, що серед них буде 2 робітника 6 розряду?

Розв'язання. Нехай $A = \{\text{серед трьох робітників, які працюють, двоє мають 6 розряд}\}$. $P(A) = \frac{m}{n}$, де $m = C_3^2 \cdot C_3^1 = 9$, $n = C_6^3 = 20$, $P(A) = \frac{9}{20} = 0,45$.

Запитання для самоперевірки

1. Дайте означення події. Які бувають події?
2. Дайте означення ймовірності події. Сформулюйте властивості ймовірності.
3. Сформулюйте правило добутку в комбінаториці.

4. Сформулюйте задачу про комбінації в комбінаториці.
5. Сформулюйте задачу про розміщення в комбінаториці.
6. Сформулюйте задачу про перестановки в комбінаториці.

2. Складні ймовірності

Теорема про додавання ймовірностей

1) Нехай A та B – дві несумісні події. Тоді ймовірність того, що відбудеться хоча б одна з двох подій, дорівнює сумі їх ймовірностей

$$P(A + B) = P(A) + P(B).$$

Доведення. Нехай A_1, A_2, \dots, A_n - повна група подій, серед яких m_1 подій сприяють появі події A , а m_2 подій сприяють появі події B ($P(A) = \frac{m_1}{n}$, $P(B) = \frac{m_2}{n}$). Так як події A та B несумісні, то ніяка подія $A_j, j = \overline{1, n}$ не може сприяти обом цим подіям, тоді події $A + B$ сприяє $m_1 + m_2$ подій:

$$P(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B),$$

що й намагалися довести. Узагальнюючи, якщо A, B, \dots, K - несумісні події, то

$$P(A + B + \dots + K) = P(A) + P(B) + \dots + P(K).$$

Наслідок. Якщо A_1, A_2, \dots, A_n утворюють повну групу подій, то

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1.$$

Частковий випадок: $P(A) + P(\bar{A}) = 1$.

2) Нехай A та B – несумісні події. Тоді ймовірність того, що відбудеться хоча б одна з двох подій, дорівнює сумі їх ймовірностей без ймовірності їх одночасної появи

$$P(A + B) = P(A) + P(B) - P(A \cdot B).$$

Доведення. Нехай A_1, A_2, \dots, A_n - повна група подій, серед яких m_1 подій сприяють появі тільки події A , m_2 подій сприяють появі тільки події B , а m_3 подій сприяють появі події $A \cdot B$ ($P(A) = \frac{m_1}{n}$, $P(B) = \frac{m_2}{n}$, $P(A \cdot B) = \frac{m_3}{n}$):

$$P(A + B) = \frac{m_1 + m_2 + m_3}{n} = \frac{m_1 + m_3 + m_2 + m_3 - m_3}{n} =$$

$$\frac{m_1 + m_3}{n} + \frac{m_2 + m_3}{n} - \frac{m_3}{n} = P(A) + P(B) - P(A \cdot B),$$

що й намагалися довести.

У випадку трьох сумісних подій ймовірність їх суми обчислюється за формулою

$$P(A + B + C) = P(A) + P(B) + P(C) - P(A \cdot B) - P(A \cdot C) - P(B \cdot C) +$$

$$+P(A \cdot B \cdot C).$$

У випадку n сумісних подій ймовірність їх суми обчислюється за формулою

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i,j} P(A_i \cdot A_j) + \sum_{i,j,k} P(A_i \cdot A_j \cdot A_k) - \dots - (-1)^{n-1} P(\prod_{i=1}^n A_i).$$

Теорема про добуток ймовірностей

Дві події називаються *незалежними*, якщо поява однієї з них не залежить від того, відбулася чи не відбулася інша подія.

Дві події називаються *залежними*, якщо поява однієї з них залежить від того, відбулася чи не відбулася інша подія.

Умовною ймовірністю події називається ймовірність появи події, яка обчислюється у припущенні, що всі попередні події відбулися: $P(B/A)$ або $P_A(B)$. Якщо події незалежні, то $P(B/A) = P(B/\bar{A}) = P(B)$.

Теорема. 1) Ймовірність одночасної появи двох подій дорівнює добутку ймовірності однієї події та ймовірності іншої події при умові, що перша мала місце:

$$P(A \cdot B) = P(A) \cdot P(B/A), \text{ якщо } A \text{ та } B - \text{ залежні події.}$$

Доведення. Якщо A_1, A_2, \dots, A_n утворюють повну групу подій, то серед них:

- 1) k подій сприяють як появі події A , так і появі події B ($A \cdot B$),
- 2) l подій сприяють появі події A , але не сприяють появі події B ($A \cdot \bar{B}$),
- 3) m подій не сприяють появі події A , але сприяють появі події B ($\bar{A} \cdot B$),
- 4) $n-k-l-m$ подій як не сприяють появі події A , так і не сприяють появі події B ($\bar{A} \cdot \bar{B}$).

Обрахуємо ймовірності: $P(A \cdot B) = \frac{k}{n}$ (перша група), $P(A) = \frac{k+l}{n}$ (перша та друга групи), $P(B) = \frac{k+m}{n}$ (перша та третя групи), $P(B/A) = \frac{k}{k+l}$ (так як подія A мала місце, то загальну кількість можливих подій складають перша та друга групи, серед яких тільки перша сприяє появі B). Запишемо очевидну тотожність $\frac{k}{n} = \frac{k+l}{n} \cdot \frac{k}{k+l}$. Підставивши відповідні ймовірності, отримаємо $P(A \cdot B) = P(A) \cdot P(B/A)$, що й намагалися довести.

Наслідок. Ймовірність сумісної появи декількох подій:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2/A_1) \cdot \dots \cdot P(A_n/A_1 A_2 \dots A_{n-1})$$

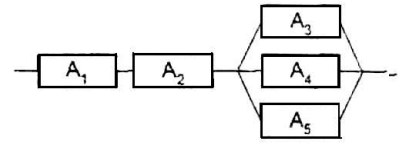
2) Якщо події A та B незалежні, то $P(A \cdot B) = P(A) \cdot P(B)$, бо

$$P(B/A) = P(B).$$

Наслідок . Ймовірність сумісної появи декількох незалежних подій:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n)$$

Приклад 2.1. Схема електричного ланцюга наведена на малюнку. Ймовірності виходу з ладу його елементів наступні: $p(A_3)=p(A_1)=p(A_2)=0,4$, $p(A_4)=p(A_5)=0,2$. Знайти ймовірність виходу з ладу всього ланцюга, якщо його елементи працюють незалежно.



Розв'язання. Нехай $A=\{\text{вихід з ладу елементів } A_1 \text{ та } A_2\}$, $B=\{\text{вихід з ладу елементів } A_3, A_4 \text{ та } A_5\}$. $P(A) = p(A_1) + p(A_2) - p(A_1) \cdot p(A_2) = 0,64$, $P(B) = p(A_3) \cdot p(A_4) \cdot p(A_5) = 0,016$, $P = p(A) + p(B) - p(A) \cdot p(B) = 0,64576$.

Запитання для самоперевірки

1. Які випадки розглядає теорема про суму подій?
2. Які випадки розглядає теорема про добуток подій?

3.Формула повної ймовірності та формула Байєса

Формула повної ймовірності

Нехай події (гіпотези) H_1, H_2, \dots, H_n утворюють повну групу подій і при настанні кожної з них подія A може з'явитися з деякою умовною ймовірністю $P(A/H_i)$, $i = \overline{1, n}$. Тоді $A = A \cdot H_1 + A \cdot H_2 + \dots + A \cdot H_n$, застосовуючи теорему про суму незалежних подій,

$$P(A) = P(A \cdot H_1) + P(A \cdot H_2) + \dots + P(A \cdot H_n),$$

застосовуючи теорему про добуток подій до кожного доданку,

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + \dots + P(H_n) \cdot P(A/H_n),$$

$$\text{де } P(H_1) + P(H_2) + \dots + P(H_n) = 1.$$

Отримали формулу повної ймовірності, за якою ймовірність події A , яка може з'явитися лише при появі однієї з несумісних подій (гіпотез) H_1, H_2, \dots, H_n дорівнює сумі добутків ймовірностей кожної з гіпотез на відповідну умовну ймовірність події A .

Формула Байєса

Формула Байєса дає відповідь на питання: якою буде ймовірність гіпотези $H_i, i = \overline{1, n}$, після випробування при припущенні, що в результаті випробування подія A відбулася.

Нехай $P(H_i), i = \overline{1, n}$ - ймовірності гіпотез до проведення випробування A , $P(H_i/A), i = \overline{1, n}$ - ймовірності тих же гіпотез після проведення випробування A ($\sum_{i=1}^n P(H_i/A) = 1$).

Користуючись теоремою про добуток подій,

$$P(A \cdot H_i) = P(H_i) \cdot P(A/H_i) = P(A) \cdot P(H_i/A), i = \overline{1, n}.$$

Звідси
$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}, i = \overline{1, n}.$$

Наслідок. Якщо $P(H_1) = P(H_2) = \dots = P(H_n)$, то

$$P(H_i/A) = \frac{P(A/H_i)}{\sum_{i=1}^n P(A/H_i)}, i = \overline{1, n}.$$

Приклад 3.1. На склад фірми надходить електронне обладнання з трьох заводів у співвідношенні 2:3:5. Ймовірність поставки бракованої партії з першого заводу – 0,1, другого – 0,2, третього – 0,3. Знайти ймовірність, що нова партія буде стандартною.

Розв'язання. Нехай $A = \{\text{нова партія стандартна}\}$, $H_1 = \{\text{партію поставив перший завод}\}$, $H_2 = \{\text{партію поставив другий завод}\}$, $H_3 = \{\text{партію поставив третій завод}\}$. Тоді $P(H_1) = 0,2, P(H_2) = 0,3, P(H_3) = 0,5, P(A/H_1) = 0,9, P(A/H_2) = 0,8, P(A/H_3) = 0,7$.

$$P(A) = 0,2 \cdot 0,9 + 0,3 \cdot 0,8 + 0,5 \cdot 0,7 = 0,77.$$

Приклад 3.2. Розглядається задача 3.1. На склад надійшла партія електронного обладнання, що задовольняє стандарту. Знайти ймовірність того, що це була поставка з першого заводу.

Розв'язання.
$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{P(A)} = \frac{0,2 \cdot 0,9}{0,77} = 0,234.$$

Запитання для самоперевірки

1. Які властивості гіпотез?
2. Яка принципова різниця у застосуванні формули повної ймовірності та формули Байєса?

4.Схема і формула Бернуллі

Повторні випробування. Формула Бернуллі

Якщо відбуваються випробування, при яких можливі лише два наслідки: подія A (успіх) або подія \bar{A} (неуспіх), при цьому ймовірність p події A є незмінною, тобто не залежить від номера випробування, то такі випробування називаються повторними, або *схемою Бернуллі*. Терміни успіх або неуспіх умовні.

Теорема Бернуллі. Ймовірність того, що в n незалежних випробуваннях, в кожному з яких ймовірність появи події A дорівнює p , подія A з'явиться рівно k разів, дорівнює $P_n(k) = C_n^k p^k (1-p)^{n-k}$.

Доведення. Підрахуємо ймовірність того, що при n незалежних випробуваннях подія A з'явиться рівно k разів у визначеному порядку $\begin{matrix} AA \dots A & \bar{A}\bar{A} \dots \bar{A} \\ k \text{ разів} & n-k \text{ разів} \end{matrix}$ $p^k (1-p)^{n-k}$ (за теоремою добутку).

Очевидно, що ймовірність того, що при n незалежних випробуваннях подія A з'явиться рівно k разів у іншому порядку буде тією ж. Кількість всіх можливих схем з n незалежних випробувань, у яких k разів зустрічається подія A у різному порядку, дорівнює кількості комбінацій C_n^k . Тому, користуючись теоремою додавання, будемо мати

$$P_n(k) = C_n^k p^k (1-p)^{n-k},$$

що й намагалися довести.

Приклад.4.1. Кидається монета шість разів. Яка ймовірність випадання номіналу 0, 1, 2, ..., 6 разів?

Розв'язання.

$$P_6(0) = P_6(6) = \left(\frac{1}{2}\right)^6 = \frac{1}{64}; \quad P_6(1) = P_6(5) = C_6^1 \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right) = \frac{3}{32};$$
$$P_6(2) = P_6(4) = C_6^2 \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^2 = \frac{15}{64}; \quad P_6(3) = C_6^3 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^3 = \frac{5}{16}.$$

З наведеного прикладу видно, що зі зростанням k ймовірність $P_n(k)$ спочатку зростає, а потім спадає. Покажемо, що ця тенденція працює і для загального випадку.

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}; \quad P_n(k+1) = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1} (1-p)^{n-k-1};$$
$$\frac{P_n(k+1)}{P_n(k)} = \frac{n!k!(n-k)!}{(k+1)!(n-k-1)!n!} \cdot \frac{p^{k+1}(1-p)^{n-k-1}}{p^k(1-p)^{n-k}} = \frac{n-k}{k+1} \cdot \frac{p}{1-p}.$$

Дріб $\frac{p}{1-p}$ є величиною сталою, а дріб $\frac{n-k}{k+1}$ зменшується зі зростанням k , тобто

$$1 < \frac{np}{1-p} = \frac{P_n(1)}{P_n(0)} > \frac{P_n(2)}{P_n(1)} > \dots > \frac{P_n(n)}{P_n(n-1)} = \frac{p}{n(1-p)} < 1.$$

Якщо n настільки велике, що $np > 1 - p$, а $p < n(1 - p)$, то існує таке

$$k = \mu, \text{ коли } \frac{P_n(\mu+1)}{P_n(\mu)} = 1 \text{ або } \frac{n-\mu}{\mu+1} \cdot \frac{p}{1-p} = 1. \text{ Тоді } np - \mu p = \mu - \mu p + 1 - p \\ \text{або } \mu = np + p - 1.$$

Якщо μ неціле, то $\frac{P_n(\mu)}{P_n(\mu-1)} > 1$, а $\frac{P_n(\mu+1)}{P_n(\mu)} < 1$. Тоді для першої нерівності $\frac{n-\mu+1}{\mu} \cdot \frac{p}{1-p} > 1$, $(n - \mu + 1)p > \mu(1 - p)$, $\mu < np + p$. Для другої нерівності $\frac{n-\mu}{\mu+1} \cdot \frac{p}{1-p} < 1$, $np - \mu p < \mu - \mu p + 1 - p$, $\mu > np + p - 1$.

Найбільш ймовірна кількість появ події A в повторних випробуваннях оцінюється за формулою: $n \cdot p + p - 1 < \mu < n \cdot p + p$.

Теорема Пуассона

У випадку, коли n велике, а p мале, замість формули Бернуллі використовується наближена формула Пуассона.

Теорема. Нехай існує схема Бернуллі, коли $n \rightarrow \infty$, $p \rightarrow 0$, при цьому $\lambda = n \cdot p$, де λ - фіксоване додатне число. Ймовірність того, що в n випробуваннях з'явиться рівно k успіхів можна обчислити за формулою

$$P_n(k) \approx \frac{\lambda^k \cdot e^{-\lambda}}{k!}.$$

Доведення.
$$P_n(k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\ \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n =$$

$$\frac{\lambda^k}{k!} \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^{\frac{n}{\lambda}(-\lambda)}.$$

Скориставшись другою визначною границею, $\lim_{n \rightarrow \infty} P_n(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$,

що й намагалися довести.

Приклад 4.2. Ймовірність похибки, яку допускає вимірювальний прилад, дорівнює 0,01. Яка ймовірність, що при 100 вимірюваннях буде допущено більше однієї похибки?

Розв'язання. Нехай $A = \{\text{допущено більше однієї похибки}\}$.

$$P(A) = 1 - P(\bar{A}). \quad P(\bar{A}) = P_{100}(0) + P_{100}(1) = \frac{\lambda^0 \cdot e^{-\lambda}}{0!} + \frac{\lambda^1 \cdot e^{-\lambda}}{1!} \approx 0,74, \text{ де}$$

$$\lambda = 0,01 \cdot 100 = 1. \quad P(A) = 1 - 0,74 = 0,26.$$

Граничні теореми Муавра-Лапласа

Ці теореми працюють в умовах схеми Бернуллі та застосовуються при великій кількості випробувань.

Локальна теорема Муавра-Лапласа. Ймовірність того, що в n незалежних випробуваннях, в кожному з яких ймовірність появи події дорівнює p , подія з'явиться рівно k разів, наближено обчислюється за формулою: $P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x)$, де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x = \frac{k-np}{\sqrt{npq}}$, $q = 1 - p$.

Таблиця значень функції $\varphi(x)$ наведена в [1].

Інтегральна теорема Муавра-Лапласа. Ймовірність того, що в n незалежних випробуваннях, в кожному з яких ймовірність появи події дорівнює p , подія з'явиться не менше k_1 разів та не більше k_2 , наближено обчислюється за формулою:

$$P_n(k_1, k_2) = F(x_2) - F(x_1),$$

де $F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-x^2/2} dx$, $x_1 = \frac{k_1 - np}{\sqrt{npq}}$, $x_2 = \frac{k_2 - np}{\sqrt{npq}}$.

Таблиця значень функції Лапласа $F(x)$ наведена в [1].

Запитання для самоперевірки

1. Опишіть схему Бернуллі.
2. Як виглядає оцінка найбільш ймовірної кількості успіхів в схемі Бернуллі?
3. Сформулюйте у порівнянні умови застосування теореми Бернуллі, теореми Пуассона та граничні теореми Муавра-Лапласа.

5. Дискретні випадкові величини

Закон розподілу дискретної випадкової величини

Дискретною називають випадкову величину X , значеннями якої є окремі ізольовані числа $x_i, i = \overline{1, n}$, котрі вона приймає з визначеною ймовірністю $p(x_i) = p_i, i = \overline{1, n}$. Наприклад, кількість очок, що випадають при киданні грального кубика, кількість пострілів до першого влучення. Якщо перераховані всі можливі значення випадкової величини X , то вони утворюють повну групу подій, тобто $\sum_{i=1}^n p_i = 1$. Співвідношення, яке

встановлює зв'язок між значеннями випадкової величини та ймовірностями цих значень, називається законом розподілу випадкової величини X , який зручно задавати у вигляді таблиці:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Якщо кількість значень випадкової величини необмежена (приклад про стрільбу), то маємо справу з безкінечною випадковою величиною.

Числові характеристики дискретних

Математичне сподівання: $M(X) = \sum_{i=1}^n x_i \cdot p_i$ або $M(X) = \sum_{i=1}^{\infty} x_i \cdot p_i$.

Властивості математичного сподівання:

- 1) математичне сподівання сталої є сама стала: $M(C) = C$;
- 2) сталий множник можна винести за знак математичного сподівання: $M(CX) = C \cdot M(X)$;
- 3) математичне сподівання суми декількох випадкових величин дорівнює сумі їх математичних сподівань: $M(X + Y + \dots + Z) = M(X) + M(Y) + \dots + M(Z)$;
- 4) математичне сподівання добутку двох незалежних випадкових величин дорівнює добутку їх математичних сподівань: $M(X \cdot Y) = M(X) \cdot M(Y)$;

Для характеристики розкидання значень випадкової величини відносно математичного сподівання вводиться поняття дисперсії.

Дисперсія: $D(X) = M(X - M(X))^2$,

$$D(X) = \sum_{i=1}^n (x_i - M(X))^2 \cdot p_i \text{ або } D(X) = \sum_{i=1}^{\infty} (x_i - M(X))^2 \cdot p_i.$$

Властивості дисперсії:

- 1) дисперсія сталої дорівнює нулю: $D(C) = 0$;
- 2) сталий множник можна винести за знак дисперсії, підвівши його до квадрату: $D(CX) = C^2 \cdot D(X)$;
- 3) дисперсія суми двох незалежних випадкових величин дорівнює сумі їх дисперсій: $D(X + Y) = D(X) + D(Y)$.

Отримаємо іншу формулу для дисперсії:

$$D(X) = M(X - M(X))^2 = M(X^2 - 2 \cdot X \cdot M(X) + (M(X))^2) = M(X^2) - 2(M(X))^2 + (M(X))^2 = M(X^2) - (M(X))^2,$$

$$D(X) = M(X^2) - (M(X))^2.$$

Середнє квадратичне відхилення: $\sigma(X) = \sqrt{D(X)}$.

Математичне сподівання, дисперсія, середнє квадратичне відхилення є основними числовими характеристиками. Узагальненням цих характеристик виступають моменти, серед яких найбільш уживаними є *початковий момент k -го порядку* $\alpha_k = M(X^k)$ (при $k = 1$ маємо математичне сподівання) та *центральний момент k -го порядку* $\mu_k = M(X - M(X))^k$ (при $k = 2$ маємо дисперсію).

Закони розподілу дискретних випадкових величин

Біноміальний закон розподілу випадкової величини функціонує в рамках схеми Бернуллі і тому підкоряється теоремі Бернуллі:

$$P_n(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n.$$

Обрахуємо числові характеристики. Для цього розглянемо випадкову величину X як суму n випадкових величин $x_i, i = \overline{1, n}$, кожна з яких відповідає одному випробуванню та підкоряється закону

x_i	0	1
p_i	$1 - p$	p

Тоді $M(x_i) = 0 \cdot (1 - p) + 1 \cdot p = p$, $D(x_i) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p \cdot (1 - p)$, $i = \overline{1, n}$. Так як випробування незалежні, то $M(X) = M(x_1 + x_2 + \dots + x_n) = M(x_1) + M(x_2) + \dots + M(x_n) = np$,
 $D(X) = D(x_1 + x_2 + \dots + x_n) = D(x_1) + D(x_2) + \dots + D(x_n) = np(1 - p)$
 або $M(X) = np, D(X) = npq$, де $q = 1 - p$.

Закон Пуассона розподілу випадкової величини функціонує в рамках схеми Бернуллі, але при умові великої кількості випробувань ($n \rightarrow \infty$), тому підкоряється теоремі Пуассона:

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, k = 0, 1, \dots$$

Обрахуємо числові характеристики.

$$M(X) = 0 \cdot e^{-\lambda} + 1 \cdot \frac{\lambda}{1} \cdot e^{-\lambda} + 2 \cdot \frac{\lambda^2}{2!} \cdot e^{-\lambda} + \dots + k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} + \dots = \lambda \cdot e^{-\lambda} \cdot$$

$$\left(1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^{k-1}}{(k-1)!} + \dots \right) = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda \quad (\text{в дужках ряд Маклорена}).$$

$$M(X^2) = 0^2 \cdot e^{-\lambda} + 1^2 \cdot \frac{\lambda}{1} \cdot e^{-\lambda} + 2^2 \cdot \frac{\lambda^2}{2!} \cdot e^{-\lambda} + \dots + k^2 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} + \dots$$

$$= e^{-\lambda} \cdot \left(\lambda + 2 \cdot \frac{\lambda^2}{1!} + \dots + k \cdot \frac{\lambda^k}{(k-1)!} + \dots \right).$$

Візьмемо ряд Маклорена $e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^{k-1}}{(k-1)!} + \dots$ та помножимо

його на λ : $\lambda \cdot e^\lambda = \lambda + \frac{\lambda^2}{1!} + \dots + \frac{\lambda^k}{(k-1)!} + \dots$ Диференціюємо ліву та праву

частини $\lambda \cdot e^\lambda + e^\lambda = 1 + 2 \cdot \frac{\lambda}{1!} + 3 \cdot \frac{\lambda^2}{2!} + \dots + k \cdot \frac{\lambda^{k-1}}{(k-1)!} + \dots$, помножимо на

λ : $\lambda \cdot e^\lambda(\lambda + 1) = \lambda + 2 \cdot \frac{\lambda^2}{1!} + \dots + k \cdot \frac{\lambda^k}{(k-1)!} + \dots$ Справа отримали вираз,

який міститься в дужках формули для $M(X^2)$. Тоді $M(X^2) = e^{-\lambda} \cdot \lambda \cdot e^\lambda(\lambda + 1) = \lambda^2 + \lambda$, а $D(X) = M(X^2) - (M(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Отже, $M(X) = \lambda, D(X) = \lambda$.

Геометричний закон розподілу випадкової величини, яка є кількістю випробувань до першого успіху, підкоряється формулі

$$P(X = k) = q^{k-1}p, k = 0, 1, \dots,$$

де $q = 1 - p$. Обрахуємо числові характеристики.

$M(X) = 0 \cdot p + 1 \cdot p \cdot q + 2 \cdot p \cdot q^2 + \dots + k \cdot p \cdot q^k + \dots = p \cdot q \cdot (1 + 2 \cdot q + 3 \cdot q^2 + \dots + k \cdot q^{k-1} + \dots)$. Візьмемо безкінечну геометричну прогресію

$1 + q + q^2 + \dots + q^k + \dots = \frac{1}{1-q}$. Диференціюємо ліву та праву частини

$1 + 2 \cdot q + 3 \cdot q^2 + \dots + k \cdot q^{k-1} + \dots = \frac{1}{(1-q)^2}$ та підставляємо у вираз $M(X)$:

$$M(X) = p \cdot q \cdot \frac{1}{(1-q)^2} = \frac{q}{p}.$$

$M(X^2) = 0 \cdot p + 1^2 \cdot p \cdot q + 2^2 \cdot p \cdot q^2 + \dots + k^2 \cdot p \cdot q^k + \dots = p \cdot q \cdot (1 + 2^2 \cdot q + 3^2 \cdot q^2 + \dots + k^2 \cdot q^{k-1} + \dots)$. Візьмемо безкінечну геометричну

прогресію $1 + q + q^2 + \dots + q^k + \dots = \frac{1}{1-q}$. Диференціюємо ліву та праву

частини $1 + 2 \cdot q + 3 \cdot q^2 + \dots + k \cdot q^{k-1} + \dots = \frac{1}{(1-q)^2}$, помножуємо на q

$q + 2 \cdot q^2 + 3 \cdot q^3 + \dots + k \cdot q^k + \dots = \frac{q}{(1-q)^2}$ та ще раз диференціюємо

$1 + 2^2 \cdot q + 3^2 \cdot q^2 + \dots + k^2 \cdot q^{k-1} + \dots = \frac{(1-q)^2 + q \cdot (1-q)}{(1-q)^4} = \frac{p+2 \cdot q}{p^3}$.

Підставляємо у вираз $M(X^2)$: $M(X^2) = p \cdot q \cdot \frac{p+2 \cdot q}{p^3} = \frac{p \cdot q + 2 \cdot q^2}{p^2}$. Тоді

$$D(X) = \frac{p \cdot q + 2 \cdot q^2}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}.$$

Отже, $M(X) = \frac{q}{p}, D(X) = \frac{q}{p^2}.$

Запитання для самоперевірки

1. Дайте означення дискретної випадкової величини.
2. Дайте означення математичного сподівання. Які його властивості?
3. Дайте означення дисперсії. Які її властивості?
4. Запишіть формули математичного сподівання і дисперсії для дискретної випадкової величини.
5. Які закони розподілу дискретної випадкової величини ви знаєте?

6. Неперервні випадкові величини

Функція розподілу та щільність ймовірності

Неперервна випадкова величина X приймає значення на скінченному або безкінечному числовому відрізку та характеризується *функцією розподілу* (інтегральний закон розподілу) $F(x) = P(X < x),$ яка є ймовірністю того, що випадкова величина прийме значення, менше за поточне значення $x.$

Властивості функцією розподілу:

- 1) $0 \leq F(x) \leq 1,$
- 2) $F(x)$ є неспадною функцією,
- 3) $P(a < X < b) = F(b) - F(a).$

Неперервна випадкова величина може ще задаватися *щільністю ймовірності* (диференціальний закон розподілу), яка визначається як похідна від функції розподілу: $f(x) = F'(x).$

Властивості щільністю ймовірності:

- 1) $F(x) = \int_{-\infty}^x f(x)dx ,$
- 2) $\int_{-\infty}^{\infty} f(x)dx = 1 ,$
- 3) $P(a < X < b) = \int_a^b f(x)dx.$

Числові характеристики неперервних випадкових величин

Математичне сподівання: $M(X) = \int_{-\infty}^{\infty} xf(x)dx$.

Дисперсія: $D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 f(x)dx$ або $D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - (M(X))^2$.

Середнє квадратичне відхилення: $\sigma(X) = \sqrt{D(X)}$.

Властивості числових характеристик неперервних випадкових величин є аналогічними властивостям дискретних випадкових величин.

Закони розподілу неперервних випадкових величин

Рівномірний закон розподілу має щільність $f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & x \notin (a, b). \end{cases}$

Обрахуємо числові характеристики.

$$M(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2},$$
$$M(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3},$$
$$D(X) = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Нормальний закон розподілу має щільність $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}$,

Обрахуємо числові характеристики.

$$M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-a)^2/(2\sigma^2)} dx = \left| \frac{x-a}{2} = t \right| = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + a) e^{-\frac{t^2}{2}} \sigma dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 0 + \frac{a}{\sqrt{2\pi}} \cdot 2 \int_0^{\infty} e^{-\frac{t^2}{2}} dt = \frac{a}{\sqrt{2\pi}} \cdot 2 \sqrt{\frac{\pi}{2}} = a,$$

$$\begin{aligned}
D(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/(2\sigma^2)} dx = \left| \frac{x-a}{2} = t \right| \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \sigma^2 e^{-\frac{t^2}{2}} \sigma dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt \\
&= \left| \begin{array}{l} u = t \\ du = dt \\ dv = t e^{-\frac{t^2}{2}} dt \\ v = -e^{-\frac{t^2}{2}} \end{array} \right| = \frac{\sigma^2}{\sqrt{2\pi}} \left(0 + \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) = \frac{\sigma^2}{\sqrt{2\pi}} \cdot 2\sqrt{\frac{\pi}{2}} \\
&= \sigma^2,
\end{aligned}$$

$$\begin{aligned}
P(\alpha < X < \beta) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-(x-a)^2/(2\sigma^2)} dx = \left| \frac{x-a}{2} = t \right| = \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{t^2}{2}} \sigma dt = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{t^2}{2}} dt = F\left(\frac{\beta-a}{\sigma}\right) - F\left(\frac{\alpha-a}{\sigma}\right),
\end{aligned}$$

де $F(x)$ – функція Лапласа.

$$P(a - \gamma < X < a + \gamma) = F\left(\frac{\gamma}{\sigma}\right) - F\left(-\frac{\gamma}{\sigma}\right) = 2F\left(\frac{\gamma}{\sigma}\right),$$

$$P(a - 3\sigma < X < a + 3\sigma) = F(3) - F(-3) = 2F(3) = 0,9973 \quad (\text{правило трьох сигм}).$$

Теорема Ляпунова (центральна гранична теорема). Якщо \bar{X} – сума великої кількості незалежних випадкових величин, які мають довільний розподіл, вплив їх на \bar{X} незначний, то \bar{X} має розподіл, близький до нормального, або нормальний, якщо кількість випадкових величин прямує до нескінченності.

Показниковий закон розподілу має щільність $f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$

Обрахуємо числові характеристики.

$$M(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \left| \begin{array}{l} u = x \\ du = dx \\ dv = \lambda e^{-\lambda x} dx \\ v = -e^{-\lambda x} \end{array} \right| = 0 + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda},$$

$$M(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \lambda \left| \begin{array}{l} u = x^2 \\ du = 2x dx \\ dv = e^{-\lambda x} dx \\ v = -\frac{1}{\lambda} e^{-\lambda x} \end{array} \right| = \lambda \left(0 + \frac{2}{\lambda} \int_0^{\infty} x e^{-\lambda x} dx \right)$$

$$= 2 \left| \begin{array}{l} u = x \\ du = dx \\ dv = e^{-\lambda x} dx \\ v = -\frac{1}{\lambda} e^{-\lambda x} \end{array} \right| = 2 \left(0 + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \right) = \frac{2}{\lambda^2},$$

$$D(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Закон великих чисел

Під законом великих чисел в теорії ймовірності розуміють ряд теорем, у кожній з яких при врахуванні окремих умов встановлюється факт наближення середніх характеристик великої кількості дослідів до деяких сталих значень.

Нерівність Чебишова. Нехай випадкова величина X має математичне сподівання $M(X)$ і скінчену дисперсію $D(X)$. Тоді для будь-якого $\varepsilon > 0$ є справедливою нерівність:

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Протилежна подія має вигляд:

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}.$$

Теорема Чебишова. Якщо $X_i, i = 1, 2, \dots, n$ - попарно незалежні випадкові величини з рівномірно обмеженими дисперсіями $D(X_i) \leq C$, то середнє арифметичне цих випадкових величин збігається по ймовірності до середнього арифметичного їх математичних сподівань при достатньо великому n :

$$\lim_{n \rightarrow \infty} P(|X - M(X)| < \varepsilon) = 1, \text{ де } X = \frac{1}{n} \sum X_i, M(X) = \frac{1}{n} \sum M(X_i).$$

Теорема Бернуллі. При необмеженому збільшенні кількості незалежних випробувань n частота появи події A збігається по ймовірності до її ймовірності p в окремому випадку (схема Бернуллі):

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1.$$

Запитання для самоперевірки

1. Дайте означення неперервної випадкової величини.
2. Дайте означення функції розподілу. Які її властивості?
3. Дайте означення щільності. Які її властивості?
4. Запишіть формули математичного сподівання і дисперсії для неперервної випадкової величини.
5. Які закони розподілу неперервної випадкової величини ви знаєте?
6. З яких теорем складається закон великих чисел?

7. Функції та системи випадкових величин

Система випадкових величин

Системою випадкових величин називається множина двох або більше випадкових величин, що розглядаються як єдине ціле при дослідженні того чи іншого явища.

Система може бути утворена дискретними або неперервними випадковими величинами, або і дискретними, і неперервними випадковими величинами.

В залежності від кількості випадкових величин, що утворюють систему, вона називається двовимірною, тривимірною, n -вимірною. Позначки: (X, Y) , (X, Y, Z) тощо. Геометрично їх можна інтерпретувати як вектори тієї чи іншої розмірності.

Дискретна двовимірна випадкова величина

Дискретна двовимірна випадкова величина - це двовимірна випадкова величина, яка приймає скінченну кількість або послідовність різних пар значень. Для її повної характеристики достатньо вказати множину можливих пар значень (точок площини) та ймовірність кожної з них, тобто *закон розподілу*, який можна представити у вигляді таблиці:

	x_1	x_2	...	x_i	...	x_n	Σ
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$...	$p(x_i, y_1)$...	$p(x_n, y_1)$	$p(y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$...	$p(x_i, y_2)$...	$p(x_n, y_2)$	$p(y_2)$
...

y_j	$p(x_1, y_j)$	$p(x_2, y_j)$...	$p(x_i, y_j)$...	$p(x_n, y_j)$	$p(y_j)$
...
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$...	$p(x_i, y_m)$...	$p(x_n, y_m)$	$p(y_m)$
Σ	$p(x_1)$	$p(x_2)$...	$p(x_i)$...	$p(x_n)$	1

Припускаючи, що всі наведені у таблиці значення єдино можливі, то

$$\sum_i \sum_j p(x_i, y_j) = 1. \text{ За теоремою додавання}$$

$$\sum_i p(x_i, y_j) = p(y_j), \sum_j p(x_i, y_j) = p(x_i).$$

Знайдемо умовну ймовірність того, що випадкова величина Y прийме значення y_j , якщо $X=x_i$. За теоремою про добуток подій:

$$p(x_i, y_j) = p(x_i) \cdot p(y_j/x_i) \text{ або } p(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)}.$$

Сукупність значень $\{y_j\}, j = 1, 2, \dots, m$ та відповідних значень $\{p(y_j/x_i)\}, j = 1, 2, \dots, m$ утворюють умовний закон розподілу Y при сталій $X=x_i$. Аналогічно сукупність значень $\{x_i\}, i = 1, 2, \dots, n$ та відповідних значень $\{p(x_i/y_j)\}, i = 1, 2, \dots, n$ утворюють умовний закон розподілу X при сталій $Y=y_j$.

Обрахуємо числові характеристики.

$$\bar{x} = M(X) = \sum_{i=1}^n x_i \cdot p(x_i), \quad \bar{y} = M(Y) = \sum_{j=1}^m y_j \cdot p(y_j),$$

$$M(X, Y) = \sum_{i=1}^n \sum_{j=1}^m x_i \cdot y_j \cdot p(x_i, y_j).$$

Точка з координатами (\bar{x}, \bar{y}) характеризує центр розсіяння двовимірної випадкової величини.

$$D(X) = \sum_{i=1}^n (x_i - \bar{x})^2 \cdot p(x_i), \quad D(Y) = \sum_{j=1}^m (y_j - \bar{y})^2 \cdot p(y_j) \text{ або}$$

$$D(X) = \sum_{i=1}^n x_i^2 \cdot p(x_i) - \bar{x}^2, \quad D(Y) = \sum_{j=1}^m y_j^2 \cdot p(y_j) - \bar{y}^2.$$

Неперервна двовимірна випадкова величина

Неперервна двовимірна випадкова величина - це двовимірна випадкова величина, яка приймає всі значення з деякої області G площини. Така випадкова величина задається функцією розподілу $F(x, y) = P(X < x, Y < y)$. Властивості функції розподілу:

- 1) $0 \leq F(x, y) \leq 1$;
- 2) $F(x, y)$ - неспадна функція;
- 3) $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$;
- 4) $F(\infty, \infty) = 1$;
- 5) $F(x, \infty) = F(x)$; $F(\infty, y) = F(y)$;
- 6) $F(x_1 < X < x_2, y_1 < Y < y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$.

Щільністю ймовірності неперервної двовимірної випадкової величини називається функція $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$.

Властивості щільності ймовірності:

- 1) $f(x, y) \geq 0$;
- 2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$;
- 3) $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$;
- 4) $f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$, $f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$ - щільності ймовірностей випадкових величин X та Y ;
- 5) $F_1(x) = \int_{-\infty}^x dx \int_{-\infty}^{\infty} f(x, y) dy$, $F_2(y) = \int_{-\infty}^y dy \int_{-\infty}^{\infty} f(x, y) dx$ - функції розподілу випадкових величин X та Y .

Умовні закони розподілу. За теоремою про добуток подій:

$$f(x, y) dx dy = f_1(x) \cdot f(y/x) dx dy = f_2(y) \cdot f(x/y) dx dy.$$

$$\text{Тоді } f(y/x) = \frac{f(x, y)}{f_1(x)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy}, \quad f(x/y) = \frac{f(x, y)}{f_2(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx}.$$

Обрахуємо числові характеристики.

$$\bar{x} = M(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy,$$

$$\bar{y} = M(Y) = \int_{-\infty}^{\infty} y f_2(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy,$$

$$M(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy,$$

$$D(X) = \int_{-\infty}^{\infty} (x - \bar{x})^2 f_1(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x, y) dx dy,$$

$$D(Y) = \int_{-\infty}^{\infty} (y - \bar{y})^2 f_2(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \bar{y})^2 f(x, y) dx dy,$$

$$D(X) = \int_{-\infty}^{\infty} x^2 f_1(x) dx - \bar{x}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f(x, y) dx dy - \bar{x}^2,$$

$$D(Y) = \int_{-\infty}^{\infty} y^2 f_2(y) dy - \bar{y}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 f(x, y) dx dy - \bar{y}^2.$$

Приклад 7.1. Точка $M(X, Y)$ розподілена у квадраті $0 \leq x \leq \frac{\pi}{2}$, $0 \leq y \leq \frac{\pi}{2}$ з щільністю $f(x, y) = 0,5 \sin(x + y)$ та $f(x, y) = 0$ поза квадратом. Знайти $F(x, y)$, $M(X)$, $M(Y)$, $D(X)$, $D(Y)$.

Розв'язування. $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy = 0$, якщо $x < 0$ або $y < 0$.

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy = 0,5 \int_0^x \int_0^y \sin(x + y) dx dy =$$

$$0,5 \int_0^x (-\cos(x + y)) \Big|_0^y dx = 0,5 \int_0^x (\cos x - \cos(x + y)) dx = 0,5(\sin x -$$

$$\sin(x + y)) \Big|_0^x = 0,5(\sin x - \sin(x + y) + \sin y), \text{ якщо } 0 \leq x \leq \frac{\pi}{2}, 0 \leq y \leq \frac{\pi}{2}.$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy = 0,5 \int_0^x \int_0^{\frac{\pi}{2}} \sin(x + y) dx dy =$$

$$0,5 \int_0^x (-\cos(x + y)) \Big|_0^{\frac{\pi}{2}} dx = 0,5 \int_0^x (\cos x - \cos(x + \frac{\pi}{2})) dx = 0,5 \int_0^x (\cos x +$$

$$\sin x) dx = 0,5(\sin x - \cos x) \Big|_0^x = 0,5(\sin x - \cos x + 1), \text{ якщо } 0 \leq x \leq$$

$$\frac{\pi}{2}, y > \frac{\pi}{2}.$$

Аналогічно,

$$F(x, y) = 0,5(\sin y - \cos y + 1), \text{ якщо } x > \frac{\pi}{2}, 0 \leq y \leq \frac{\pi}{2}.$$

$$F(x, y) = 1, \text{ якщо } x > \frac{\pi}{2}, y > \frac{\pi}{2}.$$

$$M(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy = 0,5 \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} x \sin(x + y) dx dy =$$

$$0,5 \int_0^{\frac{\pi}{2}} x (-\cos(x + y)) \Big|_0^{\frac{\pi}{2}} dx = 0,5 \int_0^{\frac{\pi}{2}} x (\cos x - \cos(x + \frac{\pi}{2})) dx =$$

$$0,5 \int_0^{\frac{\pi}{2}} x (\cos x + \sin x) dx = \left| \begin{array}{l} \text{інтегрування} \\ \text{частинами} \end{array} \right| = 0,5(x(\sin x - \cos x)) \Big|_0^{\frac{\pi}{2}} -$$

$$\int_0^{\frac{\pi}{2}} (\sin x - \cos x) dx = 0,5 \left(\frac{\pi}{2} + (\sin x + \cos x) \Big|_0^{\frac{\pi}{2}} \right) = \frac{\pi}{4}.$$

Аналогічно,

$$M(Y) = \frac{\pi}{4}.$$

$$D(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f(x, y) dx dy - \bar{x}^2 = 0,5 \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} x^2 \sin(x + y) dx dy -$$

$$\left(\frac{\pi}{4}\right)^2 = \left| \begin{array}{l} \text{інтегрування} \\ \text{частинами} \end{array} \right| = \frac{\pi}{2} + \frac{\pi^2}{16} - 2.$$

Аналогічно,

$$D(Y) = \frac{\pi}{2} + \frac{\pi^2}{16} - 2.$$

Коваріація та кореляція

Коваріацією випадкових величин X, Y та називається число

$$k = k(X, Y) = M[(X - M(X))(Y - M(Y))].$$

Інша формула $k = k(X, Y) = M(X, Y) - M(X)M(Y).$

Коваріація характеризує залежність випадкових величин та їх розсіювання навколо точки $(M(X), M(Y))$.

Коефіцієнт кореляції між випадковими величинами X, Y називається число

$$\rho_{xy} = \frac{k(X,Y)}{\sigma_x \sigma_y}, \quad \text{де } |\rho_{xy}| \leq 1.$$

Якщо $\rho_{xy} \neq 0$, то кажуть, що X, Y - корельовані, $\rho_{xy} = 0$ - некорельовані. Чим ближче $|\rho_{xy}|$ до одиниці, тим "щільніший" зв'язок між X, Y . Якщо $|\rho_{xy}| = 1$, то маємо функціональний зв'язок. Якщо $0 < |\rho_{xy}| < 1$, то маємо статистичний зв'язок. Якщо $\rho_{xy} = 0$, то лінійний зв'язок відсутній.

Запитання для самоперевірки

1. Як задається дискретна двовимірна випадкова величина? Запишіть її числові характеристики.
2. Як задається неперервна двовимірна випадкова величина? Запишіть її числові характеристики.
3. Що характеризують коваріація і коефіцієнт кореляції.

8. Загальні поняття ймовірнісних процесів

Основні означення

Ймовірнісною функцією називається функція, що в результаті досліду може прийняти той чи інший вигляд (реалізація), який наперед невідомий. Кожна реалізація не є випадковою.

Ймовірнісний процес є ймовірнісною функцією часу t . Якщо параметр t приймає дискретні значення, то маємо процес з дискретним часом; якщо t змінюється на деякому інтервалі, то маємо процес з неперервним часом.

Стани у фіксований момент часу, по суті, представляють собою випадкову величину $\{X(t)|t \in T\}$ (повну групу подій), в яких може знаходитися система в цей момент. Кількість станів може бути скінченною та безкінечною.

Ймовірнісний процес $\{X(t)|t \in T\}$, де T – скінченний або безкінечний відрізок, називається *процесом з незалежними приростами*, якщо для будь-яких n , $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$, випадкові вектори $x(0), x(t_1) - x(0), x(t_2) - x(t_1), \dots, x(t_n) - x(t_{n-1})$ взаємно-незалежні. Саме з вивчення процесів з незалежними приростами виникла теорія ймовірнісних процесів. Спочатку вивчався вінеровський процес (або процес броунівського руху), а потім і більш загальні процеси з незалежними приростами. Так як окреме зміщення мале, то можна вважати, що до їх суми застосовується центральна гранична теорема теорії ймовірності, а броунівський рух розглядати як гаусовський процес.

Гаусові ймовірнісні процеси вимагають, щоб випадкові величини, які їх утворюють, підкорялися нормальному закону розподілу.

Марковський процес (або «процес без післядії») є ймовірнісним процесом, що має наступну властивість: для кожного моменту часу ймовірність будь-якого стану системи у майбутньому залежить тільки від її стану у теперішньому часі і не залежить від того, коли та яким чином система прийшла у цей стан (тобто як розвивався процес у минулому). Марковські процеси знайшли дуже широке застосування у кібернетиці (особливо в теорії інформації). Так як конспект лекцій розробляється для студентів спеціальності 125«Кібербезпека» та 123«Комп'ютерна інженерія», то саме на цьому класі випадкових процесів буде зроблено акцент.

Закони розподілу ймовірнісних процесів

Функція $F_1(x_1, t_1) = P(X(t_1) < x_1)$ називається *одновимірною інтегральною функцією розподілу ймовірнісного процесу*. Якщо функцію $F_1(x_1, t_1)$ диференціювати по x_1 , то отримаємо одновимірну щільність ймовірності:

$$\frac{\partial F_1(x_1, t_1)}{\partial x_1} = f_1(x_1, t_1).$$

Функція $F_2(x_1, t_1, x_2, t_2) = P(X(t_1) < x_1, X(t_2) < x_2)$ називається *двовимірною інтегральною функцією розподілу ймовірнісного процесу*. Якщо функцію $F_2(x_1, t_1, x_2, t_2)$ диференціювати по x_1 та x_2 , то отримаємо двовимірну щільність ймовірності:

$$\frac{\partial F_2(x_1, t_1, x_2, t_2)}{\partial x_1 \partial x_2} = f_2(x_1, t_1, x_2, t_2).$$

Функція $F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = P(X(t_1) < x_1, X(t_2) < x_2, \dots, X(t_n) < x_n)$ називається *n-вимірною інтегральною функцією розподілу ймовірнісного процесу*. Якщо функцію $F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)$ диференціювати по x_1, x_2, \dots, x_n , то отримаємо *n* – вимірну щільність ймовірності:

$$\frac{\partial F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n} = f_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n).$$

Характеристики ймовірнісних процесів

Математичне сподівання ймовірнісного процесу $\{X(t)|t \in T\}$ називається не випадкова функція $m_x(t)$, яка при кожному значенні аргументу t дорівнює математичному сподіванню відповідного перерізу ймовірнісного процесу $m_x(t) = M(X(t))$.

Якщо відома одновимірна щільність ймовірності $f_1(x_1, t_1)$, то $m_x(t) = \int_{-\infty}^{\infty} x f_1(x, t) dx$.

Математичне сподівання $m_x(t)$ ймовірнісного процесу є не випадкова функція, навколо якої коливаються реалізації ймовірнісного процесу.

Дисперсією ймовірнісного процесу $\{X(t)|t \in T\}$ називається не випадкова функція $D_x(t)$, яка при кожному значенні аргументу t дорівнює дисперсії відповідного перерізу ймовірнісного процесу $D_x(t) = D(X(t))$.

Якщо відома одновимірна щільність ймовірності $f_1(x_1, t_1)$, то $D_x(t) = \int_{-\infty}^{\infty} (x - m_x(t))^2 f_1(x, t) dx$ або $D_x(t) = \int_{-\infty}^{\infty} x^2 f_1(x, t) dx - (m_x(t))^2$.

Математичне сподівання та дисперсія ймовірнісного процесу дають можливість мати лише деяке уявлення про ймовірнісний процес: навколо якої функції групуються та як розкидані реалізації навколо середньої функції.

Кореляційною функцією ймовірнісного процесу $\{X(t)|t \in T\}$ називається не випадкова функція $K_x(t_1, t_2)$, яка при кожній парі t_1, t_2 дорівнює коваріації відповідних перерізів ймовірнісного процесу

$$K_x(t_1, t_2) = M((X(t_1) - m_x(t_1))(X(t_2) - m_x(t_2))).$$

Якщо відома двовимірна щільність ймовірності $f_2(x_1, t_1, x_2, t_2)$, то

$$K_x(t_1, t_2) = \iint_{-\infty}^{\infty} (x_1 - (m_x(t_1)))(x_2 - (m_x(t_2))) f_2(x_1, t_1, x_2, t_2) dx_1 dx_2$$

або

$$K_x(t_1, t_2) = \iint_{-\infty}^{\infty} x_1 x_2 f_2(x_1, t_1, x_2, t_2) dx_1 dx_2 - m_x(t_1) m_x(t_2).$$

Кореляційна функція характеризує степінь зв'язку між значеннями ймовірнісного процесу $\{X(t) | t \in T\}$ в моменти часу t_1, t_2 .

Запитання для самоперевірки

1. Дайте означення ймовірнісного процесу.
2. Якими характеристиками описується ймовірнісний процес?

9. Марковські процеси

Марковський ймовірнісний процес з дискретними станами

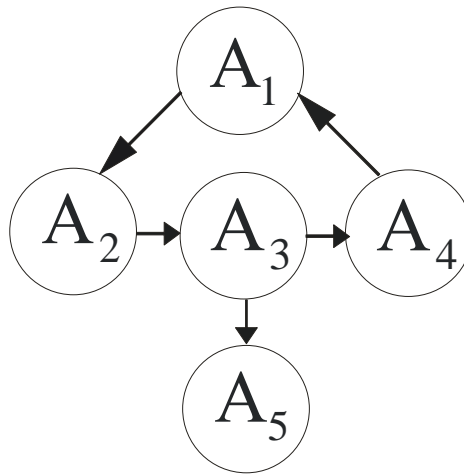
Ймовірнісний процес називається процесом з дискретними станами, якщо можливі стани системи A_1, A_2, \dots можна перерахувати (пронумерувати) один за одним, а сам процес полягає у тому, що час від часу система стрибком (миттєво) переходить з одного стану в інший.

При аналізі ймовірнісних процесів з дискретними станами зручно користуватися геометричною схемою – *графом станів*. Кожний стан будемо зображувати вершинами, а можливі переходи з одного стану в інший – орієнтованими дугами, що з'єднують ці вершини. Можливість системи залишитися у наступний момент часу в тому ж самому стані зображується петлею, що виходить і входить в одну й ту ж саму вершину.

Приклад. Система – комп'ютер, що може знаходитися в одному з можливих станів:

A_1 - комп'ютер працює; A_2 - комп'ютер вийшов з ладу, чекає на тестування; A_3 - тестування комп'ютера; A_4 - ремонт комп'ютера; A_5 списання комп'ютера.

Графічна модель функціонування системи зображена на малюнку.



Марковський ймовірнісний процес з дискретним часом

Нехай маємо систему, що може знаходитися у станах A_1, A_2, \dots, A_n , при цьому переходи системи із стану в стан можливі тільки у моменти t_1, t_2, \dots . Будемо називати ці моменти кроками, а процес, що відбувається у системі, розглядатимемо як функцію цілочисельного аргументу (номера кроку).

Умовимося позначати через A_i^k подію, яка полягає у тому, що система за k кроків опинилася у стані A_i . При будь-якому k події $A_1^k, A_2^k, \dots, A_n^k$ утворюють повну групу подій. Позначимо ймовірності цих подій через p_i^k . Тоді $p_1^k + p_2^k + \dots + p_n^k = 1$ (повна група подій).

Ці ймовірності називаються *абсолютними ймовірностями* станів.

Для будь-якого кроку існують ймовірності переходу системи з будь-якого стану у будь-який інший стан за один крок, а також ймовірності затримки системи у даному стані. Такі ймовірності називаються *перехідними* і задаються *стохастичною* матрицею:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ & & \dots & \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{pmatrix},$$

де p_{ij} – ймовірність переходу системи з стану i у стан j . Деякі з перехідних ймовірностей можуть дорівнювати нулю (це означає, що перехід системи за один крок з стану i у стан j неможливий), на головній діагоналі розташовані ймовірності затримки системи у тому ж стані на наступному кроці.

Матриця перехідних ймовірностей P разом з початковими ймовірностями p_j^0 ($j = 1, 2, \dots, n$) повністю задає *марковський ланцюг*.

Марковський ланцюг називається *однорідним*, якщо перехідні ймовірності не залежать від номера кроку, у протилежному випадку марковський ланцюг називається *неоднорідним*.

Розглянемо однорідний марковський ланцюг.

Нехай $\{p_j^k\}$ ($j = 1, 2, \dots, n$) – абсолютні ймовірності станів системи після k переходів. Величини $\{p_j^k\}$ ($j = 1, 2, \dots, n$) можна виразити через $\{p_j^0\}$ ($j = 1, 2, \dots, n$) та P , використовуючи формулу повної ймовірності :

$$\begin{aligned} p_j^1 &= p_1^0 p_{1j} + p_2^0 p_{2j} + \dots + p_n^0 p_{nj} = \sum_i p_i^0 p_{ij}, \\ p_j^2 &= \sum_i p_i^1 p_{ij} = \sum_i (\sum_m p_m^0 p_{mj}) p_{ij} = \sum_m p_m^0 (\sum_i p_{mj} p_{ij}) = \sum_m p_m^0 p_{mj}^2, \\ &\dots \\ p_j^k &= \sum_i p_i^0 (\sum_m p_{im}^{k-1} p_{mj}) = \sum_i p_i^0 p_{ij}^k, \end{aligned} \quad (2.1)$$

де p_{ij}^k – k -крокова перехідна ймовірність (ймовірність переходу системи з стану i у стан j за k кроків).

У загальному вигляді для всіх i та j :

$$p_{ij}^k = \sum_l p_{ik}^{k-l} p_{lj}^l, \quad 0 < l < k.$$

Ці рівняння відомі як рівняння Колмогорова-Чепмена.

Тобто якщо абсолютні ймовірності визначені у векторній формі $p^{(k)} = \{p_1^k, p_2^k, \dots, p_n^k\}$, то $p^{(k)} = p^{(0)} P^k$.

Марковський ймовірнісний процес з дискретними станами та неперервним часом

У попередньому розділі ми розглядали марковський ланцюг, тобто ймовірнісний процес, що відбувається у системі, яка випадковим чином може переходити з стану у стан тільки у деякі наперед визначені, фіксовані моменти часу.

На практиці значно частіше зустрічаються ситуації, коли перехід системи з стану у стан відбуваються не у фіксовані, а у випадкові моменти часу, які наперед вказати неможливо.

Для опису таких процесів може бути застосована схема марковського ймовірнісного процесу з дискретними станами та неперервним часом, який називається *неперервним ланцюгом Маркова*.

Нехай маємо ряд дискретних станів A_1, A_2, \dots, A_n , перехід системи з стану у стан може відбуватися у будь-який момент часу.

Позначимо через $p_i(t)$ ймовірність того, що в момент t система буде знаходитися у стані A_i ($i = 1, 2, \dots, n$). Очевидно, що для будь-якого моменту t сума ймовірностей станів дорівнює одиниці:

$$\sum_{i=1}^n p_i(t) = 1,$$

так як події A_1, A_2, \dots, A_n є повною групою.

Поставимо задачу – визначити для будь-якого t ймовірності $p_i(t)$ ($i = 1, 2, \dots, n$). Для того, щоб знайти ці ймовірності, необхідно знати характеристики процесу, що є аналогічними перехідним ймовірностям для марківського ланцюга. У випадку процесу з неперервним часом не прийдеться задавати визначені перехідні ймовірності, бо ймовірність переходу системи з стану у стан точно в момент t буде дорівнювати нулю (так само як ймовірність будь-якого окремого значення неперервної випадкової величини). Замість перехідних ймовірностей p_{ij} введемо інтенсивності переходу λ_{ij} .

Інтенсивністю переходу системи з стану i у стан j називається границя відношення ймовірності переходу системи за час Δt з стану i у стан j до довжини Δt :

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{p_{ij}(\Delta t)}{\Delta t},$$

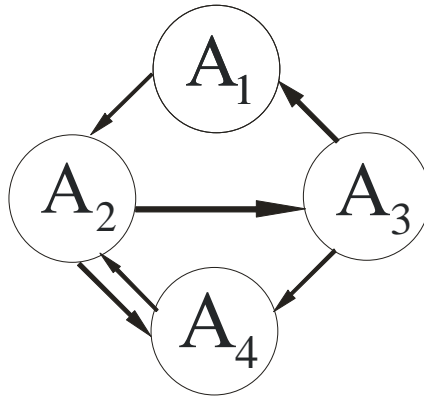
де $p_{ij}(\Delta t)$ - ймовірність того, що система, яка знаходилася в момент t у стані A_i за час Δt перейде з нього у стан A_j .

З формули випливає, що при малому Δt ймовірність переходу $p_{ij}(\Delta t)$ (з точністю до безкінечно малих вищих порядків) дорівнює $\lambda_{ij}\Delta t$.

Якщо всі інтенсивності переходів λ_{ij} не залежать від t , то марковський процес називається *однорідним*; якщо всі інтенсивності представляють собою функції часу, то процес є *неоднорідним*.

Графічною моделлю марковського ланцюга з неперервним станом є граф, у якому над кожною дугою надписана інтенсивність переходу системи з відповідного стану у відповідний стан. Знаючи такий граф, можна визначити ймовірності станів $p_1(t), p_2(t), \dots, p_n(t)$ як функції часу. А саме, ці ймовірності задовільняють диференціальним рівнянням, так званим *рівнянням Колмогорова*.

Продемонструємо методику побудови рівнянь Колмогорова для ймовірностей станів на конкретному прикладі. Нехай система має чотири можливих стани, які представлені на малюнку.



Знайдемо ймовірність $p_1(t)$. Надамо t малий приріст Δt і знайдемо ймовірність того, що у момент $t + \Delta t$ система буде знаходитися у стані A_1 . Ця подія може відбутися двома шляхами:

- в момент t система вже була у стані A_1 , а за час Δt не вийшла з цього стану або
- в момент t система була у стані A_3 , а за час Δt перейшла з нього у стан A_1 .

Ймовірність першого варіанта знайдемо як добуток ймовірності $p_1(t)$ того, що в момент t система була у стані A_1 , на умовну ймовірність того, що, знаходячись у стані A_1 , система за час Δt не перейде з нього у стан A_2 . Ця умовна ймовірність (з точністю до безкінечно малих вищих порядків) дорівнює $1 - \lambda_{12}\Delta t$.

Аналогічно, ймовірність другого варіанта дорівнює ймовірності того, що в момент t система була у стані A_3 , яка помножується на умовну ймовірність переходу за час Δt у стан A_1 :

$$p_3(t)\lambda_{31}\Delta t.$$

Застосовуючи правило додавання ймовірностей, отримаємо:

$$p_1(t + \Delta t) = p_1(t)(1 - \lambda_{12}\Delta t) + p_3(t)\lambda_{31}\Delta t.$$

Розкриємо дужки у правій частині, перенесемо $p_1(t)$ у ліву частину і розділимо обидві частини рівняння на Δt , отримаємо:

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = -\lambda_{12}p_1(t) + p_3(t)\lambda_{31}.$$

Тепер спрямуємо Δt до нуля і перейдемо до границі:

$$\lim_{\Delta t \rightarrow 0} \frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = -\lambda_{12}p_1(t) + p_3(t)\lambda_{31}.$$

Ліва частина є ні чим іншим, як похідною функції $p_1(t)$:

$$\frac{\partial p_1(t)}{\partial t} = -\lambda_{12}p_1(t) + p_3(t)\lambda_{31}.$$

Таким чином вивели диференціальне рівняння, якому задовільняє функція $p_1(t)$. Аналогічні диференціальні рівняння можна вивести і для інших ймовірностей станів $p_1(t), p_2(t), p_3(t)$.

Розглянемо другий стан A_2 . Знайдемо ймовірність того, що у момент $t + \Delta t$ система буде знаходитися у стані A_2 . Ця подія може відбутися двома шляхами:

- в момент t система вже була у стані A_2 , а за час Δt не вийшла з цього стану;

або

- в момент t система була у стані A_1 , а за час Δt перейшла з нього у стан A_2 ;

або

- в момент t система була у стані A_4 , а за час Δt перейшла з нього у стан A_2 .

Ймовірність першого варіанта обчислюється так: $p_2(t)$ перемножується з умовною ймовірністю того, що система за час Δt не перейде ні в A_3 , ні в A_4 . Так як події, які полягають у переході за час Δt з A_2 у A_3 та з A_2 у A_4 , є несумісні, то ймовірність того, що відбудеться один з цих переходів, дорівнює сумі їх ймовірностей, тобто $\lambda_{23}\Delta t + \lambda_{24}\Delta t$ (з точністю до безкінечно малих вищих порядків). Ймовірність того, що не відбудеться жодний з цих переходів, дорівнює $1 - (\lambda_{23}\Delta t + \lambda_{24}\Delta t)$. Звідси ймовірність першого варіанту:

$$p_2(t)(1 - (\lambda_{23}\Delta t + \lambda_{24}\Delta t)).$$

Додаючи сюди ймовірності другого та третього варіантів, отримаємо:

$$p_2(t + \Delta t) = p_2(t)(1 - (\lambda_{23}\Delta t + \lambda_{24}\Delta t)) + p_1(t)\lambda_{12}\Delta t + p_4(t)\lambda_{42}\Delta t.$$

Переносячи $p_2(t)$ у ліву частину, ділячи на Δt та переходячи до границі, отримаємо диференціальне рівняння для $p_2(t)$:

$$\frac{\partial p_2(t)}{\partial t} = -\lambda_{23}p_2(t) - \lambda_{24}p_2(t) + \lambda_{12}p_1(t) + \lambda_{42}p_4(t).$$

Міркуючи аналогічно для станів A_3 та A_4 , отримаємо в результаті систему диференціальних рівнянь. Відкинувши з них заради зручності аргумент t у функціях p_1, p_2, p_3, p_4 , перепишемо систему у вигляді:

$$\frac{\partial p_1}{\partial t} = -\lambda_{12}p_1 + p_3\lambda_{31},$$

$$\frac{\partial p_2}{\partial t} = -\lambda_{23}p_2 - \lambda_{24}p_2 + \lambda_{12}p_1 + \lambda_{42}p_4,$$

$$\frac{\partial p_3}{\partial t} = -\lambda_{31}p_3 - \lambda_{34}p_3 + \lambda_{23}p_2,$$

$$\frac{\partial p_4}{\partial t} = -\lambda_{42}p_4 + \lambda_{24}p_2 + \lambda_{34}p_3.$$

Ці рівняння для ймовірностей станів і є рівняннями Колмогорова.

Інтегрування цієї системи рівнянь дасть шукані ймовірності станів як функції часу. Початкові умови беруться в залежності від того, яким був початковий стан системи. Наприклад, якщо у початковий момент часу (при $t = 0$) система знаходилася у стані A_1 , то треба прийняти початкові умови: $t = 0, p_1 = 1, p_2 = p_3 = p_4 = 0$.

Зауважимо, що всі чотири рівняння можна було б і не писати, так як $p_1 + p_2 + p_3 + p_4 = 1$ для всіх t , і будь-яку з ймовірностей p_1, p_2, p_3, p_4 можна виразити через три інші. Наприклад,

$$p_4 = 1 - (p_1 + p_2 + p_3).$$

Тоді спеціального рівняння для p_4 можна не писати.

Звернемо увагу на структуру рівнянь Колмогорова. Всі вони побудовані за загальним визначеним правилом, яке можна сформулювати наступним чином.

У лівій частині кожного рівняння розташована похідна ймовірності стану, а права частина містить стільки членів, скільки дуг пов'язано з даним станом. Якщо дуга виходить з стану, то відповідний член має знак «мінус», а якщо дуга входить у стан, то – знак «плюс». Кожний член дорівнює добутку інтенсивності переходу, що відповідає даній дузі, та ймовірності того стану, з якого виходить дуга.

Запишемо систему рівнянь Колмогорова у матричному вигляді.

Матриця $\Lambda = \|\lambda_{ij}\|$ ($i, j = 1, 2, \dots, n$) називається *матрицею інтенсивностей* переходів марковського ланцюга з неперервним часом. Її властивості:

- 1) $\lambda_{ij} \geq 0$ для $i \neq j$;
- 2) $\lambda_{ij} \leq 0$ для $i = j$;
- 3) $\sum_j \lambda_{ij} = 0, i = 1, 2, \dots, n$.

Знаючи матрицю інтенсивностей переходів або граф станів, можна визначити вектор ймовірностей станів

$$\overline{p(t)} = (p_1(t), p_2(t), \dots, p_n(t))$$

через матричне рівняння $\overline{p(t)'} = \overline{p(t)} \cdot \Lambda$.

Граничні ймовірності станів

Нехай фізична система має скінченну кількість дискретних станів A_1, A_2, \dots, A_n , в якій відбувається марковський процес з неперервним часом. Припустимо, що всі інтенсивності потоків подій, що переводять систему із

стану у стан, сталі ($\lambda = const$), іншими словами, всі потоки подій – найпростіші (стаціонарні, без післядії, ординарні, тобто пуасонівські) потоки.

Записавши систему диференціальних рівнянь Колмогорова для ймовірностей станів та проінтегрувавши ці рівняння при заданих початкових умовах, отримаємо ймовірність станів, як функції часу:

$$p_1(t), p_2(t), \dots, p_n(t),$$

які при будь-якому t дають у сумі одиницю.

Поставимо наступне питання: що буде відбуватися з системою при $t \rightarrow \infty$? Чи будуть функції $p_1(t), p_2(t), \dots, p_n(t)$ прямувати до якихось границь? Такі границі, якщо вони існують, називаються *граничними ймовірностями станів*.

Система, для якої існують граничні ймовірності, називається *ергодичною*, відповідний процес – *ергодичний*.

Говорячи про граничні ймовірності, введемо ще кілька визначень.

Стан A_i називається *неістотним*, якщо знайдеться такий стан A_j , що з A_i в A_j можна перейти, а з A_j в A_i – не можна. Якщо A_i – неістотний стан, то $\lim_{t \rightarrow \infty} p_i(t) = 0$.

Два істотні стани A_j і A_i називаються зв'язними, якщо з A_i можна потрапити в A_j і навпаки.

Можна довести наступні загальні положення. *Якщо кількість станів системи скінченна і всі вони є зв'язними, то граничні ймовірності станів існують і не залежать від початкового стану системи.*

Щоб процес зі скінченим числом станів мав стаціонарний розподіл ймовірностей, необхідно і достатньо, щоб усі його істотні стани були зв'язними між собою.

Нехай сформульована умова виконується, а граничні ймовірності існують:

$$\lim_{t \rightarrow \infty} p_i(t) = p_i \quad i = 1, 2, \dots, n.$$

Граничні ймовірності будуть позначатися тими ж літерами p_1, p_2, \dots, p_n , що й самі ймовірності станів, маючи на увазі не змінні величини, а сталі числа. Для граничних ймовірностей також виконується умова:

$$\sum_{i=1}^n p_i = 1.$$

Таким чином при $t \rightarrow \infty$ у системі встановлюється деякий *граничний стаціонарний режим*: він полягає у тому, що система випадковим чином змінює свої стани, але ймовірності кожного з них вже не залежать від часу.

Який зміст цієї ймовірності? Вона представляє собою не що інше, як *середній відносний час перебування системи у даному стані*.

Для того, щоб обчислити граничні ймовірності, необхідно у системі рівнянь Колмогорова ліві частини покласти рівними нулю.

Запитання для самоперевірки

1. Дайте означення марковського процесу.
2. Опишіть марковський процес з дискретними станами і часом.
3. Опишіть марковський процес з дискретними станами і неперервним часом.
4. Як будуються рівняння Колмогорова?
5. У чому полягає стаціонарний режим марковського процесу?

10. Ймовірнісні процеси в теорії масового обслуговування

Основні означення

Заявкою називають попит на задоволення деякої потреби.

Системою масового обслуговування (СМО) називається будь-яка система, що задовольняє заявки, які надходять до неї у випадкові моменти часу.

Пристрій, що безпосередньо обслуговує заявку, називається *каналом обслуговування*.

Послідовність подій які полягають в надходженні заявок в СМО, називають *вхідним потоком*.

В залежності від поведінки заявки в СМО розглядають СМО з відмовами (заявка поступила в СМО в момент зайнятості усіх каналів, дістає відмову і залишає систему) і СМО з чергою (заявка при відсутності вільних каналів на момент її надходження в СМО стає у чергу на обслуговування).

Потік заявок називається найпростішим, якщо він задовольняє умовам:

- 1) відсутність післядії (заявки надходять в СМО незалежно одна від одної);
- 2) стаціонарність (можна говорити про середнє число заявок за одиницю часу);

3) ординарність (одночасне надходження до СМО двох або більшого числа заявок є малоімовірним).

Позначимо через T проміжок часу між надходженням двох послідовних заявок. Функція розподілу цієї випадкової величини для найпростішого потоку показникова: $F(t) = P(T < t) = 1 - e^{-\lambda t}, t > 0$, де λ – інтенсивність потоку заявок (визначає середнє число заявок, що надходять в систему за одиницю часу).

Таким чином, математичне сподівання величини $T: M(T) = \frac{1}{\lambda}$.

Щоб задати СМО необхідно визначити ймовірносні характеристики часу обслуговування однієї заявки. Призначимо цей час через $T_{\text{обсл}}$, $T_{\text{обсл}}$ – випадкова величина, в більшості практичних задач вважають закон розподілу цієї характеристики показниковим, тобто $F(t) = P(T_{\text{обсл}} < t) = 1 - e^{-\mu t}, t > 0$.

Параметр цього розподілу μ є величиною оберненою до середнього часу обслуговування: $\mu = \frac{1}{M(T_{\text{обсл}})}$, μ – інтенсивність потоку обслуговування.

Процес загибелі та розмноження

Марковський неперервний ланцюг називається «процесом загибелі та розмноження», якщо його граф станів має вигляд, що представлений на рисунку, тобто всі стани можна витягнути в один ланцюг, в якому кожний стан зв'язаний прямим та оберненим зв'язком з кожним із сусідніх станів, а крайні стани A_0 та A_n – тільки з сусідніми станами.

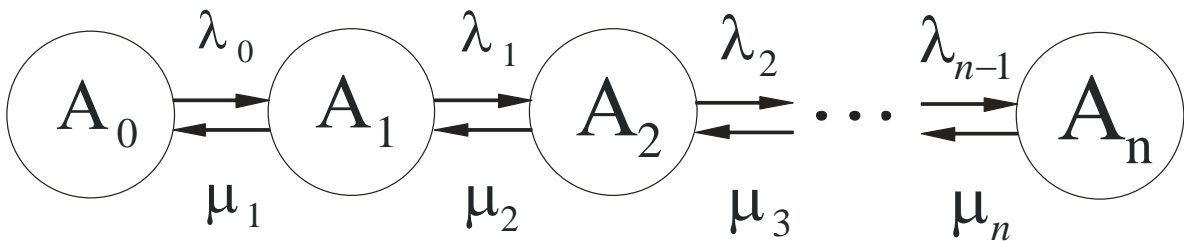


Схема загибелі та розмноження дуже часто зустрічається в самих різних практичних задачах, тому є сенс розглянути цю схему у загальному вигляді. У цій схемі $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$ – інтенсивності переходів системи зліва направо, $\mu_1, \mu_2, \dots, \mu_n$ – інтенсивності переходів системи справа на ліво.

Запишемо алгебраїчні рівняння для ймовірностей станів. Для стану A_0 маємо:

$$\lambda_0 p_0 = \mu_1 p_1.$$

Для стану A_1 :

$$\lambda_1 p_1 + \mu_1 p_1 = \lambda_0 p_0 + \mu_2 p_2.$$

Враховуючи попередню залежність, можна скоротити справа та зліва члени $\lambda_0 p_0 \mu_1 p_1$ та отримати:

$$\lambda_1 p_1 = \mu_2 p_2.$$

Аналогічні залежності для інших станів:

$$\lambda_2 p_2 = \mu_3 p_3$$

.....

$$\lambda_{n-1} p_{n-1} = \mu_n p_n.$$

Граничні ймовірності станів $p_0, p_1, p_2, \dots, p_n$ у будь-якій схемі загибелі та розмноження задовільняють системі рівнянь

$$\lambda_0 p_0 = \mu_1 p_1,$$

$$\lambda_1 p_1 = \mu_2 p_2,$$

$$\lambda_2 p_2 = \mu_3 p_3,$$

.....

$$\lambda_{n-1} p_{n-1} = \mu_n p_n.$$

Разом з нормуючою умовою $\sum_{i=0}^n p_i = 1$.

Будемо розв'язувати цю систему наступним чином. З першого рівняння виразимо p_1 :

$$p_1 = \frac{\lambda_0}{\mu_1} p_0,$$

З другого рівняння, з урахуванням p_1 , виразимо p_2 :

$$p_2 = \frac{\lambda_1}{\mu_2} p_1 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0,$$

з третього рівняння:

$$p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0,$$

і загалом

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0.$$

Підставляючи всі ймовірності $p_0, p_1, p_2, \dots, p_n$ у нормуючу умову, отримаємо

$$p_0 + \frac{\lambda_0}{\mu_1} p_0 + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0 + \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0 + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0 = 1,$$

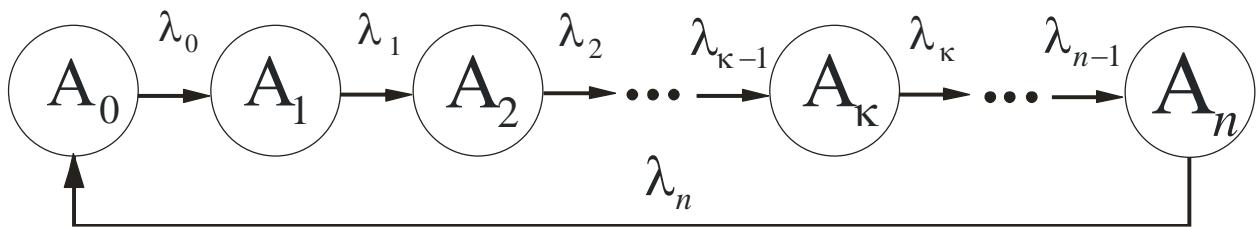
звідки
$$p_0 = \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}\right)^{-1}.$$

Інші ймовірності виражаються через p_0 :

$$p_1 = \frac{\lambda_0}{\mu_1} p_0, p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0, p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0, \dots, p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0.$$

Циклічний процес

Марковський випадковий процес називається циклічним, якщо стани системи, яку він описує, зв'язані між собою у кільце (цикл) з односторонніми переходами.



Запишемо алгебраїчні рівняння для граничних ймовірностей станів:

$$\lambda_0 p_0 = \lambda_1 p_1$$

$$\lambda_1 p_1 = \lambda_2 p_2$$

$$\lambda_2 p_2 = \lambda_3 p_3$$

.....

$$\lambda_{n-1} p_{n-1} = \lambda_n p_n$$

$$\lambda_n p_n = \lambda_0 p_0$$

разом з нормуючою умовою $\sum_{i=0}^n p_i = 1$. 3

отриманих рівнянь, відкинувши останнє, виразимо всі ймовірності p_1, p_2, \dots, p_n через p_0 та підставимо у нормуючу умову, звідки

$$p_0 = \left(1 + \lambda_0 \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_n}\right)\right)^{-1},$$

$$p_1 = \frac{\lambda_0}{\lambda_1} p_0,$$

$$p_2 = \frac{\lambda_0}{\lambda_2} p_0,$$

.....

$$p_n = \frac{\lambda_0}{\lambda_n} p_0.$$

Формули можна привести до більш зручного вигляду, якщо перейти від інтенсивності $\lambda_0, \lambda_1, \dots, \lambda_n$ до середнього часу t_i перебування системи у стані

A_i :

$$p_0 = \frac{t_0}{t_0+t_1+\dots+t_n},$$

$$p_1 = \frac{t_1}{t_0+t_1+\dots+t_n},$$

.....

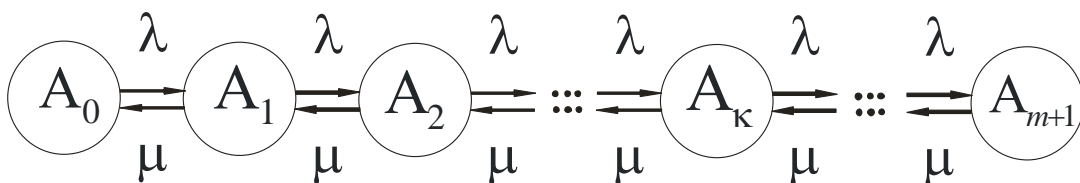
$$p_n = \frac{t_n}{t_0+t_1+\dots+t_n}.$$

Одноканальна СМО з обмеженою чергою

Розглянемо найпростішу з усіх можливих СМО з очікуванням – одноканальну систему ($n=1$), на яку поступає потік заявок з інтенсивністю λ , інтенсивність обслуговування заявок позначимо через μ . Заявка, що поступає у момент, коли канал зайнятий, стає у чергу та очікує обслуговування.

У цьому пункті розглядається система з обмеженою кількістю місць у черзі (m). Граф такої системи має наступні стани:

- A_0 – канал вільний;
- A_1 – канал зайнятий, черги немає;
- A_2 – канал зайнятий, одна заявка стоїть у черзі;
-
- A_k – канал зайнятий, $k-1$ заявки стоять у черзі;
-
- A_{m+1} – канал зайнятий, m заявок стоять у черзі.



Схема, що зображена на малюнку, є схемою загибелі та розмноження. Запишемо для неї граничні імовірності:

$$p_0 = (1 + \rho + \rho^2 + \dots + \rho^{m+1})^{-1},$$

$$p_1 = \rho p_0, p_2 = \rho^2 p_0, \dots, p_{m+1} = \rho^{m+1} p_0, \text{ де } \rho = \frac{\lambda}{\mu}.$$

Звернемо увагу, що вираз у дужках першої формули є геометричною прогресією з першим членом 1 та знаменником ρ , тоді: $p_0 = \frac{1-\rho}{1-\rho^{m+2}}$

$$p_1 = \rho p_0, p_2 = \rho^2 p_0, \dots, p_{m+1} = \rho^{m+1} p_0.$$

Ці формули є вірними при $\rho \neq 1$. Якщо $\rho = 1$, то сума геометричної прогресії дорівнює $m + 2$, $p_0 = \frac{1}{m+2}$.

Визначимо характеристики СМО:

- 1) ймовірність відмови $P_{\text{відм}} = p_{m+1} = \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}}$;
- 2) відносна пропускна спроможність $Q = 1 - P_{\text{відм}} = 1 - \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}}$;
- 3) абсолютна пропускна спроможність $A = \lambda Q$;
- 4) середня кількість заявок, що знаходяться у черзі

$$\bar{r} = 1 \cdot p_2 + 2 \cdot p_3 + \dots + (k-1) \cdot p_k + \dots + m \cdot p_{m+1} = \rho^2 p_0 (1 + 2\rho + \dots + (k-1)\rho^{k-2} + \dots + m\rho^{m-1}) = \frac{\rho^2(1-\rho^m(m+1-m\rho))}{(1-\rho^{m+2})(1-\rho)}$$
;
- 5) середня кількість заявок, що знаходяться під обслуговуванням

$$\bar{\omega} = 0 \cdot p_0 + 1 \cdot (1 - p_0)$$
;
- 6) середня кількість заявок, що знаходяться у СМО,

$$\bar{k} = \bar{r} + \bar{\omega}$$
;
- 7) середній час очікування заявки у черзі

$$\bar{t}_{\text{оч}} = \frac{1}{\rho\mu} \bar{r} = \frac{\bar{r}}{\lambda}$$
;
- 8) середній час перебування заявки у системі

$$\bar{t}_{\text{пер}} = \bar{t}_{\text{оч}} + \frac{Q}{\mu}$$
.

Багатоканальна СМО з відмовами

Для СМО з відмовами визначають наступні показники ефективності роботи:

A – абсолютна пропускна спроможність СМО (середня кількість заявок, яка обслуговує система за одиницею часу).

Q – відносна пропускна спроможність СМО (ймовірність обслуговування заявки, що надійшла)

$$Q = \frac{A}{\lambda}.$$

$P_{\text{відм}}$ – ймовірність відмови (ймовірність того, що заявка, яка надійшла, не буде обслугована), $P_{\text{відм}} = 1 - Q$.

\bar{k} – середнє число каналів, що зайняті.

Нехай СМО має k каналів, вхідний потік має інтенсивність λ , потік обслуговування одним каналом має інтенсивність μ . Пропонуємо стани СМО за кількістю зайнятих каналів:

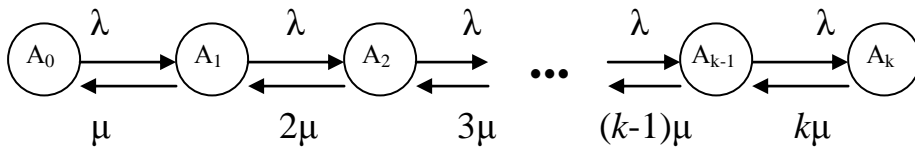
A_0 – всі канали вільні;

A_1 – один канал зайнятий;

A_2 – два канали є зайнятими;

...

A_k – всі канали зайняті.



Цей граф є графом загибелі і розмноження, для якого

$$\lambda_{i-1} = \lambda, \mu_i = i\mu \quad (i = 1, 2, \dots, k).$$

Граничний розподіл імовірностей станів можна обчислити, поклавши

$\rho = \frac{\lambda}{\mu}$ (ρ – коефіцієнт завантаження системи):

$$p_0 = \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^k}{k!}\right)^{-1};$$

$$p_1 = \frac{\rho}{1!} p_0, p_2 = \frac{\rho^2}{2!} p_0, \dots, p_k = \frac{\rho^k}{k!} p_0.$$

Отримані формули називаються формулами Ерланга. Спираючись на формули Ерланга, визначають показники ефективності СМО:

$$A = \lambda(1 - p_k); Q = \frac{A}{\lambda} = 1 - p_k; P_{\text{відм}} = p_k; \bar{k} = \frac{A}{\mu} = \rho(1 - p_k).$$

Багатоканальне СМО з необмеженою чергою

Нехай система має k каналів обслуговування. Всі потоки найпростіші, інтенсивність потоку заявок λ , потоку обслуговування однієї заявки – μ . Коефіцієнт завантаження СМО $\rho = \frac{\lambda}{\mu}$.

Позначимо відношення коефіцієнта завантаження до числа каналів системи черзі $\chi = \frac{\rho}{k}$. Граничний розподіл ймовірностей станів в СМО з необмеженою чергою існує тільки за умови:

$$\chi < 1.$$

Позначимо через p_i граничну ймовірність того, що в системі зайняті i каналів ($0 \leq i \leq k$), а через p_{k+r} – граничну ймовірність того, що в системі зайняті k каналів і r заявок стоять в черзі на обслуговування. При $\chi < 1$ граничний розподіл ймовірностей станів має вигляд:

$$p_0 = \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^k}{k!} + \frac{\rho^{k+1}}{k \cdot k!} \cdot \frac{1}{1 - \chi}\right)^{-1};$$

$$p_i = \frac{\rho^i}{i!} p_0 \quad (0 \leq i \leq k);$$

$$p_{k+r} = \frac{\rho^{k+r}}{k^r \cdot k!} p_0 \quad (r \geq 1).$$

Виходячи з того, що черга в СМО необмежена, кожна заявка згодом буде обслугована, тому очевидним є співвідношення:

$$P_{\text{відм}} = 0; \quad Q = 1; \quad A = \lambda Q = \lambda.$$

Крім розглянутих в попередній задачі для даної СМО обчислюються такі показники ефективності:

- 1) \bar{z} – середнє число заявок СМО (усі заявки, і такі, що обслуговуються, і такі, що стоять у черзі);
- 2) \bar{r} – середнє число заявок у черзі;
- 3) $\overline{t_{\text{сист}}}$ – середній час перебування заявки в системі (як у черзі, так і під обслуговування);
- 4) $\overline{t_{\text{сер}}}$ – середній час перебування заявки у черзі.

Вказані показники ефективності обчислюються за формулами:

$$\bar{r} = \frac{\rho^{k+1} \cdot p_0}{k \cdot k! (1 - \chi)^2}; \quad \bar{k} = \frac{A}{\mu} = \frac{\lambda}{\mu} = \rho; \quad \bar{z} = \bar{r} + \bar{k}; \quad \overline{t_{\text{сист}}} = \frac{\bar{z}}{\lambda}; \quad \overline{t_{\text{сер}}} = \frac{\bar{r}}{\lambda}.$$

Запитання для самоперевірки

1. Дайте означення системи масового обслуговування та її складових частин?
2. Як виглядає схема загибелі та розмноження?
3. Перерахуйте відомі вам СМО та їх характеристики.

11. Статистичний експеримент

Варіаційні ряди

Вихідним пунктом статистичних досліджень будь-якої випадкової величини X з функцією розподілу $F(x)$ є сукупність із n незалежних спостережень, в результаті яких випадкова величина X набуває значення x_1, x_2, \dots, x_n . Набір значень x_1, x_2, \dots, x_n називається *вибіркою* об'єму n , взятою з генеральної сукупності випадкової величини X з функцією розподілу $F(x)$.

Впорядкована за величиною послідовність вибірових значень $x_1 < x_2 < \dots < x_k$ називається *варіаційним рядом*. Різні значення x_i у вибірці називаються варіантами, а n_i – кількість появ x_i у вибірці – *частотою варіанта*. Очевидно, $\sum_{i=1}^k n_i = n$. Відносною частотою варіанта x_i називається відношення частоти n_i до об'єму вибірки n і позначається ω_i . Очевидно, $\sum_{i=1}^k \omega_i = 1$. Таблиця, що встановлює відповідність між варіантами та їх частотами або відносними частотами, називається *статистичним розподілом*:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

x_i	x_1	x_2	...	x_k
ω_i	ω_1	ω_2	...	ω_k

Якщо варіанти вибірки можуть відрізнятися один від одного на скільки завгодно малу величину, тобто випадкова величина набуває довільних значень у деякому інтервалі (генеральна сукупність неперервна), або коли вибірка містить досить велику кількість варіантів, то на практиці такі вибірки перегруповуються. Для цього всю ширину вибірки об'єму n розбивають інтервалами довжиною h і дані спостережень подаються таблицею частот, де вказані часткові інтервали $(a_i; a_{i+1})$ і n_i – кількість тих вибірових значень, які потрапили до i -го інтервалу розбиття:

I	$(a_1; a_2)$	$(a_2; a_3)$...	$(a_k; a_{k+1})$
n_i	n_1	n_2	...	n_k

Таблиця називається інтервальним варіаційним рядом. Для визначення оптимальної ширини інтервалу варіаційного ряду користуються формулою

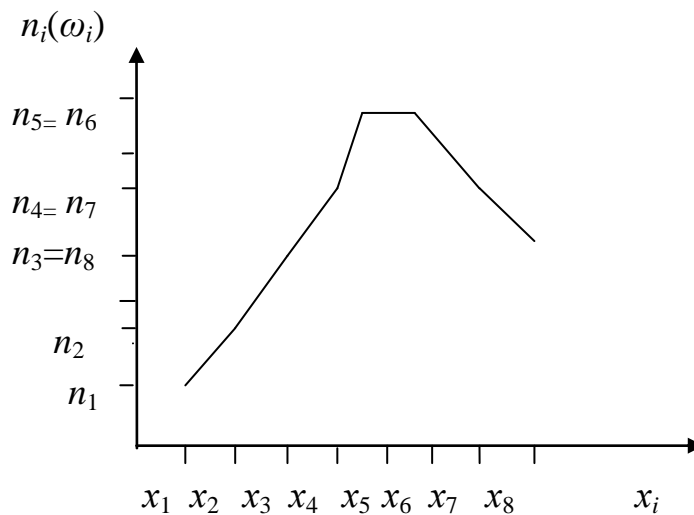
$$h = \frac{x_{\max} - x_{\min}}{1 + 3,2 * \lg n},$$

де x_{\max} , x_{\min} – максимальні та мінімальні варіанти. Якщо h – дробове число, то за h беруть найближче ціле число або найближчий простий дріб. За початок першого інтервалу беруть $a_1 = x_{\min} - h/2$, тоді $a_2 = a_1 + h$ і т.д. Процес продовжується доти, поки початок наступного інтервалу не буде більшим за x_{\max} . Число h називається *кроком* вибірки.

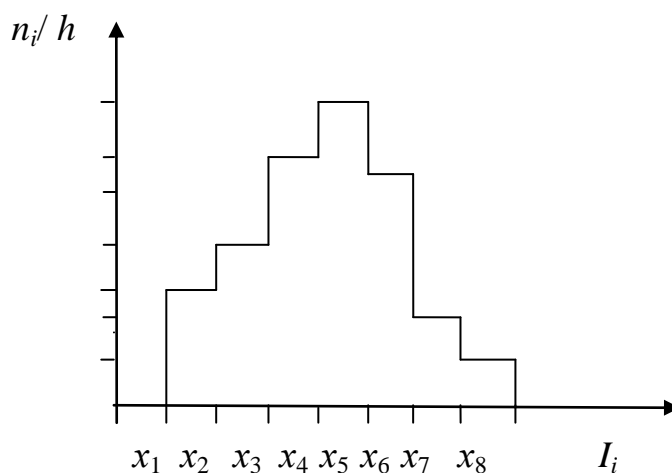
Статистичним розподілом називають таблицю, де вказують середини x_i інтервалів розбиття та їх відносні частоти.

Графічне зображення варіаційних рядів

Для ілюстрації варіаційних рядів використовують *полігон частот* ($x_i; n_i$) або *полігон відносних частот* ($x_i; \omega_i$):



Для ілюстрації інтервальних варіаційних рядів будують діаграми, які називають *гістограмами*.



Емпірична функція розподілу

Нехай x_1, x_2, \dots, x_n - вибірка об'єму n з генеральної сукупності випадкової величини X з функцією розподілу $F(x)$. Визначимо для кожного значення x функцію $\mu^*(x)$, що дорівнює кількості елементів вибірки, значення яких не перевищують x .

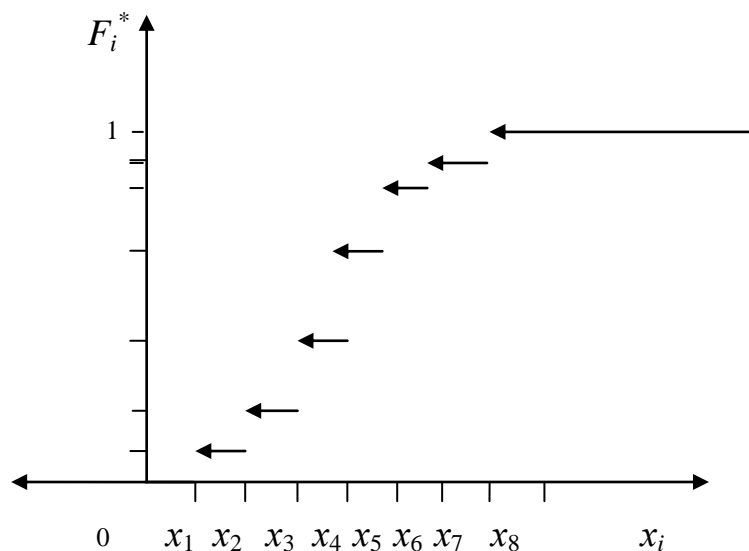
Емпіричною функцією розподілу вибірки x_1, x_2, \dots, x_n називається функція $F^*(x) = \mu^*(x)/n$. На відміну від емпіричної функції розподілу вибірки функцію розподілу $F(x)$ генеральної сукупності називають теоретичною функцією розподілу. Якщо вибірка має лише k ранжованих варіант, то

$$F^*(x) = \begin{cases} 0, & \text{при } x \leq x_1, \\ \sum_{j=1}^i \frac{n_j}{n}, & \text{при } x_i < x \leq x_{i+1}, i = 1, \dots, k-1, \\ 1, & \text{при } x > x_k \end{cases}$$

та має такі властивості:

1) $F^*(x) \in [0; 1]$; 2) $F^*(x)$ – неспадна, зростає в точках x_i стрибками $\frac{n_i}{n}$.

Графік емпіричної функції розподілу має ступінчастий характер:



Емпірична функція розподілу відіграє фундаментальну роль у математичній статистиці. Найважливіша його властивість полягає в тому, що при збільшенні кількості випробувань над випадковою величиною відбувається зближення цієї функції з теоретичною функцією розподілу.

Числові характеристики вибірки

Основні числові характеристики статистичного ряду:

- 1) вибіркоче середнє $\bar{x}_B = \frac{1}{n} \sum_{i=1}^k x_i n_i = \sum_{i=1}^k x_i \omega_i$;
- 2) вибіркова дисперсія $D_\epsilon = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x}_B)^2 n_i = \sum_{i=1}^k (x_i - \bar{x}_B)^2 \omega_i$ або $D_\epsilon = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - (\bar{x}_B)^2 = \sum_{i=1}^k x_i^2 \omega_i - (\bar{x}_B)^2$;
- 3) вибіркоче середнє квадратичне відхилення $\sigma_\epsilon = \sqrt{D_B}$;
- 4) вибіркоче виправлена дисперсія $S^2 = \frac{n}{n-1} \cdot D_\epsilon$ та виправлене середнє квадратичне відхилення $S = \sqrt{\frac{n}{n-1} \cdot D_B}$.

Основні числові характеристики інтервального статистичного ряду:

- 1) вибіркоче середнє $\bar{x}_B = \frac{1}{n} \sum_{i=1}^k \bar{x}_i n_i = \sum_{i=1}^k \bar{x}_i \omega_i$, де $\bar{x}_i = \frac{x_{i+1} + x_i}{2}$;
- 2) вибіркова медіана M_e^* - це точка частинного інтервалу (медіанного інтервалу) такого, що $F^*(x_m) < 0,5$, а $F^*(x_{m+1}) \geq 0,5$,
$$M_e^* = x_m + \frac{0,5 - F^*(x_m)}{F^*(x_{m+1}) - F^*(x_m)} \cdot h$$
;
- 3) вибіркова мода M_o^* - це точка частинного інтервалу (модального інтервалу), якому відповідає найбільше значення частоти n_m ,
$$M_o^* = x_m + \frac{n_m - n_{m-1}}{2n_m - n_{m-1} - n_{m+1}} \cdot h$$
;
- 4) вибіркова дисперсія $D_\epsilon = \frac{1}{n} \sum_{i=1}^k (\bar{x}_i - \bar{x}_B)^2 n_i = \sum_{i=1}^k (\bar{x}_i - \bar{x}_B)^2 \omega_i$ або $D_\epsilon = \frac{1}{n} \sum_{i=1}^k \bar{x}_i^2 n_i - (\bar{x}_B)^2 = \sum_{i=1}^k \bar{x}_i^2 \omega_i - (\bar{x}_B)^2$;
- 5) вибіркоче середнє квадратичне відхилення $\sigma_\epsilon = \sqrt{D_B}$;
- 6) вибіркоче виправлена дисперсія $S^2 = \frac{n}{n-1} \cdot D_\epsilon$ та виправлене середнє квадратичне відхилення $S = \sqrt{\frac{n}{n-1} \cdot D_B}$.

Запитання для самоперевірки

1. Як виглядає варіаційний ряд? Як він геометрично зображується?
2. Які числові характеристики описують варіаційний ряд?

12. Статистичні оцінки параметрів розподілу

Незсунені, ефективні та спроможні оцінки

У багатьох задачах вигляд теоретичного розподілу генеральної сукупності може вважатися відомим. Природно, виникає задача оцінки параметрів, якими характеризується розподіл. Наприклад, μ, σ - для нормального закону розподілу, λ - для розподілу Пуассона.

Нехай x_1, x_2, \dots, x_n - вибірка з генеральної сукупності X з функцією розподілу $F(x, \theta)$, функціональна форма якого відома, але залежить від невідомого параметру θ . Задача оцінювання полягає у тому, щоб, використовуючи статистичну інформацію, зробити статистичні висновки про реальне значення параметру θ .

Нехай θ^* - статистична оцінка невідомого параметру теоретичного розподілу $F(x, \theta)$. Припустимо, що за вибіркою знайдена оцінка θ_1^* . При повторному експерименті з тим же об'ємом вибірки отримана нова оцінка θ_2^* . Повторюючи дослід багаторазово, отримаємо числа $\theta_1^*, \theta_2^*, \dots, \theta_k^*$.

Якщо оцінка θ^* дає наближене значення θ з надлишком, то $M\theta^* > \theta$, якщо з нестачею, то $M\theta^* < \theta$. Природно вимагати, щоб $M\theta^* = \theta$.

Оцінка θ^* називається *незсуненою* невідомого параметру θ , якщо $M\theta^* = \theta$. В протилежному випадку оцінка θ^* називається *зсуненою*.

Незсунену оцінку $\hat{\theta}$ називається *ефективною*, якщо вона має найменшу можливу дисперсію серед усіх інших незсунених оцінок параметру θ .

Оцінка θ^* називається *спроможною*, якщо для довільного $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| > \varepsilon) = 0.$$

Отже, "точність" оцінки збільшується при збільшенні об'єму вибірки.

Точкові оцінки

Точкова оцінка θ^* невідомого параметру θ - це певне число, яке використовується замість θ .

Теорема. Вибіркове середнє $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ вибірки X_1, X_2, \dots, X_n є незсуненою і спроможною оцінкою математичного сподівання $a = MX$.

Доведення. 1) Оскільки випадкові величини X_1, X_2, \dots, X_n розподілені однаково та $MX_i = a$, то $MX = \frac{1}{n} M \sum_{i=1}^n X_i = \frac{1}{n} \cdot na = a$, тобто незсунена.

2) За законом великих чисел $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - a\right| > \varepsilon\right) = 0$, що доводить спроможність.

Теорема. Вибіркова дисперсія $D_s = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ є зсуненою і спроможною оцінкою дисперсії $DX = \sigma^2$. Величина $S^2 = \frac{n}{n-1} \cdot D_s$ буде незсуненою та спроможною оцінкою σ^2 .

Доведення. Введемо випадкові величини $Y_i = X_i - a$. Тоді $\bar{Y} = \bar{X} - a$, $MY_i = 0, DY_i = MY_i^2 = \sigma^2$, $MD_s = \frac{1}{n} M \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} M \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n} M \sum_{i=1}^n Y_i^2 - \frac{2}{n} M \sum_{i=1}^n (Y_i \bar{Y}) + M\bar{Y}^2 = \frac{1}{n} \cdot n\sigma^2 - 2M\bar{Y}^2 + M\bar{Y}^2 = \sigma^2 - M\bar{Y}^2 = \sigma^2 - M\left(\frac{1}{n} \sum_{i=1}^n Y_i\right)^2 = \sigma^2 - \frac{1}{n^2} M\left(\sum_{i=1}^n Y_i^2 - \sum_{i \neq j} Y_i Y_j\right) = \sigma^2 - \frac{1}{n^2} (n\sigma^2 - 0) = \sigma^2 \cdot \frac{n-1}{n}$.

$$M S^2 = \frac{n}{n-1} \cdot \sigma^2 \cdot \frac{n-1}{n} = \sigma^2.$$

Спроможність впливає з закона великих чисел.

Інтервальне оцінювання

При точковому оцінюванні визначається єдине значення оцінки, яке береться за наближене значення параметра θ , що оцінюється.

Довірчим (надійним) інтервалом для невідомого параметра θ розподілу ознаки генеральної сукупності називається інтервал $(\theta^* - \delta; \theta^* + \delta)$, $\delta > 0$, який містить в собі невідомий параметр θ з ймовірністю γ , тобто $\gamma = P\{|\theta^* - \theta| < \delta\}$, γ називається надійністю, число $\alpha = 1 - \gamma$ – рівнем значущості, а δ – точністю оцінки. Зазвичай $\gamma = 0,9; 0,95; 0,99$.

Якщо ознака генеральної сукупності має нормальний закон розподілу $N(a; \sigma)$ та відоме σ , то $a \in (\bar{x}_B - c_\gamma \frac{\sigma}{\sqrt{n}}; \bar{x}_B + c_\gamma \frac{\sigma}{\sqrt{n}})$, де $\Phi(c_\gamma) = \gamma/2$, $\Phi(x)$ – функція Лапласа (додаток).

Якщо ознака генеральної сукупності має нормальний закон розподілу $N(a; \sigma)$ та невідоме σ , то $a \in (\bar{x}_B - t_\gamma \frac{S}{\sqrt{n}}; \bar{x}_B + t_\gamma \frac{S}{\sqrt{n}})$, де t_γ – коефіцієнт Стьюдента (додаток).

Якщо ознака генеральної сукупності має нормальний закон розподілу $N(a; \sigma)$, то $\sigma \in (S(1-q); S(1+q))$, де q – коефіцієнт χ -розподілу (додаток).

Запитання для самоперевірки

1. Що розуміємо під поняттями незсуненої, ефективної та спроможної оцінок?
2. Які точкові оцінки є незсуненими та спроможними?
3. У чому полягає інтервальне оцінювання?
4. Запишіть інтервальну оцінку вибіркового середнього і вибіркової дисперсії для нормального закону розподілу.

13. Перевірка статистичних гіпотез

Критерій перевірки гіпотези H_0 – це правило, за яким приймається рішення залишити чи відхилити H_0 . Статистика критерію – це випадкова величина K , що використовується для перевірки гіпотези H_0 . Емпіричним значенням статистики K_e називається значення статистики K , що обчислюється за вибіркою.

Множина значень статистики K розбивається критичними точками на область прийняття гіпотези H_0 та критичну область критерію (множину значень K , при яких H_0 відхиляється).

Алгоритм застосування критерію згоди Пірсона.

1) Обчислити теоретичні частоти n_i' гіпотетичного закону розподілу ознаки генеральної сукупності: $n_i' = n(\Phi_0(z_{i+1}) - \Phi_0(z_i))$, де $z_{i+1} = (x_{i+1} - \bar{x}_B) / \sigma_B$; $z_i = (x_i - \bar{x}_B) / \sigma_B$.

2) Обчислити емпіричне значення статистики критерію Пірсона

$$\chi_B^2 = \sum_{i=1}^k (n_i - n_i')^2 / n_i'$$

3) За таблицею критичних точок розподілу χ^2 при заданому рівні значущості α та $s = k - 1 - r$ кількості ступенів свободи (r – кількість параметрів гіпотетичного розподілу, які оцінюються за вибіркою) знайти критичну точку $\chi_{кр}^2(\alpha; s)$.

4) Якщо $\chi_B^2 < \chi_{кр}^2(\alpha; s)$, то гіпотеза H_0 не відхиляється (спостереження узгоджуються з гіпотетичним розподілом); якщо $\chi_B^2 > \chi_{кр}^2(\alpha; s)$, то гіпотеза H_0 відхиляється.

Алгоритм застосування критерію згоди Колмогорова.

1) Побудувати емпіричну функцію розподілу $F^*(x)$ за вибіркою об'єму n та знайти $F^*(\bar{x}_i)$.

2) Обчислити значення гіпотетичної функції розподілу

$$F(\bar{x}_i) = 0,5 + \Phi_0\left(\frac{\bar{x}_i - \bar{x}_B}{\sigma_B}\right).$$

3) Обчислити емпіричне значення статистики Колмогорова

$$D^* = \max |F^*(\bar{x}_i) - F(\bar{x}_i)|.$$

4) За таблицею критичних точок розподілу Колмогорова

1- α	0,90	0,95	0,99
$D_{кр} \sqrt{n}$	1,224	1,358	1,628

при заданому рівні значущості α знати критичну точку $D_{кр}(\alpha; n)$.

5) Якщо $D^* < D_{кр}(\alpha; n)$, то гіпотеза H_0 не відхиляється; якщо $D^* > D_{кр}(\alpha; n)$, то гіпотеза H_0 відхиляється.

Запитання для самоперевірки

1. У чому полягає зміст критерію згоди Пірсона?
2. У чому полягає зміст критерію згоди Колмогорова?

14. Лінійна кореляція

Якщо з генеральної сукупності (X, Y) добута вибірка об'єму n , то вибіркові рівняння прямих ліній регресії Y на X та X на Y записуються наступним чином $\bar{y}_x = \bar{y} + r^* \frac{\sigma_y^*}{\sigma_x^*} (x - \bar{x})$, $\bar{x}_y = \bar{x} + r^* \frac{\sigma_x^*}{\sigma_y^*} (y - \bar{y})$,

де \bar{x} , \bar{y} , σ_x^* , σ_y^* - вибіркові середні та вибіркові середні квадратичні відхилення; r^* - вибірковий коефіцієнт кореляції, які обчислюються за

формулами $\bar{x} = \frac{\sum n_x x}{n}$, $\bar{y} = \frac{\sum n_y y}{n}$, $\sigma_x^* = \sqrt{\frac{\sum n_x x^2}{n} - \bar{x}^2}$,

$$\sigma_y^* = \sqrt{\frac{\sum n_y y^2}{n} - \bar{y}^2}, \quad r^* = \frac{\sum n_{xy} xy - n \bar{x} \bar{y}}{n \sigma_x^* \sigma_y^*}, \quad x, y, (x, y) - \text{варіанти величин } X,$$

$Y, (X, Y)$; n_x, n_y, n_{xy} - відповідно частоти варіант, для яких виконуються рівності $\sum n_x = \sum n_y = \sum n_{xy} = n$.

Рівняння прямих ліній регресії відображають залежність умовних математичних сподівань від заданих значень x та y . Для коефіцієнта кореляції має місце нерівність $|r| \leq 1$. Якщо величини X та Y незалежні, то $r = 0$. Якщо $r = 0$, то величини X та Y називаються некорельованими, але вони можуть бути залежними у випадках, коли їх розподіл

відрізняється від нормального. При лінійній функціональній залежності X та Y $r = 1$ або $r = -1$. Тому вважають, що цей коефіцієнт при $|r| < 1$ характеризує ступінь тісноти лінійної ймовірнісної залежності між випадковими величинами X та Y .

Надалі розглядатимуться вибірки з рівновіддаленими варіантами, а двомірний статистичний розподіл вибірки задається у вигляді кореляційної таблиці:

$Y \backslash X$					
	...	x	$x + h_1$...	n_y
...					
y		n_{xy}			n_y
$y + h_2$					
...					
n_x		n_x			n

У першому рядку та у першому стовпчику таблиці відповідно містяться варіаційні ряди величин X та Y , тобто зростаючі послідовності рівновіддалених варіант x та y . Останній рядок та стовпчик вказують відповідно частоти n_x, n_y варіант x та y . У правому нижньому куті вказано об'єм вибірки $n = \sum n_x = \sum n_y = \sum n_{xy}$. Решта клітинок таблиці містить частоти n_{xy} варіант (x, y) .

Щоб уникнути обчислень з великими числами у варіантах, доцільно скористатися умовними варіантами $u = \frac{x-C_1}{h_1}$, $v = \frac{y-C_2}{h_2}$, де C_1, C_2 – “фіктивні нулі” (або нові початки відліку) відповідно варіант x та y ; h_1, h_2 – кроки варіант x та y . Надалі будемо приймати значення C_1, C_2 , що відповідають найбільшому значенню частоти n_{xy} або знаходяться в околі цієї точки. Формули для $\bar{u}, \bar{v}, \sigma_u^*, \sigma_v^*, r^*$ набувають вигляду

$$\bar{u} = \frac{\sum n_u u}{n}, \quad \bar{v} = \frac{\sum n_v v}{n}, \quad \sigma_u^* = \sqrt{\frac{\sum n_u u^2}{n} - \bar{u}^2}, \quad \sigma_v^* = \sqrt{\frac{\sum n_v v^2}{n} - \bar{v}^2},$$

$$r^* = \frac{\sum n_{uv} uv - n \bar{u} \bar{v}}{n \sigma_u^* \sigma_v^*}.$$

Обчисливши ці величини, можна визначити величини, що входять до рівнянь регресії за формулами:

$$\bar{x} = h_1 \bar{u} + C_1, \quad \bar{y} = h_2 \bar{v} + C_2, \quad \sigma_x^* = h_1 \sigma_u^*, \quad \sigma_y^* = h_2 \sigma_v^*.$$

Оскільки r^* - вибірковий коефіцієнт кореляції, то виникає задача оцінки наскільки він відображає корельованість величин X та Y генеральної сукупності (X, Y) , тобто задача перевірки нульової гіпотези H_0 про рівність нулю коефіцієнта кореляції.

Ця задача розв'язується за наступним правилом. Для того, щоб для заданого рівня значущості α перевірити гіпотезу H_0 при конкуруючій гіпотезі $H_1: r \neq 0$, необхідно обчислити значення критерія $T = \frac{r^* \sqrt{n-2}}{\sqrt{1-r^{*2}}}$.

По значенню α та кількості ступенів свободи $k=n-2$ знайти в таблиці критичних точок розподілу Стюдента критичну точку $t_{кр}(\alpha, k)$ двобічної критичної області. Якщо $|T| > t_{кр}$, то гіпотезу H_0 відкидають; якщо $|T| < t_{кр}$, то гіпотезу H_0 приймають.

У випадку прийняття гіпотези H_1 (тобто кореляція між X та Y вважається суттєвою) наближений довірчий інтервал для коефіцієнта кореляції r має вигляд

$$r^* - u_\alpha \frac{1-r^{*2}}{\sqrt{n}} \leq r \leq r^* + u_\alpha \frac{1-r^{*2}}{\sqrt{n}},$$

де u_α - критична границя для нормального розподілу, що відповідає рівню значущості α ($\frac{1-\alpha}{2} = \Phi_0(u_\alpha)$, $\Phi_0(x)$ - функція Лапласа). При $n \geq 50$ можна прийняти $u_\alpha = 3$.

Запитання для самоперевірки

1. На яке основне питання відповідає кореляційний аналіз вибірок випадкових величин?
2. Про що говорять числові значення коефіцієнта кореляції?

Лабораторна робота №1

Тема заняття: Класичне визначення ймовірності. Обчислення ймовірності за допомогою формул комбінаторики. Теореми додавання та множення ймовірностей.

Мета заняття: Вивчити формули комбінаторики. Ознайомитися з основними поняттями середовища Excel (формула, функція, абсолютне та відносне посилання). Навчитися розв'язувати задачі, використовуючи класичне визначення ймовірності, теореми додавання та множення ймовірностей, а також програмувати розв'язання на робочому аркуші Excel.

Теоретичні відомості: *електронна таблиця* – це програма, яку використовують для виконання будь-яких обчислень (математичних, статистичних, фінансових та ін.), а також для накопичення інформації та її оброблення.

Вікно Excel вміщує багато стандартних елементів Windows'95, такі, як рядок меню, рядок стану та панелі інструментів. Є також елементи, що використовуються тільки у електронних таблицях:

- *рядок формул* – коли ви вводите інформацію у комірку, вона одночасно з'являється у рядку формул, де висвічуються і координати (адреса) комірки;

- *вікно робочої книги* – кожен файл Excel являє собою робочу книгу з аркушами. Одночасно можна відкрити кілька книг – кожен в своєму вікні;

- *заголовки стовпців* – стовпці позначаються літерами тільки латинського алфавіту, розміщеними у верхній їх частині;

- *заголовки рядків* – рядки відрізняються номерами, розміщеними у їх лівій частині;

- *комірка* – комірка створюється на перетині стовпця з рядком. Ім'я комірки складається з літери стовпця (обов'язково латинської) та номера рядка.

Існує кілька *типів даних* для роботи з електронними таблицями: текстові; дати; формули; числові; часу; функції. Текстові дані є послідовністю літер, цифр та пропусків між ними. За замовчуванням текст, що вводиться у комірку, вирівнюється по лівому краю.

Розглянемо *основні формати чисел*. Серед них слід виділити: "Общий" (10,6) – значення заноситься до комірки у тому ж вигляді, у якому його вводять;

“Числовой” (240,00) – за замовчування числа у цьому форматі мають два десяткових розряди, від’ємні числа позначаються зі знаком “–” червоного кольору та в круглих дужках;

“Денежный” (3,67.50 грн.) – за замовчуванням мають два десяткових знаки і назву грошової одиниці, від’ємні значення позначаються червоним кольором і дужками;

“Финансовый” (454.4 грн.) – цей формат використовується для вирівнювання чисел за знаком грошової одиниці та десятковою крапкою;

“Дата” (11/2) – за замовчуванням значення в цьому форматі записується у вигляді числа і дня місяця, що розділяються знаком “/”;

“Час” (10:12) – за замовчуванням це кількість годин і хвилин, що розділяються двома крапками.

Для того, щоб встановити потрібний формат, необхідна така послідовність дій:

- 1 Виділити комірку, формат якої потрібно змінити.
- 2 Відкрити меню Формат і вибрати у ньому пункт Ячейки.
- 3 Клацнути по вкладинці Число.
- 4 У списку Числовые форматы вибрати потрібний. У вікні Образец з’явиться приклад даного числа.
- 5 При потребі змінити тип формату.
- 6 Клацнути по ОК чи натиснути Enter.

Для того, щоб виконувати розрахунки за даними, які знаходяться у комірках, використовуються *формули*. За допомогою формул додають, віднімають, множать та ділять числа. Формула складається із одного чи кількох адрес комірок, чисел і математичних знаків: “+”, “–”, “*”, “/”, “^”. Кожна формула повинна починатися зі знаку “=”.

Полегшити введення функції у комірку можна, користуючись *Майстром функцій*. Для його застосування необхідно:

- 1 Виділіть комірку, у яку потрібно помістити результат функції.
- 2 Введіть знак рівності, а потім клацніть по кнопці Изменить формулу “=”, що розміщена у рядку формул. З’явиться панель формул.
- 3 Клацніть по кнопці, яка розкриває список функцій. Виберіть потрібну функцію. Якщо її немає, то виберіть пункт Другие функции.
- 4 Введіть аргумент функції. Якщо потрібно в аргумент ввести адресу комірки, клацніть по кнопці згорання панелі формул.

5 Введіть адреси комірок, а потім клацніть по кнопці згортання панелі формул ОК.

Завдання №1: Запрограмувати формули комбінаторики P_n , C_n^m та A_n^m за допомогою функції =ФАКТР(n), яка обчислює факторіал числа. (рекомендація: числа n и m розмістити у різних клітинках)

Завдання №2: За допомогою функцій =ПЕРЕСТ(n,m) та =ЧИСЛКОМБ(n,m), обчислити значення виразу: $\frac{A_{15}^5 + C_{21}^4}{P_6} - 36.9 \cdot \frac{C_{13}^4}{A_{20}^3 - P_6}$

Завдання №3: Розв'язати задачу, використовуючи класичне визначення ймовірності: Набираючи номер телефону, абонент забув останні три цифри, і, пам'яючи, що ці цифри різні, набрав їх навмання. Знайти ймовірність того, що набрані цифри вірні.

Завдання №4: Розв'язати задачу використовуючи теорему додавання та множення ймовірностей: Три студенти складають іспит. Ймовірність того, що перший студент здасть іспит, дорівнює 0,8; другий - 0,9; третій - 0,7. Знайти ймовірність того, що іспит складуть: а) всі три студенти; б) тільки один студент; в) тільки перший студент.

Розв'язання завдання №1:

Для початку лабораторної роботи треба запустити редактор електронних таблиць Excel. Для цього з меню кнопки **Пуск** треба вибрати пункт **Програми** \Rightarrow **MS Office** \Rightarrow **Microsoft Excel**. Після виконання цієї команди на екрані монітора з'явиться нова робоча книга з трьома чистими аркушами (таблицями).

У клітинку **A1** робочого аркуша вводимо текст: «Лабораторна робота №1». У клітинку **A2** вводимо текст: «Формули комбінаторики».

Далі нам потрібно запрограмувати такі формули:

Перестановки	Сполучення	Розміщення
$P_n = n!$	$C_n^m = \frac{n!}{m!(n-m)!}$	$A_n^m = \frac{n!}{(n-m)!}$

Для визначення кількості варіантів перестановок з, наприклад, 4-х елементів у клітинку **A4** введемо текст: «Перестановки», а у клітинку **A5** текст: « $P_4 = 4! =$ ». Далі приступимо до програмування нашої формули. У

клітинку, що знаходиться на перетині стовбчика **B** та рядка **5** (тобто клітинка **B5**) необхідно ввести формулу, що буде обчислювати факторіал числа 4, тобто $4! = 1 \cdot 2 \cdot 3 \cdot 4$.

У середовищі Excel є різні типи даних, які можна вводити до клітинок робочого аркушу. Наприклад, коли вводили до клітинок назви роботи та формул – це текстовий тип даних; якщо у клітинку ввести будь-яке число – говорять про числовий тип даних. Для того, щоб Excel зрозумів, введені у клітинку, дані як формулу, та виконав відповідні розрахунки, перед вводом виразу формули треба ввести знак рівності «=».

Отже, для обчислення $4!$ у клітинку **B5** вводимо таку формулу: $=1*2*3*4$ та натискаємо клавішу **[Enter]**. У клітинці **B5** повинен з'явитися результат обчислень, тобто число 24. Наші перші результати мають наступний вигляд:

	A	B	C	D
1	Лабораторна робота №1			
2	Формули комбінаторики			
3				
4	Перестановки			
5	$P_4=4!=$	24		

От ми і запрограмували нашу першу формулу! Але вона не дуже зручна для використання. Припустимо, що тепер нам потрібно обчислити кількість перестановок з 10-ти елементів. Що робити? Треба наново вводити в клітинку **B5** формулу, але тепер вона буде у 2 рази довше: $=1*2*3*4*5*6*7*8*9*10$. Це дуже не зручно.

Для того щоб зробити обчислення кількості перестановок з будь-якого числа елементів зручним використаємо вбудовану функцію **=ФАКТР(*n*)**. Функція **=ФАКТР(*n*)** обчислює факторіал числа, яке стоїть в аргументі функції.

Для наочності у клітинку **A6** введемо текст: «*n* =»; у клітинку **B6** введемо кількість елементів комбінації, наприклад 8; у клітинку **C6** введемо текст: « $P_n =$ ». Далі поставимо курсор миші у клітинку **D6** та виберемо з головного меню Вставка пункт **Вставка функції**. Після виконання цієї команди на екрані монітора з'явиться діалогове вікно, де потрібно знайти й обрати функцію ФАКТР і натиснути клавішу **ОК**. Далі, у наступному діалоговому вікні у поле **Число** потрібно ввести адресу клітинки, де знаходиться кількість елементів комбінації. У нашому випадку – це клітинка **B6**. Після натиснення клавіші **Enter** у клітинці **B6** повинен з'явитися результат обчислень, тобто число 40320. Отже тепер, якщо

нам буде потрібно обчислити кількість перестановок з 6-ти елементів нам не треба наново переписувати формулу, а треба тільки змінити число у клітинку **B6**. Спробуйте!

Далі нам потрібно запрограмувати формулу для обчислення кількості сполучень з n елементів комбінації по m елементів. У клітинку A8 введемо назву формули: «**Сполучення**». Передбачимо окреме введення індексів n та m у клітинках **B9** та **B10** відповідно. Для обчислення кількості сполучень з n елементів по m у клітинку **D10** потрібно ввести таку формулу:

$$=\text{ФАКТР}(B10)/(\text{ФАКТР}(B9)*\text{ФАКТР}(B10-B9))$$

Аналогічно для формули Сполучень запрограмуємо формулу для обчислення кількості розміщень з n елементів комбінації по m елементів. Після введення назви формули та індексів n і m у клітинку **D14** потрібно ввести формулу:

$$=\text{ФАКТР}(B14)/\text{ФАКТР}(B14-B13)$$

Після виконання деякого форматування робочого аркуша наша лабораторна робота має такий вигляд:

	A	B	C	D
1	Лабораторна робота №1			
2	Формули комбінаторики			
3				
4	Перестановки			
5	P4=4!=	24		
6	n=	8	Pn=	40320
7				
8	Сполучення			
9	m=	5		
10	n=	10	C=	252
11				
12	Розміщення			
13	m=	3		
14	n=	10	A=	720

Розв'язання завдання №2:

Крім функції ФАКТР(n) у середовищі Excel є функції =ПЕРЕСТ(n,m) і =ЧИСЛКОМБ(n,m), за допомогою яких можна обчислити кількість перестановок, сполучень і розміщень. Для обчислення даного виразу розіб'ємо його на два доданки **D1** і **D2**, які обчислимо в клітинках **B17** і **B18** відповідно. Отже для обчислення доданка **D1** вводимо формулу:

$$=(\text{ПЕРЕСТ}(15,5)+\text{ЧИСЛКОМБ}(21,4))/\text{ПЕРЕСТ}(6,6).$$

Для доданка **D2** формула має вигляд:

$$=36.9 * \text{ЧИСЛКОМБ}(13,4) / (\text{ПЕРЕСТ}(20,3) - \text{ПЕРЕСТ}(6,6))$$

Таким чином даний вираз дорівнює **=B17- B18**.

Результати обчислень на аркуші Excel мають вигляд:

16	Завдання №2: Обчислення значення виразу			
17	D1=	508.81		
18	D2=	4.311		
19	D=	504.5		

Розв'язання завдання №3:

Для початку опишемо подію, ймовірність якої потрібно знайти. В клітинку **A30** вводимо текст: «подія А – набрано три потрібні цифри». Тепер потрібно визначити **m** – кількість наслідків, що сприяють появі події А та **n** – загальну кількість всіх наслідків випробування. Визначимо **n** і **m** в окремих клітинках **B31** і **B32**. Отже усього можна набрати стільки різних цифр, скільки може бути складено розміщень з десяти цифр по три. Тобто для визначення загальної кількості елементарних подій $n = A_{10}^3$ у клітинку **B32** вводимо формулу **=ПЕРЕСТ(10,3)**. Сприяє події А тільки один результат – три вірні цифри, отже **m = 1**. Для визначення ймовірності події А у клітинку **A33** вводимо текст «P(A)=», а в клітинку **B33** вводимо формулу: **=B31/B32**, натискаємо клавішу **Enter**. У клітинці **B33** повинен з'явитися результат обчислень 0,0014.

Оформлення лабораторної роботи повинен мати такий вигляд:

21				
22	Класичне визначення ймовірності			
23	P(A)=	m	- кількість наслідків, що сприяють появі події А	
24		n	- загальна кількість всіх наслідків випробування	
25				
26	Основна властивість ймовірності			
27	$0 \leq P(A) \leq 1$			
28				
29	Завдання №3: Розв'язати задачу			
30	подія А - набрано три потрібні цифри			
31	m=	1		
32	n=	720		
33	P(A)=	0.0014		

Розв'язання завдання №4:

У діапазоні клітинок **A40:A48** описуємо події A, B і C. У діапазоні клітинок **A50:D53** записуємо ймовірності подій A, B і C та обчислюємо ймовірності протилежних подій за формулою: $P(\bar{A}) = 1 - P(A)$.

Використовуючи теореми додавання та множення ймовірностей у діапазоні клітинок **A55:D62** визначаємо ймовірності подій D, E і F.

Варіант оформлення звіту про виконання завдання наведено нижче.

40	Завдання №4: Розв'язати задачу		
41	подія A - перший студент здав іспит		
42	подія B - другий студент здав іспит		
43	подія C - третій студент здав іспит		
44			
45	протилежні події		
46	подія A - перший студент НЕ здав іспит		
47	подія B - другий студент НЕ здав іспит		
48	подія C - третій студент НЕ здав іспит		
49			
50	Ймовірності описаних подій		
51	P(A)= 0.8	P(A)= 0.2	
52	P(B)= 0.9	P(B)= 0.1	
53	P(C)= 0.7	P(C)= 0.3	

55	подія D - усі три студенти здали іспит					
56	а) $P(D)=P(A \cap B \cap C)=$		0.504			
57						
58	подія E - тільки один студент здав іспит					
59	б) $P(E)=P(A \cap \bar{B} \cap \bar{C}) \text{ або } P(\bar{A} \cap B \cap \bar{C}) \text{ або } P(\bar{A} \cap \bar{B} \cap C)=$					
60	0.024	+	0.054	+	0.014	=
61	= 0.092					
62						
63	подія F - тільки перший студент здав іспит					
64	в) $P(F)=P(A \cap \bar{B} \cap \bar{C})=$		0.024			

Варіанти завдань для самостійної роботи:

Варіант №1.

Завдання 1: $P_{12}; C_{38}^9; A_{23}^7$

Завдання 2: $\frac{A_{13}^2 + C_{21}^4}{P_{11}} - 4 \cdot C_{13}^4$

Завдання 3: У партії із 15 деталей 7 стандартних. Знайти ймовірність того, що серед 5 деталей, що були узяті навмання, тільки 3 є стандартними.

Завдання 4: Проводиться хімічний аналіз проб води із трьох джерел. Імовірність того, що хімічний склад води з першого джерела перебуває в межах норми, дорівнює 0,6; із другого - 0,8; із третього - 0,5. Знайти ймовірність того, що у межах норми склад води буде: а) хоча б з одного джерела; б) із всіх трьох джерел; в) тільки із двох джерел; г) тільки з другого і третього джерел.

Варіант №3.

Завдання 1: $P_{19}; C_{17}^6; A_{15}^3$

Завдання 2: $\frac{A_{28}^{10} + C_{19}^6}{P_{23}} - 4 \cdot A_{10}^3$

Завдання 3: У групі 15 юнаків і 10 дівчат. Для чергування призначають 5 студентів. Яка ймовірність того, що серед них 2 дівчини?

Завдання 4: Три менеджери влаштовуються на роботу. Перший з них на 90% відповідає вимогам роботодавця, другий – на 70%, третій – на 80%. Знайти ймовірність того,

Варіант №2.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6,3 \cdot C_{23}^{15}$

Завдання 3: Набираючи номер телефону, абонент забув останні дві цифри і, пам'ятаючи лише, що ці цифри різні, набрав їх навмання. Знайти ймовірність того, що набрано потрібні цифри.

Завдання 4: На зупинці очікують три автобуси. Імовірність того, що автобус №17 приїде на зупинку за розкладом, дорівнює 0,8, №443 – 0,6 і № 26 – 0,2. Знайти ймовірність того, що: а) за розкладом приїде хоча б один автобус; б) за розкладом приїдуть автобуси №26 і № 17; в) усі автобуси спізняться; г) спізниться один автобус.

Варіант №4.

Завдання 1: $P_{18}; C_{21}^7; A_{11}^6$

Завдання 2: $\frac{A_{23}^6 C_{18}^3}{P_{10}} - \frac{1}{2} \cdot C_{23}^{15}$

Завдання 3: У пакунку 6 білих і 4 чорних мотків пряжі. Із пакунка навмання беруть 2 мотки. Яка ймовірність того, що вони одного кольору?

Завдання 4: До складу доставили 2 партії товару. У першій партії 76% якісних виробів, у другій – 83%. Навмання вибирають по одному ви-

що: а) на роботу візьмуть хоча б одного менеджера; б) відмовлять першому і третьому менеджерам; в) на роботу візьмуть двох менеджерів; г) відмовлять одному менеджеру.

Варіант №5.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: У підручнику 207 сторінок. Яка ймовірність того, що порядковий номер навання відкритої сторінки буде закінчуватися цифрою 3?

Завдання 4: Три стрільця роблять по одному пострілу по мішені. Ймовірності влучення в мішень для першого стрільця дорівнює 0,5, для другого – 0,8 і для третього – 0,75, Визначити ймовірність того, що: а) у мішені буде тільки дві пробоїни; б) у мішень влучить хоча б один стрілець; в) у мішень влучить саме другий стрілець, г) у мішені буде одна пробоїна.

Варіант №7.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: У партії 40 стандарт-

робу з кожної партії. Яка ймовірність виявити серед них: а) хоча б одне браковане; б) два якісних; в) одне якісне й одне браковане; г) якісне з другої партії.

Варіант №6.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: В механізм входять дві однакові деталі. Механізм не буде працювати, якщо обидві деталі зменшеного розміру. У складальника 10 деталей, серед яких 3 зменшеного розміру. Знайти ймовірність того, що механізм буде працювати нормально, якщо складальник бере навання дві деталі.

Завдання 4: Із трьох гармат зробили залп по цілі. Імовірність влучення в ціль при одному пострілі з першої гармати дорівнює 0,9; для другої – 0,8, для третьої – 0,9. Знайти ймовірність того, що: а) тільки один снаряд влучить у ціль; б) другий і третій снаряди не влучать у ціль; в) хоча б один снаряд влучить у ціль; г) усі три снаряди не влучать у ціль

Варіант №8.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: Студент знає 50 з 60

них і 4 нестандартних деталей. Для контролю узяли навмання 8, які виявилися стандартними. Знайти ймовірність того, що наступна деталь буде також стандартна.

Завдання 4: Робітник обслуговує три верстати. Ймовірність того, що протягом години перший верстат не зупиниться, дорівнює 0,81; другий – 0,65; третій – 0,43. Визначити ймовірність того, що протягом години: а) не зупиниться жоден верстат; б) зупиняться будь-які два верстати; в) зупиняться хоча б один верстат; г) не зупиняться перший і третій верстати.

Варіант №9.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: На столі лежать 30 екзаменаційних білетів. Яка ймовірність того, що номер навмання узятого білета буде кратним 3 або 7?

Завдання 4: Для повідомлення про пожежу встановили три незалежно працюючих автомати. Ймовірність того, що при пожежі спрацює перший автомат, дорівнює 0,81; другий – 0,83, третій – 0,85. Знайти ймовірність того, що при аварії надійде сигнал: а) тільки від одного автомата; б) хоча б від одного автомата; в) тільки від третього автомата; г) від будь-яких двох автоматів.

питань. Знайти ймовірність того, що студент відповість на два питання.

Завдання 4: Пристрій складається із трьох незалежно працюючих елементів. Ймовірність безвідмовної роботи першого елемента дорівнює 0,5; другого – 0,4; третього – 0,7. Знайти ймовірність того, що за одну годину будуть працювати безвідмовно: а) всі три елементи; б) тільки два елементи; в) хоча б один елемент; г) тільки другий елемент.

Варіант №10.

Завдання 1: $P_{10}; C_{24}^6; A_{31}^2$

Завдання 2: $\frac{A_{20}^6 + C_{13}^{10}}{P_{15}} - 6.3 \cdot C_{23}^{15}$

Завдання 3: На 20 однакових жетонах написано 20 двозначних чисел від 11 до 30. Яка ймовірність того, що номер навмання узятого жетона буде кратним 4 або 7?

Завдання 4: Від аеровокзалу відправляються три автобуси до трапа літаків. Ймовірність своєчасного прибуття кожного автобуса дорівнює 0,8. Знайти ймовірність того, що: а) усі автобуси прибудуть своєчасно; б) два автобуса спізняться; в) тільки один автобус прибуде своєчасно; г) хоча б один автобус прибуде своєчасно

Контрольні запитання до захисту лабораторної роботи №1:

1) $A_{67}^5 + A_{23}^4 + P_4$

2) $C_{18}^6 - P_{32} + P_9$

3) $1,5 * (C_{1023}^1 + A_{26}^3) + P_{16}$

4) $\frac{2 \cdot A_6^5 + P_4}{C_8^2} - 50689$

5) $\frac{A_6^5 + A_{68}^2}{C_{100}^{97}} - 4 \cdot P_{10}$

6) $\left(\frac{3 \cdot P_{17}}{C_{75}^5} + 6 \frac{C_{152}^{150}}{A_{10}^3} \right)$

Задача 1: На совещание явилось 10 человек, и все они обменялись рукопожатиями. Сколько было рукопожатий?

Задача 2: В четырех номинациях кинофестиваля участвуют 7 фильмов. Вычислить число вариантов распределения призов, если по каждой номинации установлены:

а) различные призы (первый лучше второго и т.д.); б) одинаковые призы.

Задача 3: Сколько различных слов можно составить переставляя буквы в слове «школа» (если каждую комбинацию считать словом)?

Задача 4: В Стране Чудес есть три города: А, Б и В. Из города А в город Б ведет 6 дорог, а из города Б в город В - 4 дороги. Сколькими способами можно проехать от А до В?

Задача 5: В бригаде 52 рабочих. Необходимо выбрать бригадира, его заместителя и профорга. Сколько существует способов это сделать?

Задача 6: Двенадцать актеров разбиты на три квартета. Сколько может быть различных квартетов?

Задача 7: У Нины 7 разных шоколадных конфет, у Коли 9 разных карамелек. Сколькими способами они могут обменяться друг с другом пятью конфетами?

Задача 8: У людоеда в подвале томятся 14 пленников.

а) Сколькими способами он может выбрать 7-х из них себе для 7-ми приемов пищи? б) А сколько есть способов выбрать пятерых, чтобы отпустить на свободу?

Задача 9: Сколько восьмизначных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, 6, 7 так, чтобы цифры в числе не повторялись?

Лабораторна робота №2

Тема заняття:

Обчислення повної ймовірності події та апостеріорної ймовірності гіпотези за умов, що подія відбулася

Мета заняття:

Закріпити теоретичні знання щодо визначення повної ймовірності випадкової події та використання формули Байєсса;

Навчитися використовувати дерево станів для проведення обчислень за допомогою MS Excel.

Теоретичні положення:

При аналізі можливого розвитку подій та визначенні ризиків, що пов'язані з впливом факторів, появу яких можна передбачити, застосовується **формула повної ймовірності**. Нехай подія A може наступити лише за умови появи однієї з несумісних подій (гіпотез) H_1, H_2, \dots, H_n , які утворюють повну групу подій, тоді ймовірність цієї події визначається формулою:

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A | H_i), \quad (2.1)$$

де $P(H_i)$ – ймовірність реалізації гіпотези H_i ;

$P(A | H_i)$ – умовна ймовірність події A , тобто ймовірність події A за умови, що реалізується гіпотеза H_i .

Оскільки гіпотези H_1, H_2, \dots, H_n утворюють повну групу несумісних подій, то повинні виконуватися наступні умови:

$$H_i \cap H_j = \emptyset, \text{ якщо } i \neq j, \text{ та } \sum_{i=1}^n P(H_i) = 1. \quad (2.2)$$

Обчислення апостеріорної ймовірності здійснюється за формулою Байєса. Нехай подія A , що може мати місце при реалізації однієї з гіпотез H_1, H_2, \dots, H_n , вже відбулася. Ймовірність того, що ця подія відбулася саме в наслідок реалізації гіпотези H_i (апостеріорна ймовірність), визначається за **формулою Байєса**:

$$P(H_i | A) = \frac{P(H_i) \cdot P(A | H_i)}{\sum_{i=1}^n P(H_i) \cdot P(A | H_i)}, \quad (2.3)$$

де $P(H_i | A)$ – ймовірність i -ої гіпотези за умов, що подія A відбулася.

Приклад 2.1. Підприємець має можливість орендувати місце на одному з двох торговельних майданчиків: закритому чи відкритому. Якщо погода сонячна, то торгівля йде на обох майданчиках, при цьому ймовірність досягти певного рівня прибутковості („успіх”) складає 0,7 для відкритого майданчика та 0,6 для закритого. Якщо йде дощ, то на відкритому майданчику покупців нема, а для закритого майданчика ймовірність отримати певний рівень прибутку зберігає те ж саме значення 0,6. Протягом року дощова погода спостерігається в середньому для половини днів. Визначить ймовірність отримання певного рівня прибутку, якщо в хорошу погоду підприємець з однаковою ймовірністю обирає торгівлю на відкритому чи закритому майданчиках, а в погану – тільки на закритому.

Розв’язання прикладу 2.1. Розглянемо стани, у яких може перебувати система „Підприємець” (рис. 2.1).

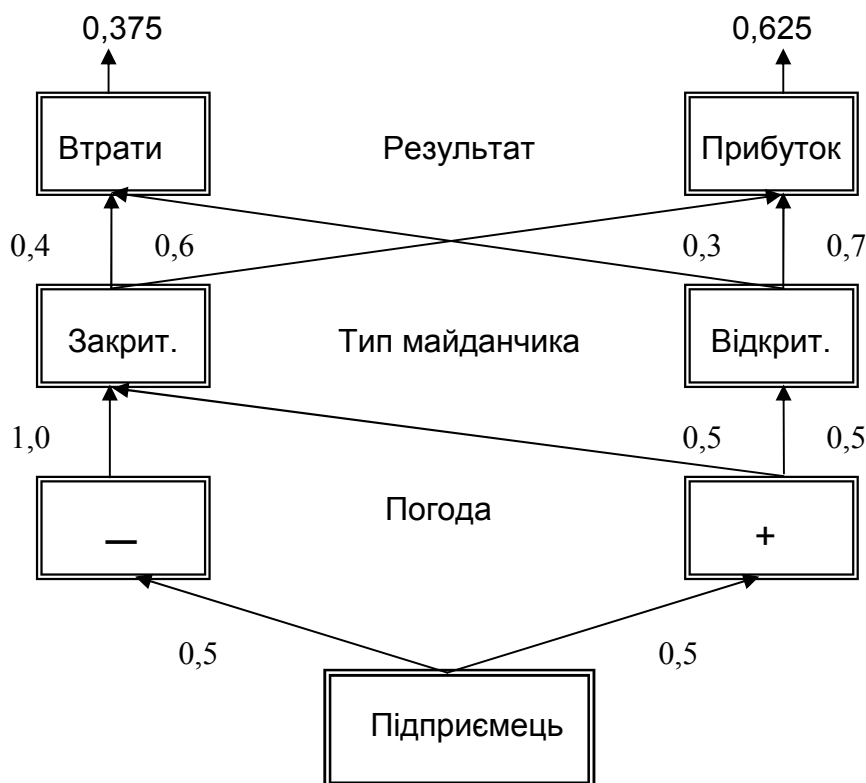


Рис. 2.1. Множина станів системи „Підприємець”

На рис. 2.1 визначена множина станів, у яких може перебувати система, показані шляхи переходів та ймовірності, з якими ці переходи здійснюються. Маємо два послідовних рівні гіпотез, на кожному з яких виконуються умови (2.2). На першому рівні маємо гіпотези відносно стану погоди – гарна чи погана. Гіпотези другого рівня визначають тип майданчика – закритий чи відкритий.

Для обчислення ймовірності успіху за допомогою MS Excel можна запропонувати такий метод. На робочому аркуші MS Excel у вигляді рядків (як показано на рис. 2.2) запишемо всі повні шляхи, якими цей успіх досягається, та відповідні їм ймовірності. Кожний шлях має вихідний стан „Підприємець” і кінцевий стан – „Успіх”. Ці шляхи визначаються за множиною станів системи (див. рис. 2.1), якщо рухатися вздовж стрілок. Стани системи позначається прямокутником з відповідною назвою, а перехід із одного стану в інший – стрілкою, над якою підписана ймовірність даного переходу. На рис. 2.2 у рядку формул приведена формула обчислення ймовірності успіху, якщо торгівля відбувається на закритому майданчику за умов поганої погоди.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		Вихідний стан		Погода		Тип майданчика		Кінцевий результат		Ймовірність	
3			0,5		0,5		0,6				
4		Підприємець	→	+	→	Закритий	→	Успіх	→	0,15	
5			0,5		0,5		0,7				
6		Підприємець	→	+	→	Відкритий	→	Успіх	→	0,175	
7			0,5		1		0,6				
8		Підприємець	→	-	→	Закритий	→	Успіх	→	=C7*E7*G7	
9											
10											

Рис. 2.2. Фрагмент робочого аркуша MS Excel, що містить схему обчислення повної ймовірності

Отже, відповідно до формули повної ймовірності (2.1) ймовірність досягти успіху (отримати певний рівень прибутку) становить:

$$P(A) = 0,5 \cdot 0,5 \cdot 0,6 + 0,5 \cdot 0,5 \cdot 0,7 + 0,5 \cdot 1 \cdot 0,6 = 0,625.$$

Приклад 2.2. Повернемося до умови прикладу 2.2. Нехай подія A вже відбулася, тобто в деякий день торгівля дала можливість підприємцю отримати певний рівень прибутку. Визначити ймовірність того, що це мало місце тоді, коли погода була поганою.

Розв'язання прикладу 2.2. Оскільки подія A вже відбулася, обчислення ймовірності здійснюється за формулою (2.3), у знаменнику міститься повна ймовірність події A (це відповідь за прикладом 2.1), а у чисельнику – ймовірність успіху за умов поганої погоди (це останній рядок на схемі на рис. 2.2). Відповідну формулу вводимо до будь-якої клітини й отримуємо:

$$P(H_i | A) = \frac{0,5 \cdot 1 \cdot 0,6}{0,625} = 0,48.$$

Можна зробити висновок, що якщо успіх є, то він майже в рівних мірах забезпечується торгівлею на відкритому та закритому майданчиках.

Варіанти завдань для самостійного розв'язання

Варіант №2.1.

Біля верстату знаходиться три деталі, до яких додали одну стандартну. Потім навмання взяли одну деталь. Визначити: а) ймовірність того, що ця деталь буде стандартною, якщо відносно якості трьох вихідних деталей можна зробити будь-яке припущення; б) апостеріорні ймовірності всіх гіпотез за умов, що випадково взята деталь виявилася нестандартною, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.2.

Фірма має трьох постачальників, обсяг постачання для кожного з яких становить відповідно 30%, 20% та 50%. Кожний з постачальників дотримується стандарту при виготовленні продукції з ймовірністю 0,8, 0,7 та 0,9, відповідно. Визначити: а) ймовірність того, що одиниця продукції, яка була вибрана для перевірки, причому цей вибір здійснювався дові-

льно, виявилася якісною; б) апостеріорні ймовірності кожної з гіпотез щодо постачальника, якщо продукція виявилася неякісною, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.3.

Три експерти роблять висновок щодо можливостей здійснення певного проекту, і відповідь надається у формі „так” чи „ні”. Перший експерт володіє інформацією на 80%, другий – на 90%, третій – на 95%. Визначити: а) ймовірність того, що відповіді всіх трьох експертів співпадуть; б) апостеріорні ймовірності кожної з гіпотез щодо правильності їх висновків за умов, що їх відповіді співпали, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.4.

У першій з шухляд знаходяться п'ять білих та сім чорних кульок, у другій – три білих та три чорних кульки, у третій – шість білих та шість чорних. З першою шухляди до другої переклали деяку кульку, після чого з другої до третьої теж переклали одну кульку, колір якої невідомий. Потім з третьої шухляди вийняли одну кульку. Визначити: а) ймовірність того, що остання кулька є білою; б) апостеріорні ймовірності кожної з гіпотез щодо кольору кульок, які переклали з першої шухляди до другої та з другої до третьої, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.5.

Під час експлуатації протягом певного терміну система може зберігати робочий стан з ймовірністю 0,7, досягти критичного стану з ймовірністю 0,2 або перейти в аварійний стан з ймовірністю 0,1. При досягненні критичного стану система самостійно повертається до робочого стану з ймовірністю 0,6 або переходить у катастрофічний стан. При досягненні катастрофічного стану вмикається зовнішнє джерело впливу, яке повертає систему до робочого стану з ймовірністю 1,0. Визначити: а) ймовірність того, що після двох термінів експлуатації система буде знаходитися у катастрофічному стані; б) апостеріорні ймовірності кожної з гіпотез щодо шляхів перетворення системи, якщо виявилось, що після двох термінів експлуатації система знаходиться в робочому стані.

Варіант №2.6.

Магазин має трьох постачальників, обсяг постачання для кожного з яких становить відповідно 10%, 20% та 70%. Кожний з постачальників дотримується стандарту при виготовленні продукції з ймовірністю 0,8, 0,7 та 1,0, відповідно. Визначити: а) ймовірність того, що одиниця продукції, яка була навмання вибрана для перевірки, виявилася неякісною; б) апостеріорні ймовірності кожної з гіпотез щодо постачальників, якщо продукція виявилася якісною, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.7.

Три консультанти роблять висновок щодо можливостей застосування певного проекту, і їх відповідь надається у формі „так” чи „ні”. Перший консультант володіє інформацією на 70%, другий – на 80%, третій – на 90%. Визначити: а) ймовірність того, що відповіді всіх трьох консультантів не співпадут; б) апостеріорні ймовірності кожної з гіпотез щодо правильності їх висновків за умов, що їх відповіді співпали, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.8.

В шухляді зберігаються три лампи, до яких додали одну стандартну. Потім навмання взяли одну з ламп. Визначити: а) ймовірність того, що ця лампа є нестандартною, якщо відносно якості трьох вихідних ламп можна зробити будь-яке припущення; б) апостеріорні ймовірності всіх гіпотез за умов, що випадково взята лампа виявилася стандартною, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.9.

Завод має три цехи, обсяг робіт для кожного з яких становить відповідно 70%, 20% та 10%. Кожний цех при виготовленні продукції дотримується стандарту з ймовірністю 0,8, 0,7 та 0,9, відповідно. Визначити: а) ймовірність того, що одиниця продукції, яка була навмання вибрана для перевірки, виявилася неякісною; б) апостеріорні ймовірності кожної з гіпотез щодо постачальника, якщо продукція виявилася якісною, та перевірити, чи виконуються відносно них умова (2.2).

Варіант №2.10.

У першій з урні знаходяться сім білих та сім чорних кульок, у другій – три білих та вісім чорних кульок, у третій – чотири білих та шість чорних кульок. З першою шухляди до другої переклали одну кульку, колір якої невідомий, після чого з другої до третьої теж переклали одну кульку, колір якої також є невідомим. Після цього з третьої шухляди вийняли одну кульку. Визначити: а) ймовірність того, що остання кулька є білою; б) апостеріорні ймовірності кожної з гіпотез щодо кольору кульок, які переклали з першої шухляди до другої та з другої до третьої, та перевірити, чи виконуються відносно них умова (2.2).

Контрольні питання:

1. Навести означення умовної та безумовної ймовірностей випадкової події та надати приклади цих подій.
2. Зображення гіпотез, що утворюють повну групу подій, за допомогою діаграми Венна – Ейлера.
3. Навести означення повної ймовірності події.
4. Надати формулу визначення повної ймовірності події.
5. Пояснити, чому за формулою повної ймовірності ймовірність суми подій $P(\sum_{i=1}^n A \cdot H_i)$ визначається як сума їх ймовірностей.
6. Навести означення апостеріорної ймовірності гіпотези (якщо випадкова подія вже відбулася).
7. Визначення ймовірності гіпотези за формулою Байєсса.
8. Надати тлумачення визначенню апостеріорної ймовірності гіпотези.

Лабораторна робота № 3 – 4

Тема занять:

Вивчення особливостей біноміального закону розподілу Бернуллі – Пуассона – Лапласа.

Мета занять:

Дослідити загальну проблему про повторення однорідних незалежних випробувань; встановити границі застосування асимптотичних формул; розглянути основні задачі теорії ймовірностей, що приводять до розподілу Бернуллі – Пуассона – Лапласа.

Теоретичні положення:

Задача теорії ймовірностей про повторення однорідних незалежних випробувань була розв'язана Бернуллі для невеликого числа випробувань $n < 30$ у вигляді $P_n(m) = C_n^m p^m (1-p)^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m}$, де n – число випробувань, p – ймовірність успіху (появи деякої події) у кожному випробуванні, $q = 1 - p$, m – загальне число успіхів, $P_n(m)$ – ймовірність появи m успіхів при n випробуваннях. При великому числі випробувань ($n \geq 30$) використовують граничні формули Пуассона й Лапласа. Комп'ютер дає можливість оцінити точність цих граничних формул. Крім того, застосування комп'ютера дозволяє наочно у зручній графічній формі представити залежності розподілу Бернуллі – Пуассона – Лапласа від його параметрів n і p .

Вивчення розподілу Бернуллі засобами Excel

На робочому аркуші Excel пропонується побудувати графіки цього розподілу при різних значеннях параметра p ($0 < p < 1$) і при різних значеннях іншого параметра n ($n = 10, 20, 30, 50$). Ці графіки дозволять помітити характерні особливості розподілу Бернуллі. Крім цього, корисно переконатися, що характеристики розподілу правильно відтворюються відомими формулами: $M(m) = np$, $D(m) = npq$. Далі корисно також переконатися в справедливості правила "3-х сигм" ($\sigma_m = \sqrt{npq}$): $M(m) - 3\sigma_m < m < M(m) + 3\sigma_m$; значення m , що виходять за межі зазначеного інтервалу, малоймовірні.

Нижче наведено фрагмент робочого аркуша таблиці Excel (табл. 3.1).

Таблиця 3.1

	A	B	C	D	E	F	G	H	I	J
1	Розподіл Бернуллі									
2	$P_n(m) = n! / (m!(n-m)!) \cdot p^m \cdot q^{(n-m)}$									
3	$P_n(m) = P_n(m-1) \cdot (n-m+1) / m \cdot p/q$									
4	$P_n(0) = q^n$									
5	n = 10		n = 10		n = 10		n = 10		n = 10	
6	p = 0,1		p = 0,3		p = 0,5		p = 0,7		p = 0,9	
7	q = 0,9		q = 0,7		q = 0,5		q = 0,3		q = 0,1	
8	M = 1		M = 3		M = 5		M = 7		M = 9	
9	D = 0,9		D = 2,1		D = 2,5		D = 2,1		D = 0,9	
10	M-3S_m = -1,84605		M-3S_m = -1,34741		M-3S_m = 0,256584		M-3S_m = 2,652587		M-3S_m = 6,15395	
11	M+3S_m = 3,84605		M+3S_m = 7,347413		M+3S_m = 9,743416		M+3S_m = 11,34741		M+3S_m = 11,84605	
12										
13	m	p=0,1	m	p=0,3	m	p=0,5	m	p=0,7	m	p=0,9
14	0	0,34868	0	0,02825	0	0,00098	0	5,9E-06	0	1E-10
15	1	0,38742	1	0,12106	1	0,00977	1	0,00014	1	9E-09
16	2	0,19371	2	0,23347	2	0,04395	2	0,00145	2	3,64E-07
17	3	0,05740	3	0,26683	3	0,11719	3	0,00900	3	8,75E-06
18	4	0,01116	4	0,20012	4	0,20508	4	0,03676	4	0,00014
19	5	0,00149	5	0,10292	5	0,24609	5	0,10292	5	0,00149
20	6	0,00014	6	0,03676	6	0,20508	6	0,20012	6	0,01116
21	7	8,75E-06	7	0,00900	7	0,11719	7	0,26683	7	0,05740
22	8	3,65E-07	8	0,00145	8	0,04395	8	0,23347	8	0,19371
23	9	9E-09	9	0,00014	9	0,00977	9	0,12106	9	0,38742
24	10	1E-10	10	5,9E-06	10	0,00098	10	0,02825	10	0,34868

Розглянемо уважно перший блок (стовпці A, B таблиці Excel).

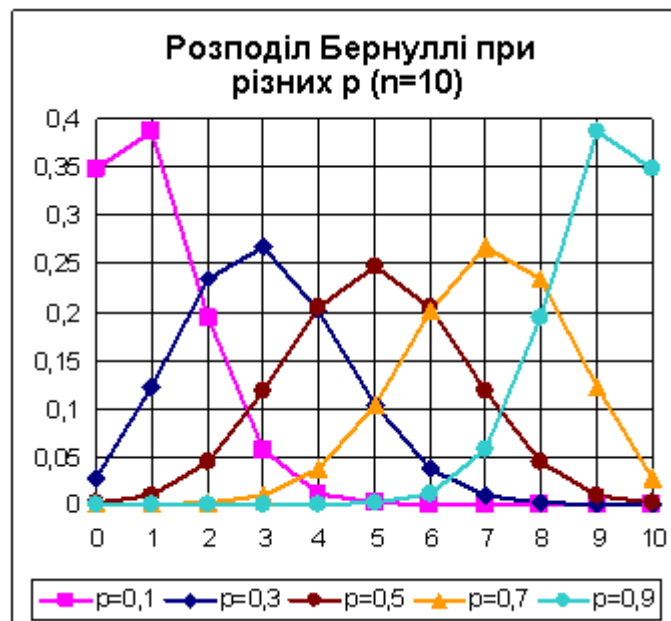
У рядках 5 і 6 задаємо значення параметрів $n = 10$, $p = 0,1$. У наступних рядках обчислюємо $q = 1 - p$, $M = \sum(m \cdot P_n(m))$, $D = \sum(m^2 \cdot P_n(m)) - M^2$, $M-3S_m = M - 3\sqrt{D}$, $M+3S_m = M + 3\sqrt{D}$. Останні 4 формули можна набрати пізніше, коли буде заповнено діапазон B14:B24, що містить значення $P_n(m)$. Відзначимо корисний прийом: у стовпці A записуємо текст і зрушуємо його праворуч, а в стовпці B обчислюємо числове значення й зрушуємо його ліворуч. Виходить зрозумілий коментар до виконаних дій. Аркуш Excel, крім усього іншого, є звітним документом, тому не варто заощаджувати на коментарях і заголовках. З інформації в рядках 8 – 11 першого блоку, видно, що, дійсно, $M = np = 10 \cdot 0,1 = 1$; $D = npq = 10 \cdot 0,1 \cdot 0,9 = 0,9$; і що всі ймовірні значення m не перевершують 4.

Значення $P_n(m)$ зручно обчислювати по рекурентній формулі, що наведена в рядку 3 робочого аркушу: $P_n(m) = P_n(m-1) \cdot (n-m+1) / m \cdot p/q$. Почат-

кове значення $P_n(0) = q^n$ обчислено в клітинці B14. При наборі рекурентної формули в клітинці B15 варто зафіксувати (знаками \$) незмінні значення n, p, q . Далі формула копіюється нижче до клітинці B24.

Заповнивши перший блок, копіюємо його кілька разів праворуч і у нових блоках заміняємо значення параметра p на $p = 0,3; p = 0,5; p = 0,7; p = 0,9$. Все автоматично перераховується. У блоках сірим фоном виділені значення $P_n(m)$, які визнані значимими за правилом "3-х сигм".

Тепер будуємо графіки. Виділяємо значення m разом із заголовком в клітинці A13, далі при натиснутій клавіші *Ctrl* виділяємо мишкою значення $P_n(m)$ для $p = 0,1; 0,3; 0,5; 0,7; 0,9$. Виділяти діапазони треба разом із заголовками в рядку 13, тоді ці заголовки автоматично будуть відображені в легенді (поясненнях до кожної лінії на графіку). Викликаємо *Майстер діаграм*, вибираємо тип діаграми – точкова, легенда – унизу, лінії сітки – основні, заголовок: "Розподіл Бернуллі при різних p ($n=10$)". У результаті одержимо графік, що майже не вимагає додаткового форматування:



Із цього графіка видно, як змінюється асиметрія розподілу при збільшенні параметра p : при $p = 0,5$ розподіл симетричний, при $p < 0,5$ – розподіл скошений ліворуч (позитивна асиметрія), а при $p > 0,5$ – скошений праворуч (негативна асиметрія).

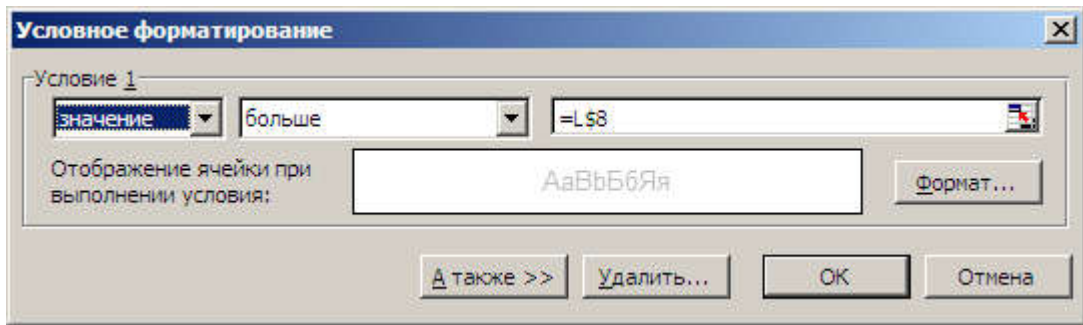
Як вже вказувалося вище, заголовки з рядка 13 автоматично переносяться в легенду діаграми. Але тоді бажане, щоб вони автоматично коректувалися при зміні параметра p . Тому як заголовок в клітинці

B13 набрана формула $=\text{"p="}\&\text{ТЕКСТ}(B9;\text{"0,0"})$. Функція $\text{ТЕКСТ}(\text{Число};\text{Формат})$ переводить у символну форму значення p з клітинки *B9*; у тексті заголовка це число буде округлене до одного знака після десяткової коми. Інші заголовки в рядку 13 коректуються автоматично при копіюванні.

Тепер переходимо до вивчення залежності розподілу Бернуллі від іншого параметра n . Скопіюємо всі 5 готових блоків праворуч, починаючи зі стовпця *K*, і замінимо в нових блоках значення параметрів: $n = 10, 20, 30, 40, 50$ і $p = 0,1$ (для всіх нових блоків). Нові таблиці треба продовжити униз до рядка 64 (вони тепер будуть мати різну довжину).

	K	L	M	N	O	P	Q	R	S	T
5	n =10		n =20		n =30		n =40		n =50	
6	p =0,1		p =0,1		p =0,1		p =0,1		p =0,1	
7	q =0,9		q =0,9		q =0,9		q =0,9		q =0,9	
8	M =1		M =2		M =3		M =4		M =5	
9	D =0,9		D =1,8		D =2,7		D =3,6		D =4,5	
10	M-3Sm=-1,84605		M-3Sm=-2,02492		M-3Sm=-1,9295		M-3Sm=-1,6921		M-3Sm=-1,36396	
11	M+3Sm=3,84605		M+3Sm=6,024922		M+3Sm=7,929503		M+3Sm=9,6921		M+3Sm=11,36396	
12										
13	m	n=10	m	n=20	m	n=30	m	n=40	m	n=50
14	0	0,348678	0	0,121577	0	0,042391	0	0,014781	0	0,005154
15	1	0,387420	1	0,270170	1	0,141304	1	0,065693	1	0,028632
16	2	0,193710	2	0,285180	2	0,227656	2	0,142334	2	0,077943
17	3	0,057396	3	0,190120	3	0,236088	3	0,200323	3	0,138565
18	4	0,011116	4	0,089779	4	0,177066	4	0,205887	4	0,180905
19	5	0,001488	5	0,031921	5	0,102305	5	0,164710	5	0,184925
20	6	0,000138	6	0,008867	6	0,047363	6	0,106756	6	0,154104
21	7	8,75E-06	7	0,00197	7	0,018043	7	0,057614	7	0,107628
22	8	3,65E-07	8	0,000356	8	0,005764	8	0,026407	8	0,064278
23	9	9E-09	9	5,27E-05	9	0,001565	9	0,010432	9	0,033329
24	10	1E-10	10	6,44E-06	10	0,000365	10	0,003593	10	0,015183
25	11	0	11	6,51E-07	11	7,38E-05	11	0,001089	11	0,006135

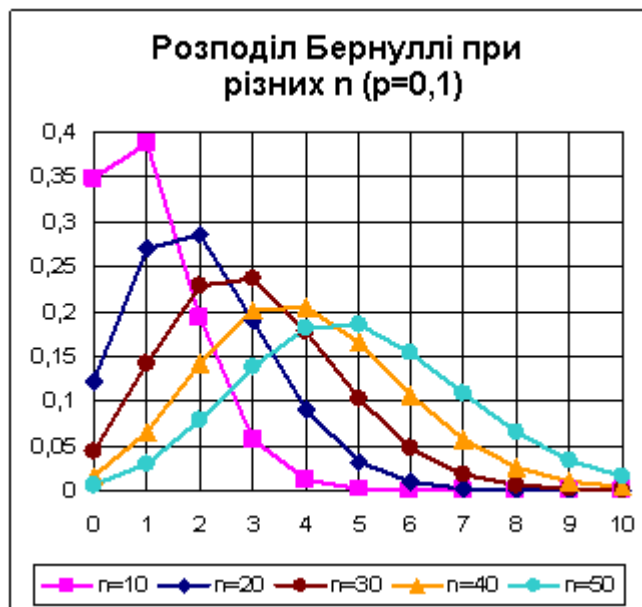
Непотрібну інформацію можна приховати за допомогою умовного форматування. Так, таблиця для $n = 10$ фактично обривається на рядку 24, тому можна зробити так, щоб подальші значення m і нульові значення $P_n(m)$ виводилися сірим кольором на білому фоні (тоді вони майже не будуть видні). Умовний формат для стовпчика m задаємо за умовою:



Зверніть увагу на те, що в посиланні на клітинку $L8$ зафіксований тільки номер рядка. Для стовпчика $P_n(m)$ із заголовком $n=10$ умова буде більш простою: "Значения равны 0". При копіюванні відформатованого блоку, копіюються також всі умовні формати.

Нарешті, треба замінити заголовки в рядку 13 на формули $=\text{"n"}\&\text{ТЕКСТ}(L8;\text{"0"})$.

Цікаво, що хоча таблиці продовжуються до рядка 64, фактично (згідно правила "3-х сигм") їх можна було обірвати на рядку 25 (це відіб'ється тільки на значеннях M і D у рядках 8, 9). Усе готово для побудови нового графіка, з якого буде видно, як зі збільшенням n розподіл Бернуллі наближається до якоїсь стандартної форми – до розподілу Лапласа, або до, так званого, нормального закону розподілу Гаусса.



Вважається, що при $n \geq 30$ розподіл вже практично нормальний. Це питання ще буде обговорюватися нижче при вивченні розподілу Лапласа. Там же розглянемо застосування кумуляти.

Вивчення розподілу Пуассона засобами Excel

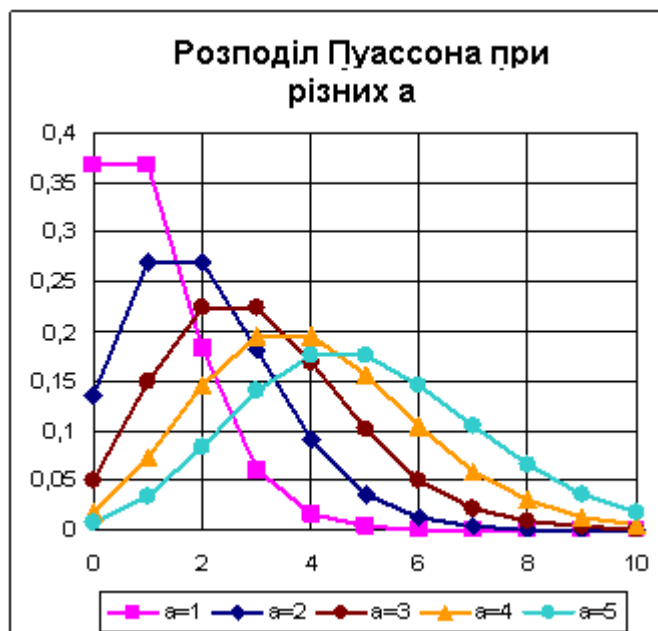
Розподіл Пуассона $P(m) = e^{-a} \frac{a^m}{m!}$ залежить від одного параметра a . Пропонується на робочому аркуші Excel побудувати графіки цього розподілу для різних значень параметра $a = 1, 2, 3, 4, 5$, які дозволять помітити характерні риси розподілу й переконатися, що при збільшенні параметра розподіл Пуассона наближається до граничного розподілу Лапласа (або ж до нормального розподілу Гаусса). Вважається, що при $a \geq 5$ розподіл вже практично нормальний. Розподіл Пуассона є асимптотичним наближенням для розподілу Бернуллі, якщо $n \gg 1$, $p \ll 1$, $a = np$. Тут вираз $n \gg 1$ означає "n – дуже велике", а вираз $p \ll 1$ – "p – дуже маленьке". Вважається, що граничним розподілом можна користуватися при $n \geq 30$, $p \leq 0,1$ і $a = np < 5$. На комп'ютері можна перевірити й уточнити ці твердження.

У розподілі Пуассона m теоретично не обмежене ($m \rightarrow \infty$). Проте правило "3-х сигм" залишається в силі: $a - 3\sigma_m < m < a + 3\sigma_m$, де $\sigma_m = \sqrt{a}$; значення m , що виходять за межі зазначеного інтервалу, малоймовірні. Корисно також переконатися, що характеристики розподілу правильно відтворюються відомими формулами: $M(m) = a$, $D(m) = a$. Цікаво, наскільки близькими до теоретичних значень будуть розрахунки за вихідними формулами $M = \sum(m \cdot P(m))$, $D = \sum(m^2 \cdot P(m)) - M^2$, якщо обмежитися діапазоном m , зазначеним правилом "3-х сигм"?

Нижче наведено фрагмент робочого аркуша таблиці Excel. Він оформлений аналогічно робочому аркушу з розрахунками за формулами Бернуллі, лише характеристики M і D тепер розраховані по діапазону $m \leq a + 3\sqrt{a}$ (правило "3-х сигм"), тому вони трохи менше теоретичних значень $M = D = a$. Послідовні значення ймовірностей зручно обчислювати за рекурентною формулою $P(m) = P(m-1) \cdot \frac{a}{m}$, де $P(0) = e^{-a}$ (ці формули наведені в рядку 3 робочого аркушу). Значення $P(m)$, що значимі за правилом 3-х сигм, виділені сірим фоном. Підготувавши перший блок (стовпці A, B), копіюємо його кілька разів вправо й у копіях заміняємо значення параметра на $a = 2, 3, 4, 5$. Практично все перераховується автоматично (за винятком характеристик M і D , де потрібно вручну уточнювати діапазони).

	A	B	C	D	E	F	G	H	I	J	
1	Розподіл Пуассона										
2	$P(m)=e^{(-a)}*a^m/m!$										
3	$P(m)=P(m-1)*a/m$										
4	$P(0)=e^{(-a)}$										
5											
6	a =1		a =2		a =3		a =4		a =5		
7											
8	M ≈0,981012		M ≈1,966873		M ≈2,964286		M ≈3,967471		M ≈4,972735		
9	D ≈0,938326		D ≈1,887672		D ≈2,875715		D ≈3,88483		D ≈4,902264		
10	M-3Sm=-2		M-3Sm=-2,24264		M-3Sm=-2,19615		M-3Sm=-2		M-3Sm=-1,7082		
11	M+3Sm=4		M+3Sm=6,242641		M+3Sm=8,196152		M+3Sm=10		M+3Sm=11,7082		
12											
13	m	a=1	m	a=2	m	a=3	m	a=4	m	a=5	
14	0	0,367879	0	0,135335	0	0,049787	0	0,018316	0	0,006738	
15	1	0,367879	1	0,270671	1	0,149361	1	0,073263	1	0,033690	
16	2	0,183940	2	0,270671	2	0,224042	2	0,146525	2	0,084224	
17	3	0,061313	3	0,180447	3	0,224042	3	0,195367	3	0,140374	
18	4	0,015328	4	0,090224	4	0,168031	4	0,195367	4	0,175467	
19	5	0,003066	5	0,036089	5	0,100819	5	0,156293	5	0,175467	
20	6	0,000511	6	0,012030	6	0,050409	6	0,104196	6	0,146223	
21	7	7,3E-05	7	0,003437	7	0,021604	7	0,059540	7	0,104445	
22	8	9,12E-06	8	0,000859	8	0,008102	8	0,029770	8	0,065278	
23	9	1,01E-06	9	0,000191	9	0,002701	9	0,013231	9	0,036266	
24	10	1,01E-07	10	3,82E-05	10	0,000810	10	0,005292	10	0,018133	
25	11	9,22E-09	11	6,94E-06	11	0,000221	11	0,001925	11	0,008242	
26	12	7,68E-10	12	1,16E-06	12	5,52E-05	12	0,000642	12	0,003434	

Усе підготовлено для побудови графіків розподілу Пуассона при різних значеннях параметра a .



З побудованої діаграми наочно видно, як зі збільшенням параметра a розподіл Пуассона наближається до симетричного розподілу Лапласа, і що при $a = 5$ розподіл вже дуже схоже на нормальний (див. також наступний розділ).

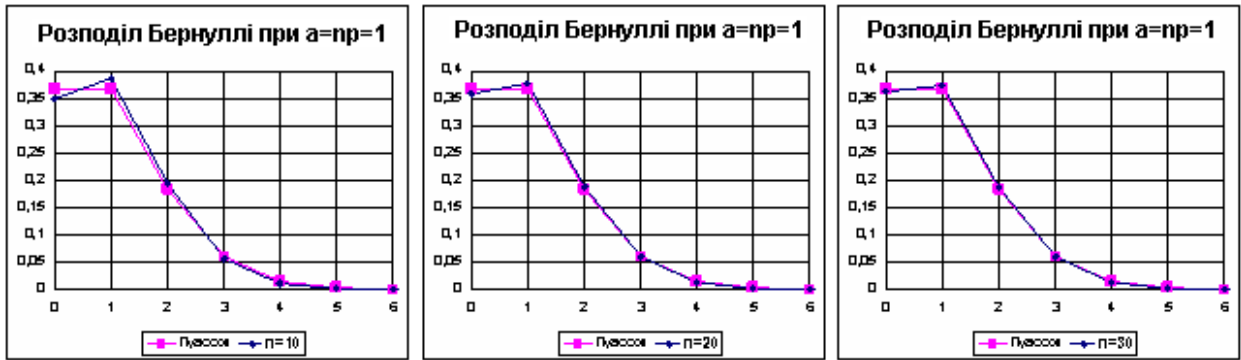
Оскільки заголовки стовпців з рядка 13 має бути перенесеними в легенду, заголовок в клітинці B13 був заданий формулою: `="a"&ТЕКСТ(В6;"0")`; при копіюванні цей заголовок автоматично коректується.

Для того, щоб перевірити, наскільки розподіл Пуассона є гарним наближенням для розподілу Бернуллі при $a = np$, скопіюємо з попереднього робочого аркуша кілька разів блоки з розрахунками розподілу Бернуллі й замінимо в нових блоках значення параметрів: $n = 10, 20, 30, 40, 50$; $p = a/n$. Формулу для p набираємо лише один раз в клітинці N6, далі її просто копіюємо. Фрагмент робочого аркуша наведений нижче.

	K	L	M	N	O	P	Q	R	S	T	U	V
4	Пуассон		Бернуллі		Бернуллі		Бернуллі		Бернуллі		Бернуллі	
5			n =10		n =20		n =30		n =40		n =50	
6	a =1		p =0,1		p =0,05		p =0,03333		p =0,025		p =0,02	
7			q =0,9		q =0,95		q =0,96667		q =0,975		q =0,98	
8	M =0,98101		M =1		M =1		M =1		M =1		M =1	
9	D =0,93833		D =0,9		D =0,95		D =0,96667		D =0,975		D =0,98	
10	M-3Sm=-2		M-3Sm=-1,8461		M-3Sm=-1,9240		M-3Sm=-1,9496		M-3Sm=-1,9623		M-3Sm=-1,9699	
11	M+3Sm=4		M+3Sm=3,84605		M+3Sm=3,92404		M+3Sm=3,94958		M+3Sm=3,96226		M+3Sm=3,96985	
12												
13	m	Пуассон	m	n=10	m	n=20	m	n=30	m	n=40	m	n=50
14	0	0,36788	0	0,34868	0	0,35849	0	0,36166	0	0,36323	0	0,36417
15	1	0,36788	1	0,38742	1	0,37735	1	0,37413	1	0,37255	1	0,37160
16	2	0,18394	2	0,19371	2	0,18868	2	0,18707	2	0,18627	2	0,18580
17	3	0,06131	3	0,05740	3	0,05958	3	0,06021	3	0,06050	3	0,06067
18	4	0,01533	4	0,01116	4	0,01333	4	0,01401	4	0,01435	4	0,01455
19	5	0,00307	5	0,00149	5	0,00225	5	0,00251	5	0,00265	5	0,00273
20	6	0,00051	6	0,00014	6	0,00030	6	0,00036	6	0,00040	6	0,00042

Тепер побудуємо порівняльні графіки розподілів Пуассона й Бернуллі при різних значеннях n і $a=np$. З цих графіків видно, що для $a = 1$ вже при $n = 20$ асимптотична формула Пуассона добре погодиться з вихідною формулою Бернуллі.

Можна поекспериментувати, змінюючи на робочому аркуші значення параметра a і спостерігаючи за відповідними трансформаціями графіків.



Вивчення розподілу Лапласа засобами Ексел

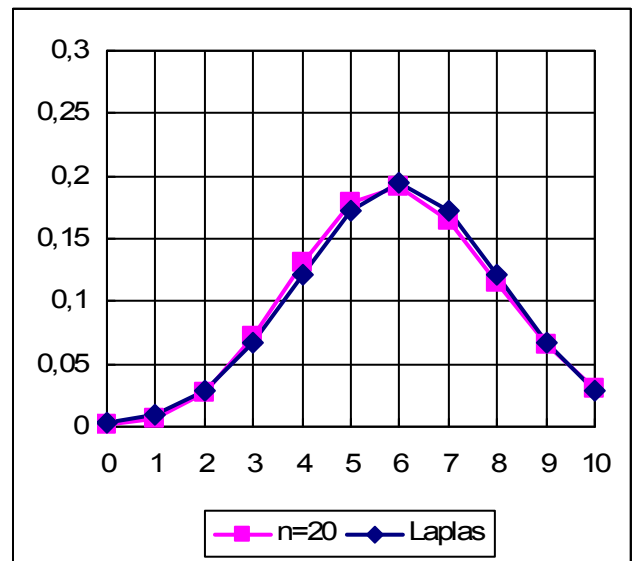
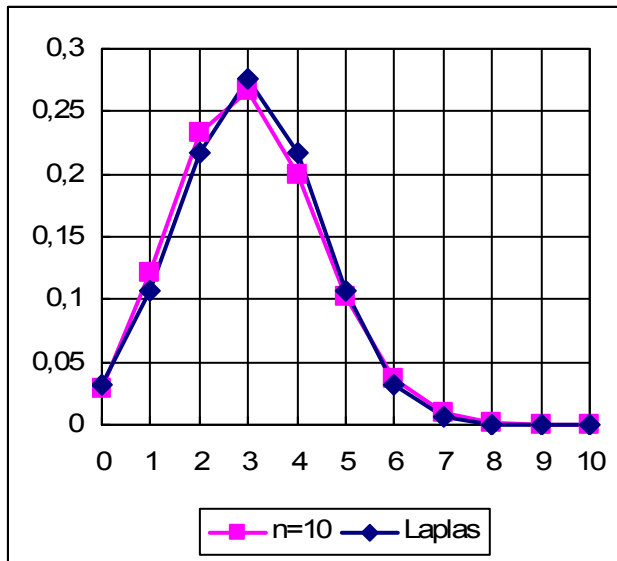
Розподіл Лапласа $P_n(m) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(m-a)^2}{2\sigma_m^2}}$ залежить від двох параметрів n і p ; характеристики розподілу виражаються через ці параметри: $a = np$, $\sigma_m = \sqrt{npq}$. При розрахунках вручну використовують таблиці диференціальної $\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$ й інтегральної $\Phi(t) = \int_0^t \varphi(s) ds = \frac{1}{\sqrt{2\pi}} \cdot \int_0^t e^{-\frac{s^2}{2}} ds$ функцій Лапласа. Тоді $P_n(m) = \frac{\varphi(t_m)}{\sigma_m}$, де $t_m = \frac{m-a}{\sigma_m}$. Ця формула звичайно називається "локальною теоремою Лапласа". Інтегральна функція Лапласа використовується для обчислення ймовірності потрапляння m у заданий інтервал: $P(m_1 \leq m \leq m_2) = \Phi(t_{m_2}) - \Phi(t_{m_1})$. Ця формула звичайно називається "інтегральною теоремою Лапласа". При розрахунках на комп'ютері для цих же цілей використовується функція НОРМРАСП(m ; a ; σ ; 0), яка повертає значення ймовірності $P_n(m)$, і НОРМРАСП(m ; a ; σ ; 1), яка повертає значення інтегральної функції розподілу Лапласа $F(m) = \Phi(t_m) + 0,5$ (функція $F(m)$ відрізняється постійним доданком від функції $\Phi(t_m)$). Вважається, що асимптотичні формули Лапласа можна застосовувати при $n \geq 30$, $a = np \geq 5$ і $nq \geq 5$. Інтегральну формулу Лапласа можна уточнити, якщо розширити інтервал $[m_1; m_2]$ на 0,5 ліворуч і праворуч: $P(m_1 \leq m \leq m_2) = F(m_2 + 1/2) - F(m_1 - 1/2)$. Пропонується прямими розрахунками перевірити точність локальної, інтегральної і скоректованої інтегральної формул Лапласа для $n = 10, 20, 30, 40, 50$ і $p = 0,3$.

Фрагмент робочого аркуша Ексел наведено нижче.

	A	B	C	D	E	F	G	H	I	J
1	Розподіл Лапласа									
2	$P_n(m)=\text{НОРМРАСП}(m;a;Sm;0)$				$a=np$	$Sm=\text{КОРЕНЬ}(npq)$				
3	$P(m_1 \leq m \leq m_2)=\text{НОРМРАСП}(m_2;a;Sm;1)-\text{НОРМРАСП}(m_1;a;Sm;1)$							Інтегральна теорема		
4	$P(m_1 \leq m \leq m_2)=\text{НОРМРАСП}(m_2+1/2;a;Sm;1)-\text{НОРМРАСП}(m_1-1/2;a;Sm;1)$ (Скоррекована)									
5	Бернуллі									
6	n =10	$\Phi(2)-\Phi(-1)=0,81859$			P([m1;m2])	n=10	n=20	n=30	n=40	n=50
7	p =0,3				Bernully	0,8401	0,87577	0,90647	0,87422	0,83588
8	q =0,7	M-Sm=1,550862			Laplas	0,73571	0,80996	0,86082	0,83047	0,79069
9	a = M =3	M+2Sm=5,898275			Correct	0,84183	0,87469	0,90419	0,87395	0,83753
10	Dm =2,1	m1 =2			Error%	-12,43%	-7,51%	-5,04%	-5,00%	-5,41%
11	Sm =1,449138	m2 =6			ErrCorr%	0,21%	-0,12%	-0,25%	-0,03%	0,20%
12	Б е р н у л л і									
13	m	n=10	F(m)	Laplas	P(a)	n=10	n=20	n=30	n=40	n=50
14	0	0,028248	0,028248	0,032298	Bernully	0,26683	0,19164	0,15729	0,13657	0,12235
15	1	0,121061	0,149308	0,106215	Laplas	0,27530	0,19466	0,15894	0,13765	0,12312
16	2	0,233474	0,382783	0,216969	Correct	0,26993	0,19275	0,15790	0,13697	0,12263
17	3	0,266828	0,649611	0,275296	Error%	3,17%	1,58%	1,05%	0,79%	0,63%
18	4	0,200121	0,849732	0,216969	ErrCorr%	1,16%	0,58%	0,39%	0,29%	0,23%
19	5	0,102919	0,952651	0,106215						
20	6	0,036757	0,989408	0,032298						
21	7	0,009002	0,998410	0,006100						
22	8	0,001447	0,999856	0,000716						
23	9	0,000138	0,999994	5,22E-05						
24	10	5,909E-06	1,000000	2,36E-06						

Ліворуч у стовпцях A, B розташований блок розрахунків за формулою Бернуллі (зараз прийнято $n = 10$). У сусідньому стовпці C наведені значення кумуляти (накопичених ймовірностей) $F(m) = \sum(P_n(k))$, обчислених за рекурентною формулою $F(m)=F(m-1)+P_n(m)$, де $F(0)=P_n(0)=q^n$. За допомогою кумуляти одним відніманням визначають ймовірності потрапляння випадкової величини m у будь-які задані інтервали $P(m_1 \leq m \leq m_2) = F(m_2) - F(m_1-1)$. Так, ймовірність $P(2 \leq m \leq 6)$ можна обчислити як різницю $F(6) - F(1) = 0,989408 - 0,149308 = 0,840140$; це ж значення можна одержати безпосередньо, підсумуючи ймовірності $P_n(m)$ для $m = 2, 3, 4, 5, 6$ (цифри на сірому фоні в стовпці B). У наступному стовпці D обчислені ймовірності за локальною формулою Лапласа. За результатами розрахунків побудовані графіки розподілу Бернуллі й розподілу Лапласа з тими самими характеристиками. Змінюючи n в клітинці B6, можна спостерігати, наскільки гарно розподіл Бернуллі описується граничним розподілом Лапласа. Нижче наведені два порівняльних

графіки для $n = 10$ і $n = 20$, звідки видно, що при $n \geq 20$ розподіл Бернуллі для $p = 0,3$ практично збігається із граничним розподілом. Таким чином, ми переконалися, що локальна формула Лапласа досить точна.



Переходимо до перевірки точності *розрахунків* за інтегральною формулою Лапласа. Обчислимо ймовірність потрапляння випадкової величини m в інтервал $[M - \sigma_m; M + 2\sigma_m]$. Теоретично зазначена ймовірність підраховується як різниця $\Phi(2) - \Phi(-1) = 0,81859$. Проте ця формула виведена для безперервного розподілу, а розподіл Лапласа – дискретний. Тому теоретичне значення (або близьке до нього) може бути отримано лише для дуже великого $n > 200$. Перевіримо це твердження. Вище над блоком розрахунків за формулою Бернуллі в стовпцях C, D обчислені границі інтервалу: $M - S_m = 1,551$ і $M + 2S_m = 5,898$; ці отримані границі округлені до найближчих цілих $m_1 = 2, m_2 = 6$. Порівняємо точне значення $P(2 < m \leq 6) = 0,8401$, що отримане раніше за допомогою кумуляти розподілу Бернуллі, з відповідним значенням за інтегральною формулою Лапласа $P(2 < m \leq 6) = \text{НОРМРАСП}(6; a; S_m; 1) - \text{НОРМРАСП}(2; a; S_m; 1) = 0,7357$. У порівнянні з точним значенням (0,8401) похибка склала -12,43%. Ці розрахунки оформлені у вигляді таблиці в рядках 6–11 у стовпцях E, F . При зміні n в клітинці $B6$ всі цифри в стовпці F автоматично коректуються. Для того, щоб зберегти результати попередніх розрахунків, вони були скопійовані *Спеціальною вставкою - Тільки значення* в продовження таблиці вправо (стовпці G, H, I, J). Розглядаючи заповнену таблицю (діапазон $E6:J11$), переконаємося у повільному наближенні ре-

зультатів граничної формули до точного значення. Тепер ми можемо оцінити ефективність скоректованої інтегральної формули Лапласа $P(2 < m \leq 6) = \text{НОРМРАСП}(6,5;a;S_m;1) - \text{НОРМРАСП}(1,5;a;S_m;1)$, яка вже для $n = 10$ привела до практично точного значення 0,8418 (похибка 0,21%).

Ефективність скоректованої формули була також перевірена на обчисленні ймовірності моди (самого ймовірного значення $m = a = np$). Ці розрахунки виконані в таблиці діапазону E13:J18. Інтегральна функція Лапласа тут дає значення 0, тому в таблиці використалася локальна формула Лапласа. Похибки локальної формули невеликі й швидко убивають зі збільшенням n (при $n = 20$ похибка становить усього 1,6%). Скоректована інтегральна формула виявилася ще більш точною; вона дає практично точні значення вже при $n = 10$.

P(a)	n=10	n=20	n=30	n=40	n=50
Bernully	0,26683	0,19164	0,15729	0,13657	0,12235
Laplas	0,27530	0,19466	0,15894	0,13765	0,12312
Correct	0,26993	0,19275	0,15790	0,13697	0,12263
Error%	3,17%	1,58%	1,05%	0,79%	0,63%
ErrCorr%	1,16%	0,58%	0,39%	0,29%	0,23%

Зразок виконання індивідуального завдання

1. Ймовірність того, що електротехнічний прилад зажадає ремонту в гарантійний строк, дорівнює 0,25. Знайти ймовірність того, що протягом гарантійного строку з семи приладів зажадають ремонту $m = 0, 1, 2, \dots$; зажадають ремонту не більш чотирьох, не менш двох, більше двох і менше шести приладів.

Рішення. У цьому завданні $p = 0,25$, $q = 1 - p = 0,75$, $n = 7$ ($n < 30$), тому застосовуємо формулу Бернуллі:

$$P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m}.$$

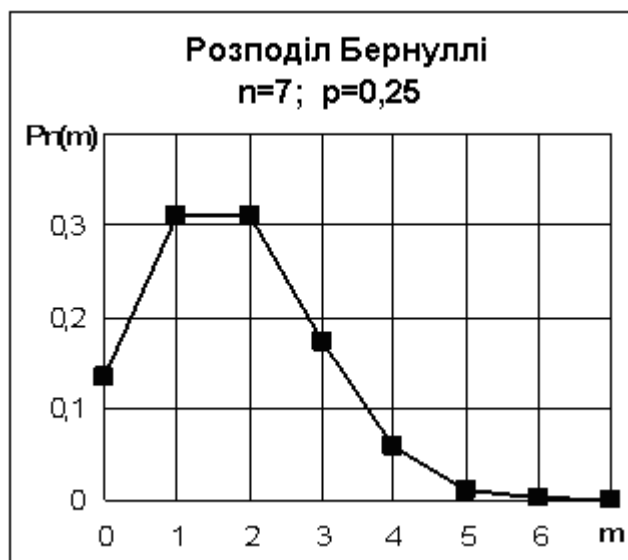
Для $m = 0, 1, 2, \dots, 7$ обчислення зведені в таблицю, в останньому стовпчику якої наведені значення накопиченої суми ймовірностей (значення кумуляти) $F(m)$. Нагадуємо, що функція розподілу (кумулята) визначена як $F(m) = P(X \leq m)$, тому $F(0) = P_n(0) = 0,133484$.

m	C_n^m	p^m	q^{n-m}	$P_n(m)$	$F(m)$
0	1	1	0,133484	0,133484	0,133484
1	7	0,25	0,177979	0,311462	0,444946
2	21	0,0625	0,237305	0,311462	0,756409
3	35	0,015625	0,316406	0,173035	0,929443
4	35	0,003906	0,421875	0,057678	0,987122
5	21	0,000977	0,5625	0,011536	0,998657
6	7	0,000244	0,75	0,001282	0,999939
7	1	0,000061	1	0,000061	1,000000

За допомогою цієї функції ймовірність потрапляння випадкової величини в інтервал $m_1 < m \leq m_2$ обчислюється як різниця значень функції розподілу на краях цього інтервалу: $P(m_1 < m \leq m_2) = F(m_2) - F(m_1-1)$.
 Обчислюємо: $P(m \leq 4) = F(4) = 0,987122$; $P(m \geq 2) = P(2 \leq m \leq 7) = F(7) - F(1) = 1 - 0,444946 = 0,555054$; $P(2 < m < 6) = P(3 \leq m \leq 5) = F(5) - F(2) = 0,998657 - 0,756409 = 0,242248$.
 Додатково обчислюємо характеристики розподілу Бернуллі: $M(m) = np = 7 \cdot 0,25 = 1,75$; $D(m) = npq = 7 \cdot 0,25 \cdot 0,75 = 1,3125$; $\sigma_m = \sqrt{D(m)} = 1,1456$.
 $M(m) - q \leq Mo \leq M(m) + p$; $1,75 - 0,75 \leq Mo \leq 1,75 + 0,25$; $1 \leq Mo \leq 2$; $P(1)=P(2)=P_{max}$.

Відповідно до правила "трьох сигм" ймовірні значення m не перевищують $M(m) + 3 \cdot \sigma_m = 1,75 + 3 \cdot 1,146 = 5,19 \approx 5$.

Нижче наведений графік полігона розподілу ймовірностей $P(m)$:



2. В одному кубічному метрі повітря в середньому знаходиться 1000 хвороботворних мікробів. На аналіз взято 2 літри (дм³) повітря. Знайти ймовірність того, що в пробі буде виявлене $m = 0, 1, 2, \dots$ хвороботворних мікробів, не більше трьох мікробів, від двох до п'яти мікробів, хоча б один мікроб.

Рішення. Ця задача на розподіл Пуассона:

$$P(m) = e^{-a} \cdot \frac{a^m}{m!},$$

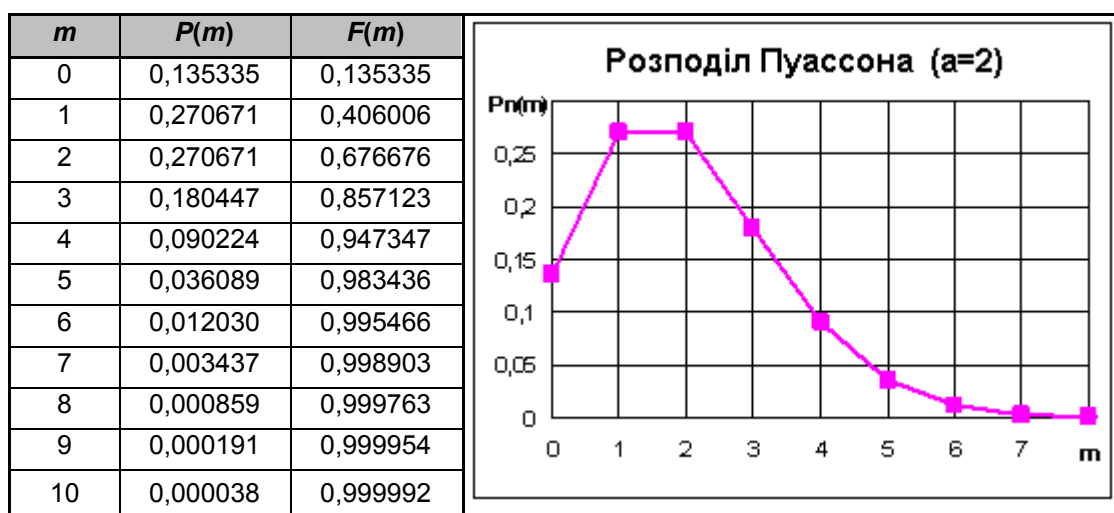
де $a = 2 \cdot 1000 / 1000 = 2$ – середній зміст мікробів в 2-х літрах повітря.

Обчислення зручно робити по рекурентним формулам:

$$P(m) = P(m-1) \cdot \frac{a}{m}; \quad P(0) = e^{-a}.$$

Можливі значення m не обмежені зверху, проте, відповідно до правила "трьох сигм", досить розрахувати ймовірності $P(m)$ для $m < a + 3\sqrt{a} = 2 + 3 \cdot 1,414 = 6,2$.

Для $m = 0, 1, 2, \dots, 10$ обчислення зведені в таблицю, поруч побудований графік:



За допомогою функції розподілу $F(m)$ обчислюємо: $P(m \leq 3) = F(3) = 0,857123$; $P(2 \leq m \leq 5) = F(5) - F(1) = 0,983436 - 0,406006 = 0,57743$; $P(m \geq 1) = 1 - P(0) = 1 - 0,135335 = 0,864665$.

Додатково обчислюємо характеристики розподілу Пуассона:

$$M(m) = a = 2; \quad D(m) = a = 2; \quad \sigma_m = \sqrt{D(m)} = 1,414.$$

$$a - 1 \leq Mo \leq a; \quad 1 \leq Mo \leq 2; \quad P(1) = P(2) = P_{max}.$$

3. Серед виробів деякого заводу 75% – першого сорту. Знайти ймовірність того, що з 800 виробів першосортних виробів буде а) 600; б) не менше 590; в) не більше 620; г) не менше 590 і не більше 620. Які відхилення відносної частоти m/n виробів першого сорту від її теоретичної долі у сукупності (75%) можна очікувати з рівнем довіри а) 95%; б) 99%? Скільки потрібно взяти виробів,

щоб з ймовірністю 90% можна було стверджувати, що відхилення відносної частоти m/n від її теоретичної долі 75% не буде перевищувати: а) 5%; б) 1%?

Рішення. У цьому завданні $p = 0,75$; $q = 1 - p = 0,25$; $n = 800 > 30$; $np = 800 \cdot 0,75 = 600 > 5$; $nq = 800 \cdot 0,25 = 200 > 5$, отже, ймовірності можна обчислювати за асимптотичними формулами Лапласа:

$$P(m) = \frac{\varphi(t_m)}{\sigma_m}; \quad t_m = \frac{m - M(m)}{\sigma_m}; \quad M(m) = np; \quad \sigma_m = \sqrt{npq},$$

де $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ – диференційна функція Лапласа;

$$P(m_1 \leq m \leq m_2) = \Phi(t_{m_2}) - \Phi(t_{m_1});$$

$$P(|m - M(m)| \leq t\sigma_m) = 2\Phi(t);$$

$$P\left(\left|\frac{m}{n} - p\right| \leq t\sqrt{\frac{pq}{n}}\right) = 2\Phi(t),$$

де $\Phi(t) = \int_0^t \varphi(s) ds = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-s^2/2} ds$ – інтегральна функція Лапласа.

Обчислюємо:

$$M(m) = np = 800 \cdot 0,75 = 600; \quad \sigma_m = \sqrt{npq} = \sqrt{800 \times 0,75 \times 0,25} = 12,2474;$$

$$t_{600} = \frac{600 - 600}{12,2474} = 0; \quad \varphi(0) = 0,3989; \quad P(600) = \frac{0,3989}{12,2474} = 0,03257.$$

$$t_{590} = \frac{590 - 600}{12,2474} = -0,816; \quad \Phi(t_{590}) = \Phi(-0,816) = -0,2924;$$

$$t_{620} = \frac{620 - 600}{12,2474} = 1,633; \quad \Phi(t_{620}) = \Phi(1,633) = 0,4488;$$

$$P(m \geq 590) = P(590 \leq m < \infty) = \Phi(\infty) - \Phi(t_{590}) = 0,5 + 0,2924 = 0,7924.$$

$$P(m \leq 620) = P(-\infty < m \leq 620) = \Phi(t_{620}) - \Phi(-\infty) = 0,4488 + 0,5 = 0,9488.$$

$$P(590 \leq m \leq 620) = \Phi(t_{620}) - \Phi(t_{590}) = 0,4488 + 0,2924 = 0,7412.$$

Відповіді на питання, що залишилися, можна одержати за допомогою "третьої форми інтегральної теореми Лапласа":

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 2\Phi(t); \quad P = 2\Phi(t); \quad \varepsilon = t\sqrt{\frac{pq}{n}}.$$

При рівні довіри $P=0,95$ $\Phi(t)=0,475$, звідки $t=1,96$ і

$$\varepsilon = t\sqrt{\frac{pq}{n}} = 1,96\sqrt{\frac{0,75 \times 0,25}{800}} = 0,030.$$

При рівні довіри $P=0,99$ $\Phi(t)=0,495$, звідки $t=2,58$ і $\varepsilon = t\sqrt{\frac{pq}{n}} = 2,58\sqrt{\frac{0,75 \times 0,25}{800}} = 0,039$.

В останнім питанні задане рівень довіри $P=0,9$ і гранична погрішність $\varepsilon = 0,05$; $\varepsilon = 0,01$.

При рівні довіри $P=0,90$ $\Phi(t)=0,45$, звідки $t=1,64$ і $\varepsilon = t\sqrt{\frac{pq}{n}} = 1,64\sqrt{\frac{0,75 \times 0,25}{n}} = \frac{0,7101}{\sqrt{n}}$.

Якщо $\varepsilon = 0,05$, то $n \geq \left(\frac{0,7101}{0,05}\right)^2 = 14,20^2 = 201,7$, тобто необхідно взяти $n=202$ вироби.

Якщо $\varepsilon = 0,01$, то $n \geq \left(\frac{0,7101}{0,01}\right)^2 = 71,01^2 = 5042,4$, тобто необхідно взяти $n=5043$ виробів.

Контрольні запитання

1. Сформулюйте завдання про повторення однорідних незалежних випробувань.
2. Приведіть формули для розрахунку характеристик розподілу Бернуллі.
3. Опишіть особливості розподілу Бернуллі при різних значеннях параметрів.
4. Сформулюйте закон рідких подій Пуассона, опишіть особливості цього розподілу.
5. У яких випадках розподіл Бернуллі припустимо апроксимувати граничним розподілом Пуассона?
6. Що таке "кумулята", як вона розраховується, де (в основному) застосовується?
7. Сформулюйте локальну і інтегральну теореми Лапласа.
8. Сформулюйте три основних форми інтегральної теореми Лапласа і типові задачі, які вирішуються за допомогою цих формул.
9. У яких випадках розподіл Бернуллі припустимо апроксимувати граничним розподілом Лапласа?
10. Опишіть уточнююче виправлення до інтегральної теореми Лапласа.

Варіанти індивідуальних завдань

№1

1. У групі студентів 84% устигаючих по вищій математиці. Для перевірки відібрали 10 студентів. Яка ймовірність того, що серед них $m=0,1,2,\dots$ устигаючих по вищій математиці? Яка ймовірність, що встигаючих по вищій математиці серед них не менш п'яти й не більше восьми? Скільки студентів треба відібрати для перевірки, щоб саме ймовірне число встигаючих по вищій математиці дорівнювало десяти? Яка ймовірність того, що встигаючих по вищій математиці не менш чотирьох?

2. Середнє число заявок, що надходять на підприємство побутового обслуговування за 1 годину, дорівнює чотирьом. Знайти ймовірність того, що в годину надійде $m=0,1,2,\dots$ заявок.

Яка ймовірність того, що за годину надійде: а) більше трьох заявок; б) менш чотирьох; в) не більше шести й не менш двох заявок?

3. Ймовірність порушення стандарту при штампуванні карболітових кілець дорівнює 0,3. Визначити ймовірність того, що з 800 готових кілець число не придатних: а) дорівнює 240; б) не менше 210; в) не більше 250; г) не менше 210 і не більше 270?

Які відхилення фактичної частоти порушення стандарту при штампуванні карболітових кілець від їх ймовірності (0,3) можна чекати із упевненістю: а) 90%; б) 95%?

Яким повинен бути обсяг вибірки n , щоб із упевненістю 95% можна було стверджувати, що відхилення фактичної частоти m/n від $p=0,3$ не перевершує: а) 2,5%; б) 5%?

№2

1. Ймовірність того, що абонент правильно набере номер телефону, дорівнює 0,9. Знайти ймовірність того, що з восьми викликів абонент набере правильно $m=0,1,2,\dots$. Знайти ймовірність того, що абонент набере правильно не менш п'яти номерів; не більше п'яти номерів; не менш трьох і не більше шести.

2. Ймовірність того, що витрата електроенергії протягом однієї доби не перевищить встановленої норми, дорівнює 0,05. Знайти ймовірність того, що в найближчий місяць витрата електроенергії не перевищить норми протягом $m=0,1,2,\dots$ доби; протягом не більше чотирьох доби; протягом не менш двох доби; протягом менш трьох доби.

3. Ймовірність появи події в кожному з 200 незалежних випробувань дорівнює 0,8. Знайти ймовірність, що подія з'явиться: а) 160 разів; б) не менш 150; в) не більше 170; г) не менш 150 і не більше 170 разів.

Які відхилення відносної частоти появи події від його ймовірності 0,8 можна гарантувати зі рівнем довіри а) 0,90; б) 0,95?

Скільки потрібно зробити випробувань, щоб з ймовірністю 95% можна було чекати відхилення відносної частоти появи події від його ймовірності 0,8 не більш, ніж: а) на 2%; б) на 3%?

№ 3

1. Імовірність виграшу по облігації 3%-ної Державної позики в деякому тиражі дорівнює 0,01. Знайти ймовірність виграшу з десяти облігацій по $m=0,1,2,\dots$ облігаціям (обчислення зробити за формулою Бернуллі). Яка ймовірність виграшу не більш ніж по двох облігаціях? Яка ймовірність виграшу не менш чим по однієї й не більш ніж по трьох облігаціях? Яка ймовірність виграшу більш ніж по двох облігаціях?

2. На телефонній станції протягом певної години дня виникає в середньому 180 викликів. Знайти ймовірність того, що за дану хвилину виникає рівно $m=0,1,2,\dots$ викликів. Яка ймовірність того, що за хвилину виникає не менш трьох викликів? Яка ймовірність того, що за хвилину виникає не менш двох і не більше шести викликів?

3. Частка ознаки в генеральній сукупності дорівнює $p=75\%$. У вибірку взято 200 одиниць. Яке серед них саме ймовірне число появ цієї ознаки? Яка ймовірність цього найімовірнішого числа?

З якою впевненістю можна затверджувати, що число одиниць із даною ознакою у вибірці виявиться: а) не менше 280; б) не більше 320; в) не менше 280 і не більше 320?

Які відхилення фактичної частоти появи ознаки у вибірці m/n від його ймовірності $p=75\%$ можна чекати з упевненістю: а) 93%; б) 95%?

Яким повинен бути обсяг вибірки n , щоб із упевненістю 95% можна було стверджувати, що відхилення фактичної частоти від ймовірності $p=75\%$ будуть не більше, ніж: а) 3%; б) 4%?

№4

1. На електроламповому заводі випускається 90% стандартних виробів, з них 80% виробів першого сорту. Знайти ймовірність того, що: 1) з десяти випадковим чином обраних ламп буде $m=0,1,2,\dots$ першого сорту; 2) хоча б одна лампа буде першого сорту; 3) не менш восьми ламп будуть першого сорту; 4) менш семи ламп будуть першого сорту.

2. Середнє число літаків, що прибувають в аеропорт за 1хв, дорівнює 2. Знайти ймовірність того, що за 2хв в аеропорт прибудуть $m=0,1,2,\dots$ літаків. Яка ймовірність того, що за 2 хв в аеропорт прибуде: а) менш чотирьох літаків; б) більше п'яти літаків; в) хоча б два літаки?

3. Завод випускає 75% електролампочок першого сорту. У вибірку взято 400 лампочок. Яке серед них саме ймовірне число лампочок першого сорту? Яка ймовірність цього найімовірнішого числа?

З якою впевненістю можна затверджувати, що число лампочок першого сорту буде: а) не менше 280; б) не більше 320; в) не менше 280 і не більше 320?

Які відхилення фактичної частоти появи лампочок першого сорту від їх ймовірності 75% можна чекати із упевненістю: а) 90%; б) 85%?

Яким повинен бути обсяг вибірки n , щоб із упевненістю 95% можна було стверджувати, що відхилення фактичної частоти m/n від 75% не перевершує: а) 4%; б) 5%?

№5

1. У ящику лежать 100 електроламп, з них: 45 стоватних, 25 шестидесятиватних і 30 сорокаватних. Навмання витягають 10 електроламп. Яка ймовірність, що з них шестидесятиватних виявиться $m=0,1,2,\dots$? Яка ймовірність того, що шестидесятиватних електроламп буде не більше 6; не менш чотирьох; не більше восьми й не менш трьох шестидесятиватних електроламп?

2. Ймовірність того, що виріб не витримає випробування, дорівнює 0,001. Знайти ймовірність того, що з 5000 виробів $m=0,1,2,\dots$ не витримають випробування; більш ніж один виріб не витримає випробування; не менш чотирьох виробів не витримають випробування; не менш трьох і не більше шести виробів не витримають випробування.

3. Відділ технічного контролю перевіряє 1000 деталей на стандартність. Ймовірність того, що деталь стандартна, дорівнює 0,95. Яке саме ймовірне число стандартних деталей? Яка ймовірність цього найімовірнішого числа?

З якою впевненістю можна стверджувати, що число стандартних деталей буде не менше 940 і не більше 960?

Знайти з ймовірністю 96% границі, у яких буде укладене число стандартних деталей серед перевірених.

Яким повинен бути обсяг вибірки, щоб із упевненістю 97% можна було стверджувати, що відхилення фактичної частоти від 95% не перевершує: а) 5%; б) 6%?

№6

1. На одному з факультетів 0,4 числа всіх студентів займаються в спортивних секціях. Для перевірки відібрали 10 студентів. Яка ймовірність, що серед них $m=0,1,2,\dots$ займаються в спортивних секціях? Яка ймовірність, що серед них студентів, що займаються в спортивних секціях: а) не менш трьох і не більше восьми; б) більше чотирьох; в) менш п'яти?

2. Середнє число кораблів, що заходять у порт за 1 годину, дорівнює трьом. Знайти ймовірність того, що за 1 годину в порт заїде $m=0,1,2,\dots$ кораблів. Яка ймовірність того, що в порт заїде: а) більше п'яти кораблів; б) менш чотирьох; в) не більше шести й не менш двох кораблів?

3. Частка ознаки в генеральній сукупності дорівнює $p=70\%$. У вибірку взято 300 одиниць. Яке серед них саме ймовірне число появ цієї ознаки? Яка ймовірність цього найімовірнішого числа?

З якою впевненістю можна затверджувати, що число одиниць із даною ознакою у вибірці виявиться: а) не менше 190; б) не більше 220; в) не менше 190 і не більше 220?

Які відхилення фактичної частоти появи ознаки у вибірці m/n від його ймовірності $p=70\%$ можна чекати з упевненістю: а) 90%; б) 95%?

Яким повинен бути обсяг вибірки n , щоб із упевненістю 96% можна було стверджувати, що відхилення фактичної частоти від ймовірності $p=70\%$ будуть не більше, ніж: а) 3%; б) 1%?

№7

1. Оптова база обслуговує 10 магазинів, від кожного з яких може надійти вимогу наступного дня з імовірністю 0,4. Визначити ймовірність того, що протягом дня база одержить $m=0,1,2,\dots$ вимог. Яка ймовірність того, що за день база одержить: а) більше чотирьох вимог; б) менш п'яти; в) не менш двох і не більше чотирьох вимог?

2. Лічильник Гейзера реєструє частки, що вилітають із деякого радіоактивного джерела, з імовірністю 0,0001. Припустимо, що за час спостереження із джерела вилетіло 30000 часток. Яка ймовірність того, що лічильник зареєстрував $m=0,1,2,\dots$ часток? Яка ймовірність того, що лічильник зареєстрував: а) більше трьох часток; б) менш п'яти часток; в) не більше семи часток і не менш трьох?

3. У великій партії валиків 85% вищого сорту. З партії взято 200 валиків. Яке серед них саме ймовірне число валиків вищого сорту? Яка ймовірність цього найімовірнішого числа?

З якою впевненістю можна затверджувати, що число валиків вищого сорту буде: а) не менш 240; б) не більше 260; в) не менше 240 і не більше 260?

Які відхилення фактичної частоти появи валиків вищого сорту від їх ймовірності 85% можна чекати із упевненістю: а) 95%; б) 93%?

Яким повинен бути обсяг вибірки, щоб із упевненістю 90% можна стверджувати, що відхилення фактичної частоти m/n від 85% будуть не більш, ніж: а) на 3%; б) на 4%?

№8

1. На автобазі є 8 машин. Ймовірність виходу на лінію кожної з них дорівнює 0,8. Знайти ймовірність виходу в найближчий день $m=0,1,2,\dots$ машин. Яка ймовірність нормальної роботи автобазі в найближчий день, якщо для цього необхідно мати на лінії не менш 6 автомашин? Обчислити ймовірність виходу на лінію не більше 5 автомашин; не менш 5 і не більше 7 автомашин.

2. Середнє число літаків, що прибувають в аеропорт за 1 хв, дорівнює 2. Знайти ймовірність того, що за 2 хв в аеропорт прибудуть $m=0,1,2,\dots$ літаків. Яка ймовірність того, що за 2 хв в аеропорт прибуде: а) менш чотирьох літаків; б) більше п'яти літаків; в) хоча б два літаки?

3. Кількість робітників, що перевиконують норму на підприємстві, дорівнює 70%. У вибірку взято 200 робітників.

З якою впевненістю можна затверджувати, що робітників, що перевиконують норму на підприємстві, буде: а) не менше 140; б) не більше 160; в) не менше 140 і не більше 160?

Які відхилення фактичної частоти від 70% можна чекати із упевненістю: а) 95%; б) 99%?

Яким повинен бути обсяг вибірки, щоб із упевненістю 90% можна було стверджувати, що відхилення фактичної частоти від 70% не перевершує: а) 5%; б) 4%.

№9

1. Схожість насіння становить 80%. Визначити ймовірність, що з 8 посіяних насінь зійде $m=0,1,2,\dots$; зійде не менш трьох насінь; зійде не більше п'яти й не менш трьох насінь; зійде менше семи насінь.

2. Середнє число заявок, що надходять на підприємство побутового обслуговування за 1 годину, дорівнює п'яти. Знайти ймовірність того, що в годину надійде $m=0,1,2,\dots$ заявок.

Яка ймовірність того, що за годину надійде: а) більше трьох заявок; б) менш чотирьох; в) не більше шести й не менш двох заявок?

3. При даному технологічному процесі 80% всієї продукції виявляється продукцією першого сорту.

Обчислити ймовірність того, що в партії з 300 виробів виявиться 250 першого сорту.

Обчислити ймовірність того, що у вибірці 300 виробів першого сорту буде не менше 230 і не більше 250.

Які відхилення фактичного відсотка продукції першого сорту у вибірці від їх частки в генеральній сукупності (80%) мають ймовірність: а) 96%; б) 94%?

Який повинен бути обсяг вибірки, щоб з ймовірністю 95% можна було стверджувати, що відхилення фактичної частоти m/n від частки ознаки (80%) будуть не більш, ніж: а) на 2%; б) на 3%?

№10

1. Ймовірність того, що телевізор зажадає ремонту протягом гарантійного строку, дорівнює 0,1. Знайти ймовірність того, що протягом гарантійного строку з 10 телевізорів вимагають ремонту $m=0,1,2,\dots$; зажадає ремонту не більше одного; хоча б один не потребує ремонту; не менш трьох зажадають ремонту.

2. Ймовірність того, що в деякім автогосподарстві одна автомашинна зазнає аварії протягом місяця, дорівнює 0,01. В автогосподарстві 500 автомашин.

Визначити ймовірність того, що протягом місяця зазнають аварії $m=0,1,2,\dots$ автомашин.

Яка ймовірність того, що за місяць зазнають аварії: а) більше трьох автомашин; б) менш дев'яти автомашин; в) не більше п'яти й не менш двох?

3. Ймовірність порушення стандарту при штампуванні карболітових кілець дорівнює 0,25. Визначити ймовірність того, що з 600 готових кілець число не придатних: а) дорівнює 125; б) не менше 100; в) не більше 150; г) не менше 125 і не більше 140?

Які відхилення фактичної частоти порушення стандарту при штампуванні карболітових кілець від їх ймовірності (0,25) можна чекати із упевненістю: а) 95%; б) 99%?

Яким повинен бути обсяг вибірки n , щоб із упевненістю 95% можна було стверджувати, що відхилення фактичної частоти m/n від $p=0,25$ не перевершує: а) 2%; б) 3%?

Лабораторна робота №5

Тема заняття:

Оцінювання параметрів розподілу одновимірної випадкової величини за даними вибіркової сукупності

Мета заняття:

Закріпити теоретичні знання щодо визначення точкових та інтервальних оцінок параметрів розподілу за незгрупованими даними; розглянути методику групування даних вибіркової сукупності та побудову гістограми й кумуляти;

навчитися використовувати такі вбудовані функції MS Excel як **МИН**, **МАКС**, **Срзнач**, **Дисп**, **Доверти**, **Стандотклон**, **Счет**, **Мода**, **Медиана**, **Ексцесс** а також процедурами **Гистограмма** та **Описательная статистика**.

Теоретичні положення:

Множина всіх значень, які приймає (або може приймати) випадкова величина X , утворює **генеральну сукупність**. Оскільки об'єкти поєднані у сукупність за певною ознакою, це передбачає, що вплив економічних факторів, дослідження яких і ставиться за мету, в межах сукупності підпорядковується єдиному закону. Якщо вивчається вплив тільки одного з факторів, то така сукупність є одновимірною. Множина значень випадкової величини, які вона приймає під час обмеженої кількості вимірювань, називається **вибірковою сукупністю** або **вибіркою**. Кількість вимірювань n становить **обсяг вибірки**. За результатами обчислення вибірових даних роблять висновок щодо кількісних характеристики випадкової величини у генеральній сукупності. Числові характеристики, що обчислені за вибіровими даними, є **статистичними оцінками** відповідних числових характеристик генеральної сукупності.

Статистичною оцінкою математичного сподівання випадкової величини X є **вибіркова середня** \bar{x} (центр вибіркової сукупності), що обчислюється як середнє виважене **варіант** x_i , кожна з яких береться з вагою, що відповідає її відносній частоті m_i :

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k x_i m_i. \quad (5.1)$$

Вибіркова середня є **незсунутою оцінкою**, тобто математичне споді-

вання вибіркової середньої дорівнює математичному сподіванню випадкової величини.

Вибіркова дисперсія D_g визначається як середнє арифметичне квадратів відхилення значень випадкової величини від її вибіркової середньої:

$$D_g = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 \cdot m_i \quad (5.2)$$

Вибіркова дисперсія є зсунутою оцінкою дисперсії випадкової величини генеральної сукупності. Для обчислення **виправленої вибіркової дисперсії** S_x^2 користуються формулою:

$$S_x^2 = \frac{n}{n-1} D_g. \quad (5.3)$$

Виправлене середнє квадратичне відхилення визначається як $S_x = \sqrt{S_x^2}$. Саме воно є оцінкою середнього квадратичного відхилення теоретичного розподілу, що характеризує розпорошення випадкової величини навколо центра сукупності.

Вибіркова середня та виправлене середнє квадратичне відхилення вибірки є **точковими оцінками** $\tilde{\Theta}$ відповідних числових характеристик Θ теоретичного розподілу, оскільки кожна з них визначена одним числом (точкою). Оскільки оцінки числових характеристик генеральної сукупності самі є випадковими величинами, тому окрім точкових розглядають ще **інтервальні оцінки**, тобто **довірчий інтервал** $(\tilde{\Theta}_1; \tilde{\Theta}_2)$, до якого з ймовірністю P влучає невідоме значення числової характеристики теоретичного розподілу. Ймовірність того, що числова характеристика теоретичного розподілу належатиме певному довірчому інтервалу, є **надійністю оцінки** γ , яка, в свою чергу, пов'язана з **рівнем значущості** α : $\alpha = 1 - \gamma$. За означенням довірчого інтервалу маємо:

$$P(\tilde{\Theta}_1 < \Theta < \tilde{\Theta}_2) = 1 - \alpha. \quad (5.4)$$

В припущенні нормального закону розподілу в генеральній сукупності маємо довірчий інтервал для математичного сподівання a :

$$\bar{x} - t_\alpha S_x / \sqrt{n} < a < \bar{x} + t_\alpha S_x / \sqrt{n}. \quad (5.5)$$

З довірчою ймовірністю $P = \gamma$ **похибка оцінки** ε , з якою математичне сподівання оцінюють за вибірковою середньою, не перевищує значення $\varepsilon = t_\alpha \cdot S_x / \sqrt{n}$.

Довірчий інтервал для середнього квадратичного відхилення теоретичного розподілу можна обчислити за формулою:

$$S_x(1 - q) < \sigma < S_x(1 + q), \quad (5.6)$$

де чинник $q = q(\gamma, n)$, який залежить від обсягу вибірки n та надійності γ , знаходиться за відповідною довідковою таблицею.

Статистичний розподіл неперервної випадкової величини можна представити графічно за допомогою гістограми. Для побудови **гістограми частот** за інтервальним рядом розподілу на кожному інтервалі, як на основі, креслять прямокутник, висота якого дорівнює m_i/h (або, відповідно, w_i/h для **гістограми частостей**). Отже, площа i -го прямокутника гістограми частот становить $h \cdot m_i/h = m_i$, звідси загальна площа всіх прямокутників цієї гістограми чисельно дорівнює обсягу вибіркової сукупності n . Відповідно загальна площа всіх прямокутників гістограми відносних частот (частостей) дорівнює одиниці. Гістограма відносних частот відображує оцінку функції щільності розподілу ймовірностей випадкової величини.

Емпіричною функцією розподілу називають функцію $F_{\text{емп.}}(x)$, яка для будь-якого наперед заданого значення x визначає відносну частоту події, яка полягає в тому, що випадкова величина прийме значення, менше за x . Згідно з цим означенням $F_{\text{емп.}}(x) = m_x/n$, де m_x – сума варіант, значення яких менші за x . Отже, емпірична функція розподілу визначається шляхом послідовного додавання відносних частот, що відповідають варіантам, меншим за x . Графіком цієї функції є **кумулята**.

Приклад 5.1. Вихідні дані, що містяться у таблиці (табл. 5.1), описують місячну продуктивність вугледобувної бригади. Ця вибірка сукупність містить 168 варіант і є репрезентативною.

Статистичні дані потрібно занести до робочого аркуша MS Excel. Оскільки статистика виконується для одновимірної вибірки, то її набирають у одному стовпчику. Систематизація вибірки, тобто підрахунок кіль-

кості варіант, що містяться у певному інтервалі (**карманах**), передбачає завдання меж цих інтервалів. Це може виконуватися автоматично (за допомогою вбудованого алгоритму) або явно (якщо навести перелік інтервалів в окремому стовпчику).

Таблиця 5.1

Довжина горизонтальної проходки (м) при вуглевидобуванні в умовах, близьких за категорією гірських робіт

56	139	54	160	133	125	196	188	93	123	148	105
216	99	197	186	119	129	145	140	63	78	124	111
402	196	199	105	105	50	331	88	141	83	130	185
118	134	72	101	125	104	243	165	77	90	115	188
53	161	117	115	76	344	58	133	142	114	143	141
121	71	147	106	89	320	61	75	150	127	176	179
161	117	132	118	193	102	122	81	107	235	223	99
70	86	108	140	119	60	169	138	112	230	244	66
137	106	151	350	55	100	124	127	222	309	116	154
147	127	91	92	197	72	103	128	178	156	66	81
112	470	182	176	141	106	115	176	301	144	159	85
126	233	135	60	72	87	107	136	250	65	149	126
290	51	101	89	116	131	200	187	118	104	137	114
203	117	59	194	138	126	120	173	184	80	107	110

При явному завданні меж інтервалів за допомогою функцій **МИН** і **МАКС** визначають розмах вибірки. Курсор встановлюємо до вільної клітини, і у списку функцій категорії **Статистические** вибираємо **МАКС**. У її діалоговому вікні у полі **Число 1** вказують діапазон вихідних даних у стовпчику. За допомогою функції **МАКС** ми отримали, що найбільше значення у вибірковій сукупності дорівнює 470. Аналогічно функція **МИН** повертає мінімальне значення. Для даного прикладу воно становить 50. Розмах визначається як різниця максимального та мінімального значень ($R = 420$). Якщо нижню межу вибіркової сукупності зсунути до $x = 0$, а верхню – до $x = 500$, то, поділивши новий розмах на 10 інтервалів, отримаємо довжину кроку $h = 50$. Межі інтервалів вказуємо в окремому стовпчику.

Для побудови гістограми можна скористатися однойменною процедурою, що міститься в **Пакете Аналіза**. Визначаємо шлях: **Сервіс** \Rightarrow **Аналіз даних** \Rightarrow **Гистограма**, та вказуємо необхідні дані в діалого-

вому вікні (рис. 5.1). Після натискання **Ок** виводиться таблиця результатів групування (табл. 5.2). Це і є емпіричний розподіл випадкової величини X . Зауважимо, що при визначенні частот влучення випадкової величини до певного інтервалу MS Excel автоматично враховує такі умови:

$$x \leq 50; 50 < x \leq 100; 100 < x \leq 150; \dots; 450 < x \leq 500; x > 500.$$

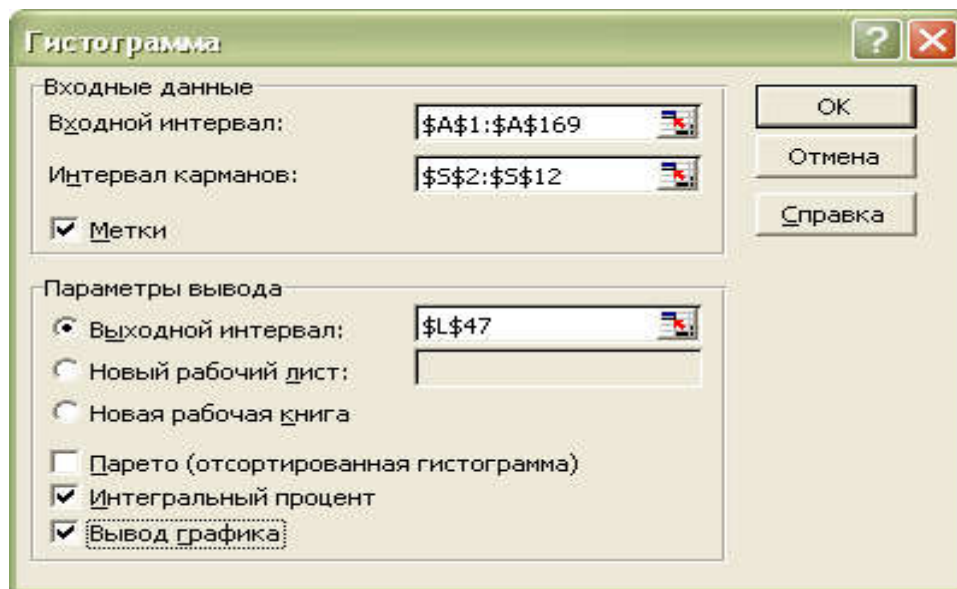


Рис. 5.1. Діалогове вікно процедури Гистограмма

Таблиця 5.2

Групування даних за визначеними інтервалами

<i>Карман</i>	<i>Частота</i>
50	2
100	40
150	77
200	30
250	10
300	1
350	6
400	0
450	1
500	1
Еще	0

Слід зазначити, що в явному завданні границь інтервалів нема потреби, оскільки процедура **Гистограмма**, що міститься в пакеті **Анализ**

данных, сама здатна їх визначити. При цьому розмах автоматично поділяється на 12 інтервалів, тобто в нашому прикладі шаг становить 37,75. Для цього у відповідному діалоговому вікні вічко **Інтервал карманов** залишаємо вільним. Натисканням клавіші **ОК** виводяться таблиця з результатами систематизації (табл. 5.3), гістограма й кумулята (рис. 5.2).

Таблиця 5.3

Результат систематизації вибіркової сукупності

№ Інтервала	Карман	Частота
1	50	1
2	85	28
3	120	48
4	155	44
5	190	20
6	225	12
7	260	6
8	295	1
9	330	3
10	365	3
11	400	0
12	435	1
13	Еще	1

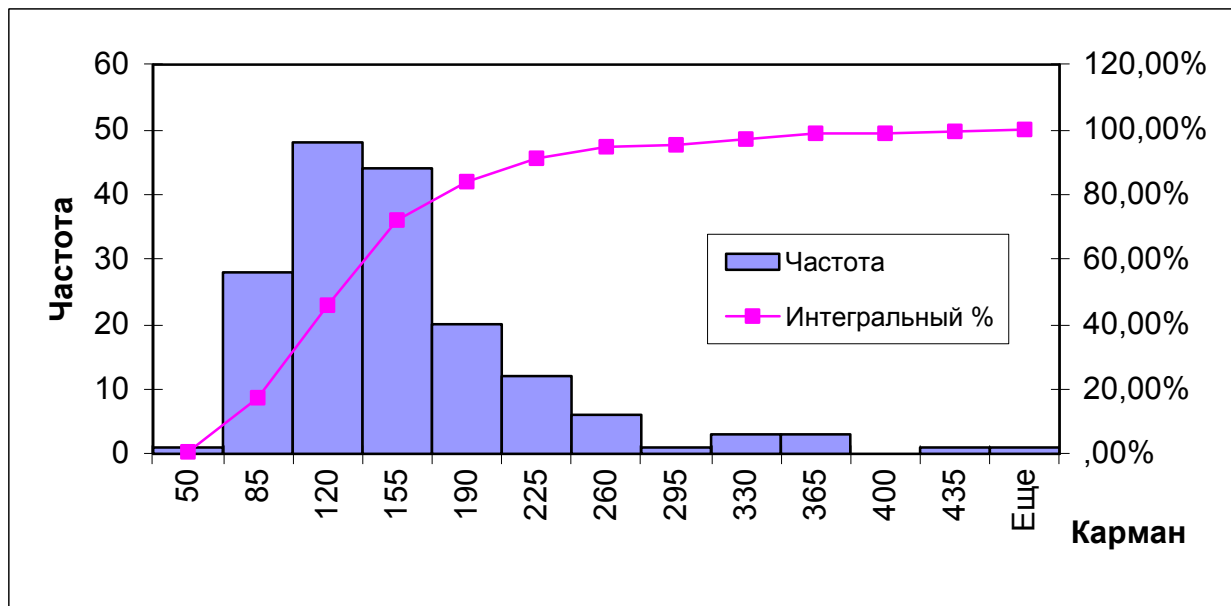


Рис. 5.2. Статистична обробка вихідних даних за допомогою процедури Гистограмма

У першому стовпчику табл. 5.3, що має назву **Карман** містяться значення верхньої межі інтервалу. Видно, що автоматично було вибрано шаг 37,75. Другий стовпчик, що має назву **Частота**, містить кількості влучень значень випадкової величини у вибраний інтервал. За даними перших двох стовпців побудована гістограма. Третій стовпець, що має назву **Інтегральний %**, містить кількість значень (у відсотках) випадкової величини, що менші за нижню межу відповідного інтервалу, і за цими даними побудована кумулята.

Слід зауважити, що гістограма, яка виводиться після виконання відповідної процедури, не відповідає двом вимогам. По-перше, ширина стовпців менша за шаг. Цей недолік виправляється за допомогою команди **Формат ряда данных** (рис. 5.3), де у відповідному діалоговому вікні знаходимо **Параметры** та вказуємо, що **Ширина зазора** є нуль.

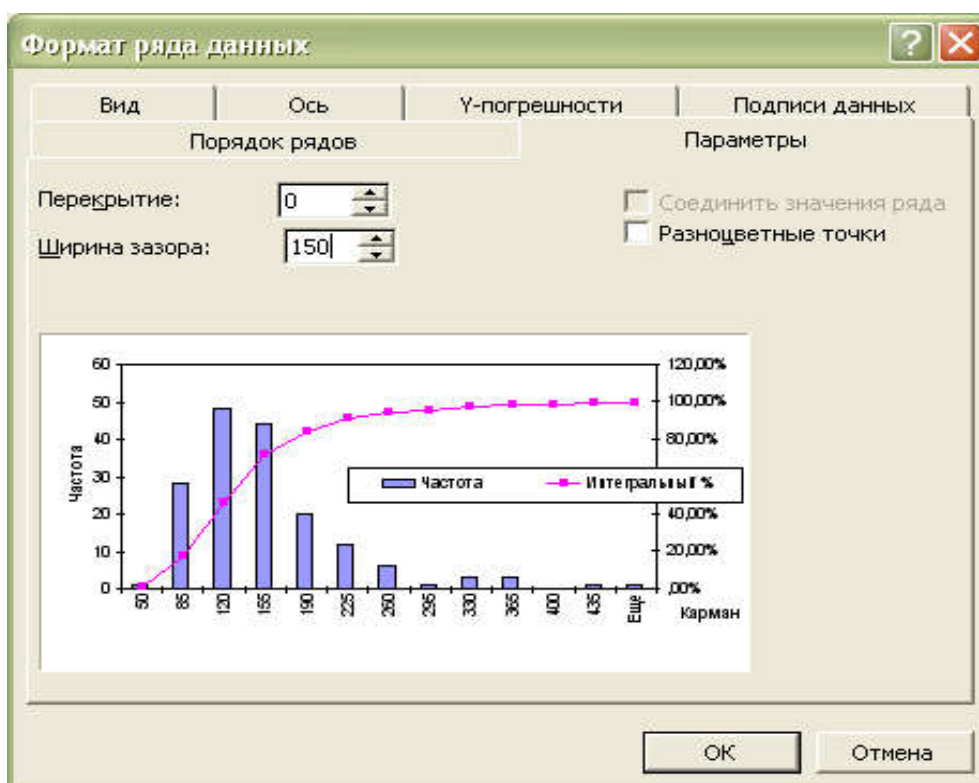


Рис. 5.3. Діалогове вікно **Формат ряда данных**

По-друге, площа під гістограмою частот повинна чисельно дорівнювати обсягу вибірки. При побудові гістограми за допомогою процедури **Гистограмма** ця умова порушується. Для того, щоб умова нормування виконувалася, необхідно замість частот використовувати відношення ча-

стот до довжини інтервалу (або до шагу, якщо інтервали є рівними за довжиною). Це можна здійснити за допомогою додаткових обчислень.

Оскільки гістограма має один максимум, а також частоти монотонно спадають як при збільшенні, так і при зменшенні значення випадкової величини відносно максимуму, то вибірккову сукупність можна вважати такою, що наближена до однорідної, та перейти до оцінки параметрів вибіркової сукупності, тобто до описової статистики.

Електронні таблиці MS Excel дозволяють визначати числові характеристики емпіричного розподілу кількома способами, наприклад, використовуючи функцій MS Excel, за допомогою процедури **Описательная статистика**, або шляхом організації відповідних розрахунків. Визначаємо шлях: **Сервис** ⇒ **Анализ данных** ⇒ **Описательная статистика**. Після заповнення діалогового вікна на робочому аркуші у вихідному інтервалі виводиться таблиця з результатами обчислення статистичних параметрів та їх назвами (табл. 5.4).

Таблиця 5.4

**Підсумкова таблиця, що відповідає виконанню процедури
Описательная статистика**

Столбец 1		Коментар
Среднее	139,7381	Среднее
Стандартная ошибка	5,219274	Стандартная ошибка
Медиана	125,5	Медиана
Мода	118	Мода
Стандартное отклонение	67,64952	Стандартное отклонение
Дисперсия выборки	4576,458	Дисперсия выборки
Эксцесс	4,827949	Эксцесс
Асимметричность	1,828735	Асимметричность
Интервал	420	Интервал
Минимум	50	Минимум
Максимум	470	Максимум
Сумма	23476	Сумма
Счет	168	Счет

Результати обчислення, що виводяться на екран, містяться в перших двох стовпцях табл. 5.4, у третьому ми додали коментар, що пояснює зміст оцінок. Кожну з характеристик окремо можна вивести за допомогою вбудованих функцій. Пропонуємо самостійно визначити оцінки чи-

слових характеристик розподілу за допомогою вбудованих функцій **Срзнач, Дисп, Доверти, Стандотклон, Медиана, Мода, Счет, Ексцесс** і порівняти отримані дані з результатами застосування процедури **Описательная статистика**. За формулами 5.5 та 5.6 визначити довірчий інтервал, до якого відповідні числові характеристики генеральної сукупності влучають з надійністю $\gamma = 95\%$.

Варіанти для самостійного розв'язання

Вибіркова сукупність містить відомості про середню вартість однієї робочої години (в грн) робітників-верстатників механічного цеха протягом поточного року:

- 5.1. 6,12 6,25 5,88 6,20 6,20 6,95 6,82 6,15 6,23 6,70 5,75 6,30
6,30 6,45 6,20 6,40 6,30 6,20 6,31 6,25 6,45 6,20 6,50 6,25
6,50 6,40 6,25 6,35 6,08 6,25 6,01 6,28 6,07 6,30 6,45 6,12
5,91 6,20 6,00 6,25 6,20 6,40 6,01 6,15 6,40 6,25 6,20 6,12
6,70 6,10 5,98 6,20 5,89 6,13 6,21 6,35 6,14 6,40 6,25 6,10
6,21 6,25 6,50 6,60 6,35 6,40 6,30 6,35 6,25 6,10 6,15 6,75
6,20 6,25 6,08 6,20 6,15 6,30 5,84 6,47 6,43 6,45 6,35 6,40
5,95 6,20 6,50 5,60 6,21 6,21 6,15 5,98 6,20 6,21 6,15 6,00
6,50 6,15 6,70 6,10.
- 5.2. 5,5 5,9 5,8 5,7 6,0 5,5 5,0 6,5 5,7 5,6 6,2 5,4 5,4 5,0 5,7 5,5
6,2 5,4 5,6 4,2 5,6 5,4 5,3 4,7 4,5 5,0 5,1 5,5 5,6 5,1 6,3 5,4
5,9 5,6 6,4 5,7 5,0 5,9 5,7 6,0 6,7 4,1 6,2 5,8 5,5 6,2 6,2 5,8
5,8 5,6 5,5 4,7 5,8 4,9 6,1 6,2 6,0 4,3 5,9 6,9 4,5 6,8 5,6 5,3
4,7 5,5 5,0 5,3 5,2 4,8 5,5 5,9 6,5 5,2 5,7 6,0 5,4 5,8 5,8 5,6
4,8 5,8 6,2 5,3 4,5 5,8 5,6 5,4 6,4 6,0 6,1 5,9 6,4 6,3 5,4 5,8
6,1 6,3 5,7 4,7.
- 5.3. 4,5 6,1 5,7 5,3 5,8 5,5 5,7 5,5 5,7 5,9 5,7 4,6 5,1 6,0 6,4 6,4
5,1 6,5 4,7 5,6 6,3 6,5 5,0 5,4 5,8 6,3 4,9 5,3 5,3 6,0 5,7 5,4
6,0 5,5 6,4 5,7 4,8 5,5 5,7 6,6 5,0 5,8 7,0 5,9 4,7 6,2 5,9 6,2
5,0 5,5 5,5 5,1 5,8 6,9 5,7 5,7 6,2 5,3 4,3 5,9 5,8 5,7 5,3 4,3
6,7 6,7 5,6 6,1 5,9 5,2 5,0 4,9 5,6 6,0 5,8 5,2 5,9 5,5 6,0 5,3
5,9 6,3 6,3 5,2 5,4 6,5 6,5 5,4 5,6 5,8 6,3 7,3 4,7 7,1 6,3 5,6
5,6 5,5 5,5 5,7.
- 5.4. 4,0 5,8 5,4 4,9 4,2 4,7 4,5 6,4 7,3 5,5 6,7 5,9 4,0 6,8 5,0 4,7
6,5 6,0 6,3 6,5 6,0 4,6 7,9 4,3 7,4 7,1 7,5 4,6 6,8 7,3 6,9 5,4
5,7 4,3 7,5 4,4 4,5 4,9 5,0 5,2 5,5 5,9 4,5 5,4 4,5 6,5 7,3 6,1
7,3 5,7 6,2 6,6 7,1 7,5 3,9 6,7 6,6 7,0 6,8 6,4 5,4 4,4 4,8 3,9
5,3 5,5 5,8 6,6 5,6 5,0 4,8 4,7 7,2 5,1 6,5 5,3 5,2 4,2 7,0 6,4

3,9 4,4 6,2 4,1 6,4 5,0 6,7 6,0 5,4 5,6 5,8 6,3 4,5 4,9 4,0 4,1
 7,5 6,1 5,3 5,1.
 5.5. 3,5 3,7 3,8 3,7 3,6 3,6 3,5 3,5 3,6 3,6 3,6 3,6 4,0 3,5 4,0 3,6
 3,6 3,8 4,2 3,5 3,7 3,8 3,7 3,6 3,5 4,3 3,8 3,5 3,7 3,9 3,9 3,5
 4,3 3,8 4,5 3,9 3,5 3,5 4,1 3,8 3,5 4,3 3,7 3,8 3,5 3,7 3,9 3,5
 3,5 3,6 3,8 4,7 3,7 3,7 3,9 3,5 3,8 3,5 4,0 3,6 3,9 3,5 4,2 3,7
 3,9 3,7 4,0 3,7 3,5 4,8 3,5 4,2 3,5 3,8 3,5 4,4 3,9 4,6 3,6 3,5
 3,5 3,8 3,6 3,7 3,9 3,6 3,5 4,0 3,5 3,9 3,5 3,8 3,5 3,8 4,4 3,6
 3,5 3,7 3,5 4,0.
 5.6. 6,7 5,3 7,5 4,4 4,5 4,9 5,0 5,2 5,5 5,9 4,5 5,4 4,5 6,5 7,3 6,1
 6,3 5,7 6,2 6,6 7,1 7,5 3,9 6,7 6,6 7,0 6,8 6,4 5,4 4,4 4,8 3,9
 5,8 5,5 5,8 6,6 5,6 5,0 4,8 4,7 7,2 5,1 6,5 5,3 5,2 4,2 7,0 6,4
 3,9 4,4 6,2 4,1 6,4 5,0 6,7 6,0 5,4 5,6 5,8 6,3 4,5 4,9 4,0 4,1
 5,5 6,0 6,3 6,5 6,0 4,6 7,9 4,3 7,4 7,1 7,5 4,6 6,8 7,3 6,9 5,4
 7,1 6,3 7,5 4,9 4,5 4,9 5,0 5,2 5,5 5,9 4,5 5,4 4,5 6,5 7,3 6,1
 5,3 5,7 6,2 4,8.
 5.7. 4,5 3,7 3,8 4,7 3,6 4,6 4,5 4,5 3,6 3,6 3,6 3,6 4,0 3,5 4,0 3,6
 4,6 4,8 5,2 3,5 3,7 3,8 3,7 4,6 3,5 4,3 3,8 3,5 3,7 4,9 3,9 4,5
 4,3 4,8 4,5 3,9 3,5 3,5 5,1 3,8 4,5 4,3 3,7 3,8 4,5 4,7 3,9 3,5
 4,5 4,6 3,8 4,7 4,7 3,7 4,9 3,5 3,8 4,5 4,0 4,6 4,9 3,5 4,2 3,7
 4,9 4,7 5,0 3,7 4,5 4,8 3,5 4,2 4,5 3,8 3,5 4,4 3,9 4,6 3,6 3,5
 4,7 3,8 3,6 4,7 3,9 4,6 4,5 5,0 4,5 3,9 4,5 4,8 4,5 3,8 4,4 3,6
 4,5 3,7 4,5 4,0.
 5.8. 5,5 4,7 5,8 5,7 5,6 5,6 5,5 4,5 4,6 4,6 4,6 3,6 4,0 3,5 4,0 4,6
 5,6 5,8 4,2 4,5 4,7 5,8 3,7 4,6 4,5 4,3 3,8 4,5 3,7 4,9 3,9 4,8
 4,3 4,8 4,5 3,9 4,5 4,5 4,1 3,8 5,5 4,3 4,7 4,8 3,5 3,7 3,9 4,5
 5,5 5,6 4,8 4,7 4,7 3,7 4,9 4,5 5,8 5,5 4,0 3,6 4,9 3,5 4,2 4,7
 4,9 4,7 4,0 5,7 5,5 4,8 4,5 4,2 4,5 4,8 3,5 4,4 3,9 4,6 3,6 4,5
 4,5 3,8 4,6 4,7 4,9 4,6 5,5 5,0 4,5 4,9 3,5 4,8 5,5 4,8 4,4 4,6
 4,5 4,7 3,5 4,0.
 5.9. 3,9 3,7 3,8 3,7 3,6 4,6 3,5 4,5 3,6 4,6 3,6 3,6 4,0 3,5 4,0 3,6
 4,6 3,8 4,7 4,5 3,9 3,8 4,7 3,6 3,5 4,3 4,8 3,5 3,7 3,9 3,9 3,5
 4,3 3,8 4,5 3,9 3,5 3,5 4,1 3,8 3,5 4,3 3,7 3,8 3,5 3,7 3,9 3,5
 4,3 3,6 3,8 4,7 3,7 3,7 3,9 3,5 4,8 4,5 5,0 3,6 3,9 3,5 4,2 3,7
 3,9 4,7 4,2 4,7 4,5 4,8 3,5 4,2 3,5 3,8 3,5 4,4 3,9 4,6 3,6 3,5
 4,7 3,8 4,6 3,7 3,9 3,6 3,5 5,0 4,5 3,9 4,5 3,8 5,5 3,8 4,4 3,6
 4,5 4,7 3,5 4,0.
 5.10. 5,5 5,9 5,8 5,7 6,0 5,5 5,0 6,5 5,7 5,6 6,2 5,4 5,4 5,0 5,7 5,5
 6,2 5,4 5,6 4,2 5,6 5,4 5,3 4,7 4,5 5,0 5,1 5,5 5,6 5,1 6,3 5,4
 5,9 5,6 6,4 5,7 5,0 5,9 5,7 6,0 6,7 4,1 6,2 5,8 5,5 6,2 6,2 5,8
 5,8 5,6 5,5 4,7 5,8 4,9 6,1 6,2 6,0 4,3 5,9 6,9 4,5 6,8 5,6 5,3
 4,7 5,5 5,0 5,3 5,2 4,8 5,5 5,9 6,5 5,2 5,7 6,0 5,4 5,8 5,8 5,6
 4,8 5,8 6,2 5,3 4,5 5,8 5,6 5,4 6,4 6,0 6,1 5,9 6,4 6,3 5,4 5,8
 6,1 6,3 5,7 4,7.

Контрольні питання:

1. Означення генеральної та вибіркової сукупностей. Визначити вимоги, яким повинна відповідати вибірка сукупність.
2. Статистичні розподіли: означення, геометричні надання.
3. Точкові оцінки основних числових характеристик розподілу та вимоги, яким вони повинні відповідати.
4. Зсунуті та незсунуті статистичні оцінки. Виправлена дисперсія.
5. Інтервальні оцінки невідомих параметрів розподілу: означення, постановка задачі.
6. Надійний інтервал для математичного сподівання та середнього квадратичного відхилення випадкової величини, що розподілена за нормальним законом.
7. Оцінки додаткових числових характеристик: мода, медіана, коефіцієнт варіації.
8. Оцінювання додаткових числових характеристик розподілу за допомогою гістограми або полігона та кумуляти.
9. Емпіричні моменти. Використання початкових та центральних моментів для оцінювання числових характеристик розподілу.
10. Емпірична функція розподілу та її основні властивості.
11. Незгруповані та згруповані емпіричні дані. Інтервальний варіаційний ряд. Вплив групування на вибіркові числові характеристики.
12. Обчислення обсягу вибіркової сукупності, що з наперед заданою точністю та надійністю дозволить визначити математичне сподівання нормально розподіленої випадкової величини.

Лабораторна робота №6

Тема заняття: Перевірка гіпотези про нормальний розподіл генеральної сукупності за критерієм Пірсона.

Мета заняття: Навчитися перевіряти нульову гіпотезу відносно теоретичного закону розподілу за допомогою критерію узгодженості Пірсона, використовувати такі вбудовані функції MS Excel як **НОРМРАСП**, **ХИ2РАСП**.

Теоретичні положення:

Однією з найважливіших задач математичної статистики є визначення теоретичного закону розподілу випадкової величини, яка характеризує вивчаєму ознаку за емпіричним розподілом, що представляє варіаційний ряд. Для розв'язку цієї задачі необхідно визначити вигляд і параметри закону розподілу. В загальному випадку закон розподілу випадкової величини в генеральній сукупності невідомий до початку досліджень, а певні припущення відносно його характеру можна зробити по вигляду гістограми.

Після того, як будуть обчислені числові характеристики вибірки, можна обчислити **вирівнюючі частоти** \tilde{m}_i , які будуть відрізнятися від емпіричних частот m_i даної вибірки.

Нульова гіпотеза H_0 полягає в тому, відхилення емпіричних частот даної вибіркової сукупності від вирівнюючих обумовлене випадковим розпорошенням. Перевірка нульової гіпотези і відповідно припущення відносно теоретичного закону розподілу здійснюється за допомогою критеріїв узгодженості.

Найбільш поширеним з них є **критерій Пірсона** або **критерій** χ^2 . Зручність цього критерію полягає в тому, що його можна використовувати не тільки для перевірки гіпотези про відповідність нормальному закону розподілу, але і при інших законах розподілу у генеральній сукупності.

Критерій узгодженості Пірсона є випадковою величиною, що має розподіл χ^2 , який визначається за формулою

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i},$$

і має $s = k - r - 1$ ступенів свободи,

де k – число часткових інтервалів інтервального статистичного розподілу вибірки;

r – число параметрів, якими визначається закон розподілу ймовірностей генеральної сукупності згідно з нульовою гіпотезою.

Так, наприклад, для нормального закону $r = 2$, оскільки він визначається двома параметрами: a – математичне сподівання та σ – середнє квадратичне відхилення. Для рівномірного закону теж $r = 2$, а саме: a – початок та b – кінець інтервалу, до якого належать значення випадкової величини. Однак для показникового закону $r = 1$, оскільки $M(X) = \sigma$, і роль параметру відіграє параметр $\lambda = \frac{1}{M(X)}$.

Для визначення критичної області критерію Пірсона прийняті два рівня значущості: 0,05 та 0,01. Якщо емпіричне значення χ^2 менш ніж критичне значення, що визначено при $\alpha = 0,05$ (позначається $\chi_{0,05}^2$), то нульова гіпотеза про випадковість розбіжностей між вирівнюючими та емпіричними частотами приймається. Навпаки, якщо емпіричне значення критерію χ^2 перевищує критичне, яке визначене при $\alpha = 0,01$ ($\chi_{0,01}^2$), то нульову гіпотезу треба відкинути. Якщо емпіричне значення критерію χ^2 належить до області невизначеності $\chi_{0,05}^2 < \chi^2 < \chi_{0,01}^2$, то прийняти чи не прийняти нульову гіпотезу залежить від дослідника.

Слід пам'ятати, що відкидаючи вірну нульову гіпотезу, ми робимо **помилку першого роду**, а приймаючи хибну гіпотезу, робимо **помилку другого роду**, яка в загальному випадку більш небезпечна.

Отже $\chi^2 < \chi_{0,05}^2$ є довірчим інтервалом для критерію Пірсона. Розширення довірчого інтервалу супроводжується збільшенням ймовірності саме помилки другого роду. Для визначення статистики χ^2 зручно скласти розрахункову таблицю за зразком (табл. 6.1). Таблиця 6.1 заповнюється таким чином:

стовпець 1 – номер відповідного інтервалу;

стовпець 2 – частота;

стовпець 3 – середина відповідного інтервалу;

стовпець 4 – значення стандартизованої випадкової величини t'_i ;

стовпець 5 – значення диференціальної функції Лапласа, аргументом якої є t'_i ;

стовпець 6 – значення функції $\tilde{f}_i = \frac{\varphi(t'_i)}{\sigma}$;

стовпець 7 – ймовірності \tilde{P}_i влучення випадкової величини до i -го інтервалу. Контрольна сума цього стовпця не повинна відрізнятись від одиниці більш, ніж на 5%;

стовпець 8 – вирівнюючі частоти, обчислюються за формулою $\tilde{m}_i = n \cdot \tilde{P}_i$;

стовпець 9 – частка квадрата відхилення емпіричної частоти від вирівнюючої у відношенні до вирівнюючої. Емпіричне значення критерію Пірсона обчислюється як сума рядків цього стовпця $\sum_{i=1}^k \chi_i^2 = \sum_{i=1}^k \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}$.

Таблиця 6.1

Визначення вирівнюючих частот розподілу випадкової величини з використанням функції Гауса у припущенні нормального закону

№	m_i	x'_i	$t'_i = \frac{x'_i - \bar{x}}{\sigma}$	$\varphi(t'_i)$	$\tilde{f}_i = \frac{\varphi(t'_i)}{\sigma}$	$\tilde{P}_i = h_i \tilde{f}_i$	$\tilde{m}_i = n \cdot \tilde{P}_i$	$\chi_i^2 = \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}$
1	2	3	4	5	6	7	8	9

Вирівнюючі частоти можна також обчислити за допомогою інтегральної функції розподілу. Для цього також зручно використовувати таблицю, зразок якої наведено (табл. 6.2).

Таблиця 6.2

Визначення вирівнюючих частот розподілу випадкової величини з використанням інтегральної функції розподілу у припущенні нормального закону

№	m_i	x'_i	$\tau_i = \frac{x'_i - \bar{x}}{\sigma}$	$\Phi(\tau_i)$	$\tilde{F}(x_i) = \frac{1}{2} + \Phi(\tau_i)$	$\tilde{P}_i = \tilde{F}_{i+1} - \tilde{F}_i$	$\tilde{m}_i = n \cdot \tilde{P}_i$	$\chi_i^2 = \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}$
1	2	3	4	5	6	7	8	9

Таблиця 2 заповнюється таким чином:

стовпець 1 – номер відповідного інтервалу;

стовпець 2 – частота;

стовпець 3 – ліва межа відповідного інтервалу;

стовпець 4 – значення стандартизованої випадкової величини τ_i ;

стовпець 5 – значення інтегральної функції Лапласа, аргументом якої є τ_i . Обчислюються за довідковою таблицею.

стовпець 6 – значення функції $\tilde{F}(x_i) = \frac{1}{2} + \Phi(\tau_i)$;

стовпець 7 – ймовірності \tilde{P}_i влучення випадкової величини до i -го інтервалу. Контрольна сума цього стовпця не повинна відрізнятись від одиниці більш, ніж на 5%;

стовпець 8 – вирівнюючі частоти, обчислюються за формулою $\tilde{m}_i = n \cdot \tilde{P}_i$;

стовпець 9 – частка квадрата відхилення емпіричної частоти від вирівнюючої у відношенні до вирівнюючої. Емпіричне значення критерію Пірсона обчислюється як сума рядків цього стовпця

$$\sum_{i=1}^k \chi_i^2 = \sum_{i=1}^k \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}.$$

Приклад 6.1. Для наданого емпіричного розподілу

$[x_i, x_{i+1})$	2,7-3,3	3,3-3,9	3,9-4,5	4,5-5,1	5,1-5,7	5,7-6,3	6,3-6,9	6,9-7,5	7,5-8,1
m_i	1	7	17	24	23	16	9	2	1

Підібрати відповідний теоретичний закон і на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про узгодженість двох розподілів за допомогою критерію χ^2 .

Розв'язання.

Для побудови гістограми можна скористатися однойменною процедурою, що міститься в **Пакеті Аналіза**. Визначаємо шлях: **Сервіс** \Rightarrow **Анализ данных** \Rightarrow **Гистограмма**, та вказуємо необхідні дані в діалоговому вікні (рис. 6.1).

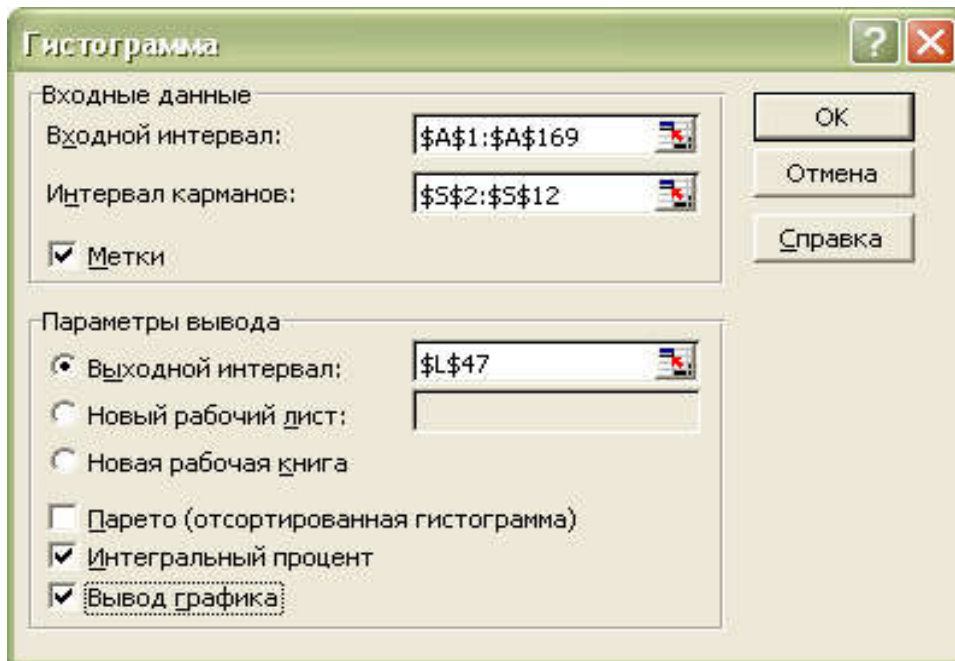
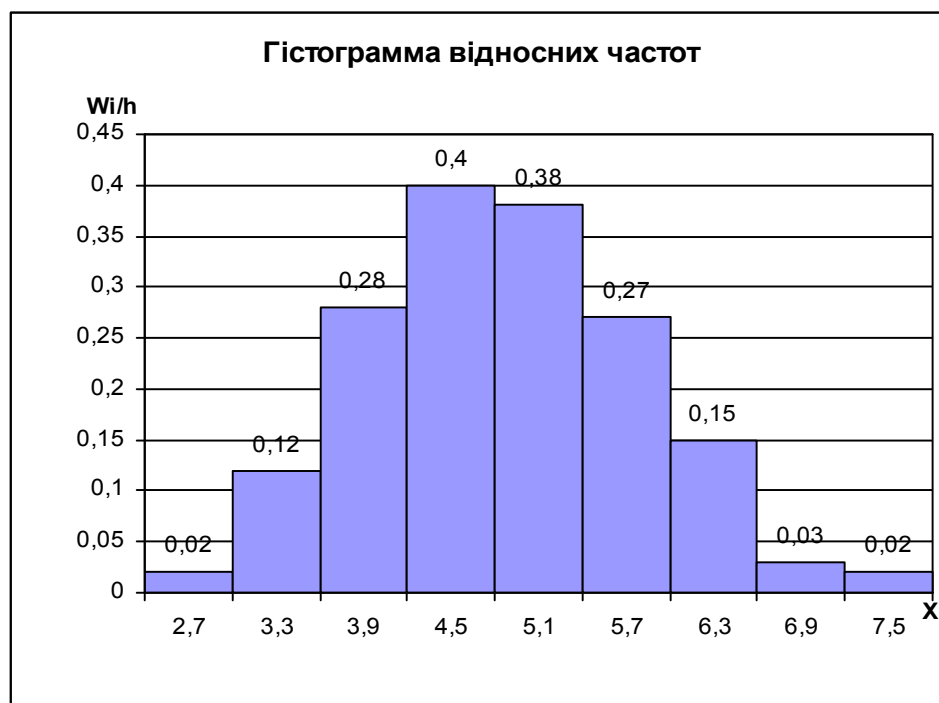


Рис. 6.1. **Діалогове вікно процедури Гистограмма**

За виглядом гістограми розподілу можна припустити нормальний закон.



За допомогою функцій **СРЗНАЧ** та **СТАНДОТКЛОН** обчислюємо вибіркове середнє \bar{x} та середнє квадратичне відхилення σ . Так як кількість спостережень $n = 100$, то замість виправленого середньоквадратич-

ного відхилення можна використати «звичайне» середнє квадратичне відхилення σ . $\bar{x} = 5.166$, $\sigma = 0.93$. Для визначення статистики χ^2 зручно скласти розрахункову таблицю (табл. 6.3).

Таблиця 6.3

Визначення вирівнюючих частот розподілу випадкової величини у припущенні нормального закону

№	x_i'	m_i	$t_i' = \frac{x_i' - \bar{x}}{\sigma}$	$\varphi(t_i')$	$\tilde{f}_i = \frac{\varphi(t_i')}{\sigma}$	$\tilde{P}_i = h_i \tilde{f}_i$	$\tilde{m}_i = n \cdot \tilde{P}_i$	$\chi_i^2 = \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}$
1	2	3	4	5	6	7	8	9
1	3,0	1	-2,32	0,0270	0,029	0,0174	1,74	0,3147
2	3,6	7	-1,68	0,0973	0,104	0,0624	6,24	0,0926
3	4,2	17	-1,04	0,2323	0,249	0,1494	14,94	0,2840
4	4,8	24	-0,39	0,3697	0,396	0,2376	23,76	0,0024
5	5,4	23	0,25	0,3867	0,414	0,2484	24,84	0,1363
6	6,0	16	0,89	0,2685	0,288	0,1728	17,28	0,0948
7	6,6	9	1,54	0,1219	0,131	0,0786	7,86	0,1653
8	7,2	2	2,18	0,0371	0,040	0,024	2,4	0,0667
9	7,8	1	2,82	0,0075	0,008	0,0048	0,48	0,5633
		$\sum_{i=1}^9 m_i =$				$\sum_{i=1}^9 \tilde{P}_i = 0.99$		$\sum_{i=1}^9 \chi^2 = 1,7202$

Таким чином маємо, що $\chi_{\alpha; i}^2 = 1,7202$. Визнаємо число ступенів свободи $df = k - r - 1 = 9 - 3 - 1 = 5$ і за довідковою таблицею знаходимо критичне значення критерію Пірсона, що відповідає рівню значущості 0,05.

Маємо $\chi^2(0.05; 5) = 11.1$ емпіричне значення χ^2 менш, ніж теоретичне і гіпотезу про нормальний розподіл генеральної сукупності треба прийняти. Тобто розбіжність між емпіричними і вирівнюючими частотами, коли останні визначені у припущенні нормального розподілу випадкової величини, можна пояснити випадковим розпорощенням.

Обчислимо вирівнюючі частоти за допомогою інтегральної функції розподілу. Для цього зручно використовувати таблицю (табл. 6.4).

Таблиця 6.4

**Визначення вирівнюючих частот розподілу випадкової величини
з використанням інтегральної функції розподілу
у припущенні нормального закону**

Інтер- вали	m_i	i	x'_i	$\tau_i = \frac{x'_i - \bar{x}_g}{\sigma}$	$\Phi(\tau_i)$	$F_i = \frac{1}{2} + \Phi(\tau_i)$	$\tilde{P}_i = F_{i+1} - F_i$	$\tilde{n}_i = n \cdot \tilde{P}_i$	$n_i - \tilde{n}_i$	$\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$	
		1	$-\infty$	$-\infty$	-0,5	0	0,00415	0,415	-0,415	0,415	
2,7 – 3,3	1	8	2	2,7	-2,64	-0,49585	0,00415	0,08276	8,276	-0,276	0,009
3,3 – 3,9	7										
3,9 – 4,5	17	3	3,9	-1,36	-0,41309	0,08691	0,15194	15,194	1,806	0,215	
4,5 – 5,1	24	4	4,5	-0,71	-0,26115	0,23885	0,23325	23,325	0,675	0,020	
5,1 – 5,7	23	5	5,1	-0,07	-0,02790	0,4721	0,24356	24,356	-1,356	0,075	
5,7 – 6,3	16	6	5,7	0,57	0,21566	0,71566	0,17311	17,311	-1,311	0,099	
6,3 – 6,9	9	7	6,3	1,22	0,38877	0,88877	0,07979	7,979	1,021	0,131	
6,9 – 7,5	2	3	8	6,9	1,86	0,46856	0,96856	0,03144	3,144	-1,144	0,416
7,5 – 8,1	1										
		9	$+\infty$	$+\infty$	0,5	1					
$\sum_i \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = \chi^2_{\text{расчетное}}$										1,38	

Враховуючи, що в даному емпіричному розподілі частоти першого і останнього інтервалів ($n_1=1, n_9=1$) менше 5, при використанні критерію Пірсона доцільно об'єднати вказані інтервали з сусідніми (див. табл. 6.4). Таким чином маємо, що $\chi^2_{\text{аіі}}=1,38$. Визнаємо число ступенів свободи $df = k - r - 1 = 9 - 3 - 1 = 5$ і за таблицею знаходимо критичне значення критерію Пірсона, що відповідає рівню значущості 0,05.

Маємо $\chi^2(0.05;5)=11.1$ і емпіричне значення χ^2 менш, ніж теоретичне. Отже, нема підстав відкинути гіпотезу про нормальний розподіл генеральної сукупності. В цьому випадку емпіричні і теоретичні частоти розрізняються не значущо.

Зауваження. Значення диференціальної та інтегральної функції нормального розподілу можна обчислити за допомогою вбудованої функції **НОРМРАСП**. Емпіричне значення критерію Пірсона χ^2 для кожного значення змінної x_i і визначеного числа ступенів свободи можна обчислити за допомогою вбудованої функції **ХИ2РАСП**. Пропонуємо самостійно це зробити і порівняти отримані дані з результатами в розрахункових таблицях 3 і 4.

Варіанти для самостійного розв'язання

За заданими статистичними розподілами вибірки висунути гіпотезу про закон розподілу ознаки генеральної сукупності і при рівні значущості $\alpha = 0.05$ перевірити її правильність:

1. Вимірювалась маса новонароджених дітей у пологовому будинку. Результати вимірювань подано інтервальним статистичним розподілом:

x_i , кг	1,0- 1,2	1,2- 1,4	1,4- 1,6	1,6- 1,8	1,8- 2,0	2,0- 2,2	2,2- 2,4	2,4- 2,6	2,6- 2,8	2,8- 3,0	3,0- 3,2
n_i	5	12	18	22	36	24	19	15	11	9	2

2. Вимірювався місячний дохід робітника певної галузі. Результати вимірювання подано інтервальним статистичним розподілом:

x_i , грн.	100- 120	120- 140	140- 160	160- 180	180- 200	200- 220	220- 240	240- 260	260- 280
n_i	10	15	20	25	30	40	10	4	2

3. Вимірювання зросту юнаків віком 18-20 років наведено у вигляді інтервального статистичного розподілу:

x_i , см	154- 160	160- 166	166- 172	172- 178	178- 184	184- 190	190- 196	196- 202
n_i	8	20	30	42	34	21	9	2

4. Вимірювався знос автошин за місяць. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i , мм	0- 0,02	0,02- 0,04	0,04- 0,06	0,06- 0,08	0,08- 0,1	0,1- 0,12	0,12- 0,14	0,14- 0,16
n_i	48	42	34	26	18	10	6	4

5. Вимірювався час неперервного горіння електролампочок, виготовлених фірмою, до виходу їх з ладу. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i , ГОД.	22- 22,5	22,5- 23	23- 23,5	23,5- 24	24- 24,5	24,5- 25	25- 25,5	25,5- 26	26,- 26,5	26,5- 27	27- 27,5
n_i	4	12	16	24	36	28	22	18	16	8	4

6. Вимірювалась кількість опадів, які випали протягом веснянопольових робіт у північних регіонах України. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i , мм	440- 452	452- 464	464- 476	476- 488	488- 500	500- 512	512- 524	524- 536	536- 548	548- 560
n_i	24	18	16	14	12	10	8	6	4	2

7. Вимірювалась плинність кадрів у відсотках на підприємствах певної галузі виробництва протягом року. Результати вимірювання показано як інтервальний статистичний розподіл:

x_i , %	0,12- 0,18	0,18- 0,24	0,24- 0,30	0,3- 0,36	0,36- 0,42	0,42- 0,48	0,48- 0,54	0,54- 0,6	0,6- 0,66
n_i	46	38	32	28	24	18	16	8	6

8. Проводився експеримент з вимірювання потужності відбиття від поверхні моря сигналу на виході радіолокаційного приймача. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i ,	0-	0,05-	0,1-	0,15-	0,2-	0,25-	0,3-	0,35-	0,4-
ВТ	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45
n_i	88	64	58	42	30	22	18	6	4

9. Вимірювалась жирність молока корів x_i із навмання вибраної ферми. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i ,	3,45-	3,55-	3,65-	3,75-	3,85-	3,95-	4,05-	4,15-	4,25-	4,35-
%	3,55	3,65	3,75	3,85	3,95	4,05	4,15	4,25	4,35	4,45
n_i	10	16	22	30	34	20	14	10	6	4

10. Результати вимірювання граничного навантаження на сталевий болт наведено інтервальним статистичним розподілом:

x_i ,	4,5-	5,5-	6,5-	7,5-	8,5-	9,5-	10,5-	11,5-	12,5-
КГ/ММ ²	5,5	6,5	7,5	8,5	9,5	10,5	11,5	12,5	13,5
n_i	40	32	28	24	20	18	16	12	4

Контрольні питання:

1. Дати означення нульової і альтернативної гіпотез.
2. Що називають простою та складною статистичною гіпотезами?
3. Що називається статистичним критерієм?
4. Що називається емпіричним значенням критерію?
5. Що таке рівень значущості α ?
6. Помилки першого та другого роду.
7. Що таке потужність критерію?
8. Що називають емпіричними частотами?
9. Що називають теоретичними частотами?
10. Записати формули для обчислення теоретичних частот, якщо ознака X генеральної сукупності має нормальний закон розподілу.
11. Критерій узгодженості Пірсона.
12. Загальна методика перевірки правильності гіпотези про закон розподілу ознаки генеральної сукупності.

Лабораторна робота №7

Тема заняття: Перевірка статистичної гіпотези про рівність середніх двох вибірових сукупностей

Мета заняття: Освоєння статистичного інструментарію перевірки гіпотез про оцінку параметрів вибірового розподілу

Основні завдання практичного заняття:

- визначення мети перевірки гіпотез про рівність середніх двох вибірових сукупностей;
- вивчення принципів побудови та схеми перевірки гіпотез;
- приклади перевірки гіпотез.

Теоретичні положення

Під **гіпотезою** розуміють деяке наукове припущення про властивості досліджуваних явищ, яке потребує перевірки та доказів.

Статистичною гіпотезою (H) називається припущення відносно параметрів або форми розподілу генеральної сукупності, яке перевіряється на основі даних вибірового спостереження.

У ході перевірки статистичної гіпотези необхідно встановити, чи узгоджуються дані спостереження з висунутою гіпотезою, чи можливо відмінності між гіпотезою і результатами спостереження віднести до випадкових або ж ці відмінності викликані впливом деяких систематично діючих причин. У результаті перевірки гіпотеза або приймається, або відхиляється.

Найбільш часто перевіряється припущення про те, що отримана за вибіркою величина незначно відрізняється від гіпотетичної (теоретично припущеної) або встановленої величини в генеральній сукупності. Для перевірки цього положення висувається гіпотеза про те, що істинна різниця між фактичними і гіпотетичними показниками дорівнює нулю. У зв'язку з цим гіпотезу, що перевіряється, називають **нульовою** і позначають H_0 . Нульову гіпотезу ще називають **основною** або **робочою гіпотезою**.

У кожному випадку нульовій гіпотезі протиставляється **альтернативна (конкуруюча) гіпотеза (H_a)**, яка заперечує нульову гіпотезу. Наприклад,

якщо нульова гіпотеза полягає в припущенні, що середня врожайність зернових культур в генеральній сукупності дорівнює 35 ц/га, то альтернативна гіпотеза, зокрема, може полягати в тому, що середня врожайність у генеральній сукупності не дорівнює 35 ц/га. Зміст гіпотези записується після двокрапки, а сам запис має вигляд:

$$H_0 : \bar{x} = 35; H_a : \bar{x} \neq 35 .$$

По відношенню до нульової гіпотези може бути вказана незлічима множина альтернативних гіпотез.

За формою побудови гіпотези можуть бути простими і складними. **Простою** називають гіпотезу, яка містить тільки одне припущення, **складною** - гіпотезу, яка містить два і більше припущення.

В результаті перевірки статистичної гіпотези, основаної на даних вибірки обмеженого обсягу, можна відхилити і прийняти нульову гіпотезу (відповідно вибіркові дані суперечать і узгоджуються з H_0). Звідси видно, що перевірка статистичних гіпотез зв'язана з ризиком прийняття помилкових рішень.

Неправильне рішення може бути прийняте у двох випадках. У зв'язку з цим розрізняють помилки двох родів.

Помилка першого роду полягає в тому, що нульова гіпотеза H_0 відхиляється, хоча в дійсності вона правильна.

Помилка другого роду полягає в тому, що приймається нульова гіпотеза H_0 , хоча в дійсності правильна альтернативна гіпотеза H_a .

Правильні і неправильні рішення можуть бути отримані в двох випадках. Це ілюструє таблиця 7.1.

Таблиця 7.1

Можливі результати перевірки нульової гіпотези

Результат перевірки H_0	Можливий стан гіпотези, що перевіряється	
	правильна гіпотеза H_0	правильна гіпотеза H_a
H_0 відхиляється	Помилка першого роду α	Правильне рішення
H_0 приймається	Правильне рішення	Помилка другого роду β

Ймовірність зробити помилку першого роду (невиправдане відхилення H_0) отримала назву **рівня значущості** і позначається α . Ймовірність зробити помилку другого роду (прийняття неправильної гіпотези) позначається β . Отже, можна сказати, що при великому числі вибірок частка хибних висновків дорівнює α , якщо правильна H_0 і дорівнює β , якщо правильна H_a .

Помилки I і II роду за своїми наслідками нерівнозначні і ведуть до різних матеріальних втрат. Тому вибір рівня значущості повинен ґрунтуватись на обліку можливих втрат: чим більші ці втрати, тим меншим повинен бути рівень значущості. Однак, якщо знижується рівень значущості, збільшується ймовірність появи помилок другого роду. В цьому розумінні помилки I і II роду є конкуруючими.

Оскільки помилки I і II роду практично виключити неможливо, то в кожному випадку необхідно прагнути до зменшення втрат від цих помилок. При практичній перевірці гіпотез прагнуть до того, щоб за помилку I роду прийняти ту із можливих помилок, яка матиме більш серйозні наслідки на практиці.

Рівень значущості встановлюється самим дослідником залежно від характеру і важливості завдань, що розв'язуються (за так званим принципом практичної впевненості). Рівень значущості являє собою ту мінімальну ймовірність, починаючи з якої можна визнати подію практично неможливою. Можна користуватися стандартними значеннями $\alpha = 0,10$; $0,05$; $0,01$; $0,001$; $0,0001$ та ін. Найбільш часто α встановлюють на рівні $0,05$ і $0,01$. При більш відповідальних рішеннях α підвищують до $0,001$. Рівень значущості, наприклад, $\alpha = 0,05$, означає, що в середньому в 5 випадках із 100 є ризик допустити помилку I роду, тобто відхилити правильну гіпотезу (H_0).

Встановлюючи певний рівень значущості, дослідник контролює ймовірність помилки I роду: чим він нижчий, тим частіше H_0 буде визнаватися правильною. Однак, як було відзначено вище, зниження рівня значущості веде до появи помилок II роду. В більшості випадків єдиним шляхом одночасного зменшення ймовірності появи помилок двох родів є збільшення чисельності вибірки.

Для перевірки нульової гіпотези і прийняття висновку про сумісність вибірових даних з висунутою гіпотезою використовують спеціальні **статистичні критерії**, які являють собою зведення правил, за якими гіпотеза, що перевіряється, або приймається, або відхиляється. Для кожного виду гіпотез, що перевіряються, розроблені спеціальні критерії, серед яких найбільш часто використовуються t -критерії нормального розподілу і розподілу Стюдента, F -критерій Фішера-Снедекора, χ^2 (хі-квадрат)- розподілу Пірсона та інші

У більшості можливих значень вибраного критерію можна виділити дві неперетинні підмножини, одна з яких містить значення критерію, а друга ні. Перша підмножина називається **критичною областю**, а друга **областю допустимих значень**.

Критичною областю називають ті значення критерію, при яких нульова гіпотеза відхиляється. **Областю допустимих значень** (областю прийняття H_0) називають сукупність значень критерію, що використовується, при яких нульова гіпотеза приймається.

Точки, які відділяють критичну область від області допустимих значень, називаються **критичними точками**.

При формулюванні висновків за результатами перевірки керуються таким принципом (правилом): якщо фактичне значення критерію попадає в критичну область, то H_0 відхиляють, якщо ж фактичне значення критерію належить до області допустимих значень, то H_0 приймають.

Для кожного критерію складені спеціальні таблиці за якими знаходять його табличне значення (критичні точки), які відділяють критичну область від області допустимих значень. Знайдене табличне значення критерію співставляють з його фактичним значенням. Якщо фактичне значення критерію, розраховане за даними вибірки, буде більше від табличного значення, то нульова гіпотеза повинна бути відхилена і прийнята альтернативна гіпотеза. Якщо ж фактичне значення критерію буде меншим або рівним табличному, то робиться висновок про узгодження даних, що спостерігаються, з нульовою гіпотезою, тобто підстави для відмови від H_0 немає і тому вона повинна бути прийнята.

При цьому слід мати на увазі, що згода з нульовою гіпотезою не доводить її абсолютної справедливості. Це лише свідчення про необхідність подальшої її перевірки, зокрема шляхом збільшення обсягу вибірки або поки більш переконливі дослідження не дозволять зробити протилежний висновок. Тому при формулюванні остаточних висновків у цьому випадку правильніше говорити про те, що дані спостереження не суперечать нульовій гіпотезі і, отже, не дають підстав для її відхилення.

Критерій перевірки гіпотези повинен бути підібраний так, щоб ризик допущення помилок був мінімальним. При цьому дуже важливо визначати ймовірність того, що не буде допущена помилка II роду. Ця ймовірність характеризує чутливість критерію до помилок II роду і отримала назву потужності критерію.

Потужністю критерію називається ймовірність відхилення гіпотези H_0 , що випробовується, коли правильна альтернативна гіпотеза $H_a(1-\beta)$. Отже, потужністю критерію є ймовірність того, що не буде допущена помилка II роду. Очевидно, що бажано мати найбільш потужний критерій, оскільки це забезпечить мінімальну ймовірність допущення помилки II роду.

Потужність (чутливість) критерію може бути підвищена двома способами:

а) шляхом збільшення рівня значущості. Проте цей шлях не зовсім прийнятний, так як необгрунтовано підвищується ймовірність помилок I роду;

б) шляхом збільшення чисельності вибірки.

Перевірка статистичної гіпотези передбачає послідовне виконання таких етапів:

1. Оцінка вихідної інформації і опис статистичної моделі вибіркової сукупності.
2. Формулювання нульової і альтернативної гіпотез.
3. Встановлення рівня значущості, за допомогою якого контролюється помилка I роду.
4. Вибір найбільш потужного критерію для перевірки нульової гіпотези.
5. Розрахунок за певним алгоритмом фактичного значення критерію.

6. Визначення критичної області і області згоди з нульовою гіпотезою, тобто встановлення табличного значення критерію.
7. Співставлення фактичного і табличного значень критерію і формулювання висновків за результатами перевірки нульової гіпотези.

Вибір конкретного методу перевірки гіпотези залежить від характеру досліджуваної гіпотези, властивостей вихідної інформації та інших умов.

Розглянемо основні з них :

1.Обсяг вибіркової сукупності. При перевірці гіпотез за даними великих вибірок ($n > 30$) доцільно використовувати t -критерій нормального розподілу, а за даними малих вибірок ($n < 30$) - t -критерій розподілу Стьюдента.

2.Рівність вибірок за чисельністю. Вибіркові сукупності за чисельністю можуть бути рівними і нерівними. Ці властивості необхідно враховувати при практичній перевірці гіпотез про істотність відмінностей між середніми, зокрема, при розрахунку середньої помилки двох вибірових середніх.

3.Схема формування вибірок. Прийоми перевірки статистичних гіпотез залежать від характеру формування вибірових сукупностей. Якщо спостереження однієї вибірки не протиставляються спостереженням іншої вибірки, то такі вибірки називаються **незалежними**. Якщо ж спостереження однієї вибірки деякою мірою пов'язані зі спостереженням іншої вибірки, то такі вибірки називаються **залежними**. Прикладом незалежних вибірок може служити дослід з двома групами тварин, одна з яких є контрольною, а на другій випробовується якийсь препарат. При цьому дослідна і контрольна група формується випадково. Однак, якщо при проведенні цього ж досліді тварини будуть попередньо розподілені на групи за будь-якими ознаками (наприклад, біологічними властивостями, продуктивністю тощо), а потім від кожної пари аналогів буде відібрано по одному представнику в контрольну і дослідну групи, то спостереження за такими вибірками, вже не можна розглядати як незалежні.

Формування вибірових сукупностей зумовлює різницю прийомів оцінки вірогідності між середніми двох малих вибірок. Якщо вибірки

незалежні, то статистичній оцінці підлягає різниця середніх, якщо залежні - середня різниця.

4. Рівність дисперсій. При перевірці гіпотез щодо середніх можливі два випадки по відношенню до вибірових дисперсій:

- 1) дисперсії рівні ($\sigma_1^2 = \sigma_2^2$);
- 2) дисперсії нерівні ($\sigma_1^2 \neq \sigma_2^2$).

У зв'язку з цим виникає спеціальне завдання перевірки гіпотези про істотність відмінностей двох дисперсій. Для перевірки гіпотези про рівність двох дисперсій у генеральних сукупностях використовується критерій F-розподілу Фішера, який оснований на співвідношенні двох вибірових скорегованих дисперсій: S_1^2 і S_2^2 , що заміняють значення дисперсій у генеральних сукупностях, які, як правило, невідомі:

$$F = \frac{S_1^2}{S_2^2}, \text{ де } S_1^2 > S_2^2.$$

Критичні значення критерію Фішера (F_α) знаходять за спеціальними таблицями при відповідному числі ступенів свободи і заданому рівні значущості α . Перевірка H_0 зводиться до такого. Спочатку визначають фактичне значення критерію $F_{\text{факт}}$. Потім за таблицями знаходять його критичне значення F_α . Якщо виявиться, що $F_{\text{факт}} > F_\alpha$, то H_0 про рівність двох дисперсій у генеральних сукупностях відхиляється, якщо ж $F_{\text{факт}} < F_\alpha$, то H_0 приймається.

Належно від того, рівні чи нерівні дисперсії в генеральних сукупностях, необхідно видозмінювати схему перевірки гіпотези.

Приклад 7.1. Перевірка гіпотези відносно двох середніх. Вибірki незалежні з рівними чисельностями ($n_1 = n_2$) і рівними дисперсіями в генеральних сукупностях ($\sigma_1^2 = \sigma_2^2$). У приміському держгоспі досліджується раціон з добавкою вітамінів при відгодівлі тварин. Були організовані дослідна і контрольна група по 5 голів у кожній групі. Протягом місяця дослідна група тварин отримала понад звичайного

раціону комплекс вітамінів. В кінці місяця для кожного виду визначено приріст у живій масі. Дані досліді подані у табл. 7.2.

Співставлення середніх добових приростів по двох групах тварин показує, що більш високий добовий приріст дали тварини дослідної групи. Однак, у зв'язку з тим, що чисельність вибірок невелика ($n = 5$), не виключена можливість, що розходження між добовими приростами отримані в результаті дії випадкових причин.

Потрібно статистично оцінити різницю між середніми по двох групах тварин.

За результатами перевірки гіпотези зробити висновок про те, що різниця між середніми лежить у межах випадкових коливань, або ж ця різниця настільки значна, що не узгоджується з нульовою гіпотезою про випадковий характер відмінностей між середніми.

Таблиця 7.2.

Добовий приріст живої маси відгодівельного поголів'я, г

Тварини	Основний раціон + вітаміни (дослід) x_{1i}	Основний раціон (контроль) x_{2i}
1	632	520
2	691	586
3	570	490
4	694	609
5	683	590

Якщо буде доведено друге положення і відхилене перше, можна стверджувати, що умови росту тварин дослідної групи істотно відрізнялись від умов росту тварин контрольної групи, тобто комплекс вітамінів стимулював підвищення добових приростів.

Розв'язання:

Умова задачі припускає, що обидві вибірки взяті із нормально розподіленої генеральної сукупності. Формування груп випадкове (незалежне),

тому оцінюватись повинна різниця між середніми. В середовищі Excel створюємо форму вихідних даних (Рис.7.1).

	A	B	C	D	E
1		Перевірка гіпотези відносно двох середніх.			
2		Тварини	Основний раціон + вітаміни (дослід)	Основний раціон (контроль)	
3		1	632	520	
4		2	691	586	
5		3	570	490	
6		4	694	609	
7		5	683	590	
8					

Рис.7.1. Форма вихідних даних задачі

Визначимо середні добові прирости у дослідній і контрольній групах за допомогою вбудованої функції **СРЗНАЧ()**, параметри якої є масиви даних відповідних вибірок. Вказана функція обчислює середні за формулою:

$$\overline{x_i^*} = \frac{\sum x_{ij}}{n_i} . \text{ Так, в рядку 8 отримуємо: значення } \overline{x_1^*} = 654, \overline{x_2^*} = 559, \text{ а}$$

також в рядку 9 обчислимо фактичну різницю між середніми: $\left| \overline{x_1^*} - \overline{x_2^*} \right|$

(Рис.7.2).

	A	B	C	D	E
1		Перевірка гіпотези відносно двох середніх.			
2		Тварини	Основний раціон + вітаміни (дослід)	Основний раціон (контроль)	
3		1	632	520	
4		2	691	586	
5		3	570	490	
6		4	694	609	
7		5	683	590	
8		Середнє значення	654	559	
9		Різниця	95		
10					

Рис.7.2. Обчислення середніх та їх різниці

Істотність цієї різниці повинна бути оцінена. Для цього необхідно перевірити гіпотезу про рівність двох середніх.

Розглянемо докладно всі етапи схеми перевірки гіпотези.

1. Сформулюємо нульову H_0 і альтернативну H_a гіпотези:

$$H_0 : \bar{x}_1 = \bar{x}_2; \quad H_a : \bar{x}_1 \neq \bar{x}_2.$$

2. Прийmemo рівень значущості $\alpha=0,05$, гарантуючи прийняття гіпотези або відмови від неї з ймовірністю помилки тільки в 5 випадках із 100.

3. Найбільш потужним критерієм для перевірки такої гіпотези H_0 є t -критерій Стюдента.

4. Сформулюємо правило прийняття рішення за результатами перевірки H_a . Оскільки за альтернативною гіпотезою \bar{x}_1^* може бути або більше або менше \bar{x}_2^* , то критична область повинна бути встановлена з двох сторін: $|t| > t_\alpha$ (так звана двостороння критична область завдання критерію).

Критична область при $\alpha=0,05$ буде міститися в межах - всі значення вищі, ніж верхнє 2,5% і нижче, ніж 2,5% точки розподілу t -критерію Стюдента.

Висновок при перевірці H_0 можна сформулювати таким чином: гіпотеза H_0 відхиляється, якщо вибіркове значення критерію $t_{факт}$ виявиться в критичній області, тобто якщо $t_{факт} > t_\alpha$. В протилежному випадку H_0 повинна бути прийнята.

5. Щоб перевірити H_0 , потрібно визначити фактичне значення t -критерію Стюдента і порівняти його з табличним значенням.

Фактичне значення t -критерію Стюдента визначимо за формулою:

$$t_{факт} = \frac{\left| \bar{x}_1^* - \bar{x}_2^* \right| - \left| \bar{x}_1 - \bar{x}_2 \right|}{\mu_{1-2}},$$

де \bar{x}_1^* - середнє значення ознаки за першою вибіркою (середній добовий приріст за даними дослідної групи); \bar{x}_2^* - середнє значення

ознаки за другою виборкою (середній добовий приріст за даними контрольної групи); \bar{x}_1 і \bar{x}_2 - середнє значення ознаки відповідних генеральних сукупностей, з яких зроблені вибірки (оскільки згідно з H_0 : $\bar{x}_1 = \bar{x}_2$, то $\bar{x}_1 - \bar{x}_2 = 0$); $\overline{\mu_{1-2}}$ - узагальнена середня помилка різниці середніх.

Для визначення фактичного значення t -критерію Стюдента зробимо такі обчислення.

6. Обчислимо за кожною вибіркою скореговані на втрату ступенів свободи варіації дисперсії за допомогою вбудованої функції **ДИСП()**, параметри якої є масиви даних відповідних вибірок (рядок 10). Обчислення за

вказаною функцією проводиться на основі формули: $S_i^2 = \frac{\sum x_{ij}^2 - n_i \left(\overline{x_i^*}\right)^2}{n_i - 1}$.

7. Розрахуємо квадрати середніх помилок за кожною виборкою й узагальнену середню помилку різниці середніх (рядки 11, 12 Рис.7.3.):

$$\mu_1^2 = \frac{S_1^2}{n_1} = \frac{2837,5}{5} = 567,5; \quad \mu_2^2 = \frac{S_2^2}{n_2} = \frac{2618,0}{5} = 523,6;$$

$$\overline{\mu_{1-2}} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{567,5 + 523,6} = 33,0$$

8. Розрахуємо фактичне значення t -критерію Стюдента (рядок 13 Рис.7.3.):

$$t_{\text{факт}} = \frac{\left[\overline{x_1^*} - \overline{x_2^*} \right]}{\overline{\mu_{1-2}}} = \frac{654 - 559}{33,0} = 2,88.$$

9. Встановимо табличне значення t -критерію Стюдента (рядок 14 Рис.7.3.), виходячи із рівня значущості $\alpha=0,05$ і загального числа ступенів свободи для двох вибірок: $k = (n_1 - 1) + (n_2 - 1) = 5 - 1 + 5 - 1 = 8$ за допомогою вбудованої функції **СТЮДРАСПОБР(0,05;8)**: $t_{0,05} = 2,31$.

10. Співставимо фактичне і табличне значення t -критерію Стюдента: $t_{\text{факт}} > t_{0,05}$; $2,88 > 2,31$.

Практичне заняття на ПК № 8

Тема заняття:

- Побудова однофакторної лінійної регресійної моделі.

Мета заняття:

- Закріпити основні теоретичні положення регресійного аналізу;
- Оволодіти технікою використання функції "ЛИНЕЙН" в Excel;
- Використати статистичні дані функції "ЛИНЕЙН" для перевірки коефіцієнтів регресії на значимість.

Зміст заняття.

Використання даних. Завантажити дані на перший аркуш в Excel і означити всі показники.

Використання функції "ЛИНЕЙН". Для цього:

- Визвати файл с даними;
- Переписати дані у свій файл на перший аркуш;
- На першому аркуші поруч з даними відмітити поле розміром 5×2 та натиснути клавішу "=";
- Визвати функцію "ЛИНЕЙН" та заповнити поля. У третє та четверте поля слід записати одиниці. У цьому разі програма видасть нульовий коефіцієнт регресії та статистику.
- Замість "ОК" слід натиснути "Ctrl+Shift+Enter";
- У виділеному полі розміром $5 \times (m+1)$ появляться:
 1. У першої стрічці коефіцієнти регресії по зростанню справа вліво;
 2. У другої стрічці скоректовані помилки коефіцієнтів регресії;
 3. У третій стрічці, зліва вправо, коефіцієнт детермінації та стандартна помилка розрахункових значень;
 4. У четвертій стрічці, зліва вправо, спостережене значення критерію Фішера та число ступенів волі;
 5. У п'ятій стрічці, зліва вправо, сума квадратів відхилень розрахункових значень та сума квадратів відхилень залишкової похибки.

Перевірка параметрів моделі (коефіцієнтів регресії) на значимість. Для цього слід використати першу та другу стрічку вихідної статистики функції "ЛИНЕЙН". Вони дають можливість розрахувати статистику Стьюдента

$$t = \frac{b}{\tilde{S}_b}$$

для коефіцієнтів регресії та порівняти її до критичних значень розподілу Стьюдента $t_{\alpha}(n-2)$ для рівня значимості α . Якщо $t > t_{\alpha}(n-2)$, відповідний коефіцієнт регресії є значимим з вірогідністю $\gamma = 1 - \alpha$.

Демонстраційне рішення проблеми значимості в Excel.

y - Середньомісячна продуктивність лави

x - довжина лави

Варіант 1			0,070251	2,397937	t= 3,055501
y	x		0,022992	4,793154	
23,556	348		0,329477	7,200867	
7,644	94		9,336088	19	
14,3	195		484,0994	985,1973	
18,928	300				
33,618	164				

На даному аркуші Excel приведені дані середньомісячної продуктивності лави у гірничодобувній промисловості та довжина лави. Результат застосування функції "ЛИНЕЙН" приведено поруч. Тоді t-статистика для першого коефіцієнта регресії

$$t = 0,070251 / 0,022992 = 3,055501$$

Ця величина розрахована в Excel праворуч з таблицею статистики. Критичне значення статистики Стюдента на рівні 0,01, тобто $t_{0,01}(19) = 2,86$ є меншим ніж спостережене значення. Це означає, що коефіцієнт регресії є значимим на цьому рівні.

Звіт. Звіт про виконану роботу повинен містити:

- Постанову завдання;
- Досліджувані дані;
- Методику рішення;
- Отримані результати.

Контрольні питання:

- Чим відрізняється регресійний аналіз від кореляційного?
- Що означає "модель випадкового явища" у генеральній сукупності?
- Що означає "модель випадкового явища" у вибірковій сукупності?
- Що мається на увазі під розрахунковою функцією?
- Запишіть двохфакторну лінійну модель.
- Якою є головна ідея розрахунку параметрів цієї моделі?
- Відкіля береться залишкова похибка?
- Якщо дані тільки значення показників X та Y чи можна розрахувати значення залишкової похибки та саме як?
- Що означає перевірка на значимість коефіцієнтів регресії?
- Як можна перевірити на значимість коефіцієнти регресії двохфакторної лінійної моделі?

Практичне заняття на ПК № 10

1 Тема заняття: Перевірка лінії регресії на значущість. Побудова довірчого інтервалу на лінію регресії.

2 Мета заняття:

2.1 Перевірка лінії регресії на значущість за допомогою критерію Фішера для визначення, наскільки мінливість залежного фактора Y , що пов'язана з впливом регресора X , перевищує мінливість випадкової похибки.

2.2 Визначення з ймовірністю 95 % можливих змін лінії регресії від вибірки до вибірки, визначаючи величини залишкової похибки лінійної регресійної моделі.

3 Зміст практичного заняття

Для виконання даного завдання та обчислення необхідних параметрів використовуються наступні функції з категорії "Статистичні":

а) функція "ЛИНЕЙН" обчислює параметри лінійного приближення за методом найменших квадратів;

Для її застосування треба відмітити поле розміром 5x2 та натиснути клавішу "=". Визивати функцію "ЛИНЕЙН" та заповнити поля. При цьому слід пам'ятати, що у друге поле записуються регресори у вигляді однієї матриці. У третє та четверте-поля слід записати одиниці. У цьому разі програма видасть нульовий коефіцієнт регресії та статистику. Замість "OK" слід натиснути "Ctrl+Shift+Enter".

b_1	b_0
$S_{e_{b_1}}$	$S_{e_{b_0}}$
R^2	S_e
F	$df = n - 2$
SSR	SSE

У виділеному полі розміром 5x2 появляться наступні значення:

1 рядок - коефіцієнти лінії регресії: кутовий коефіцієнт b_1 (ліворуч) та вільний член b_0 (праворуч);

2 рядок – скоректовані помилки коефіцієнтів регресії $S_{e_{b_1}}$ та $S_{e_{b_0}}$;

3 рядок – коефіцієнт детермінації (квадрат коефіцієнту регресії R^2 , ліворуч) та стандартна помилка розрахункових значень;

4 рядок – спостережене значення критерію Фішера F (ліворуч) та число ступенів свободи $df = n - 2$ (праворуч);

5 рядок – сума квадратів відхилень розрахункових значень SSR (ліворуч) та сума квадратів відхилень залишкової похибки SSE (праворуч).

б) функція “*FPACPOBP*” обчислює значення для F-розподілу ймовірностей. Вона має три аргументи: *FPACPOBP(Вероятность; степени_свободы1; степени_свободы2)*, де *Вероятность* обирається 0,05 або 0,01; *степени_свободы1* дорівнюють 1 (критерій Фішера), *степени_свободы2* дорівнюють $df = n - 2$.

в) функція “*ДИСПР*” обчислює дисперсію для генеральної сукупності масиву випадкової величини;

Обираємо функцію “*ДИСПР*”, виділяємо необхідний стовпець даних, натискаємо “ОК”.

г) функція “*CP3HACH*” обчислює середнє значення масиву випадкової величини;

Обираємо функцію “*CP3HACH*”, виділяємо необхідний стовпець даних, натискаємо “ОК”.

Для побудови необхідних ліній використовується *Мастер диаграмм / точечная / точечная диаграмма со значениями, соединенными сглаживающими линиями / далее / ряд /* У розділі „ряд” далі задаються стовпці даних, які необхідно зобразити на графіку.

4 Демонстраційне розв'язання задач на ПК з проміжними та кінцевим результатами:

- Заповнюється таблиця даних. У 1-й та 2-й стовпці вносяться значення двох випадкових величин x та y (табл. 1).
- Відмітити нижче (під таблицею 1) поле розміром 5x2 та натиснути клавішу “=”. Визвати функцію “*ЛИНЕЙН*” та заповнити поля. У виділеному полі розміром 5x2 з’явиться таблиця 2.

Таблица 1

x	y	y_p	$X = x - \bar{x}$	$Y = \tilde{S}_{y_p}$	$y_p - Y$	$y_p + Y$	y_{ac1}	y_{ac2}
-----	-----	-------	-------------------	-----------------------	-----------	-----------	-----------	-----------

1	2	3	4	5	6	7	8	9
2	30,291	23,599	-8,2857	6,540	17,060	30,139	30,157	17,042
2	13,44	23,599	-8,2857	6,540	17,060	30,139	30,157	17,042
4	35,99	21,659	-6,2857	5,372	16,287	27,031	26,634	16,684
5	20,714	20,689	-5,2857	4,831	15,858	25,520	24,872	16,506
8	25,756	17,779	-2,2857	3,535	14,243	21,314	19,588	15,970
8	13,613	17,779	-2,2857	3,535	14,243	21,314	19,588	15,970
10	10,076	15,838	-0,2857	3,169	12,669	19,008	16,064	15,612
10	28,585	15,838	-0,2857	3,169	12,669	19,008	16,064	15,612
10	13,35	15,838	-0,2857	3,169	12,669	19,008	16,064	15,612
10	7,596	15,838	-0,2857	3,169	12,669	19,008	16,064	15,612
11	9,817	14,868	0,7143	3,201	11,667	18,070	14,303	15,433
11	9,555	14,868	0,7143	3,201	11,667	18,070	14,303	15,433
11	13,762	14,868	0,7143	3,201	11,667	18,070	14,303	15,433
12	14,386	13,898	1,7143	3,378	10,520	17,276	12,541	15,255
12	3,72	13,898	1,7143	3,378	10,520	17,276	12,541	15,255
12	11,35	13,898	1,7143	3,378	10,520	17,276	12,541	15,255
12	9,519	13,898	1,7143	3,378	10,520	17,276	12,541	15,255
13	13,51	12,928	2,7143	3,677	9,251	16,605	10,780	15,076
13	21,64	12,928	2,7143	3,677	9,251	16,605	10,780	15,076
18	10,894	8,077	7,7143	6,197	1,880	14,274	1,972	14,183
22	9,22	4,197	11,7143	8,689	-4,492	12,885	-5,074	13,468

Таблиця 2

-0,97014	25,53968
0,345406	3,888879
0,293383	7,247652
7,888689	19
414,3807	998,0408

• Перевіряємо значущість моделі, тобто впливовість факторів, які враховуються цією моделлю, за допомогою критерію Фішера. Із отриманої таблиці 5x2 у четвертому рядку беремо спостережене значення критерію $F = 7,888689$.

Знаходимо критичні точки для критерію Фішера визначаємо за допомогою функція “ФРАСПОБР” з урахуванням числа ступенів свободи для відповідних дисперсій: $F_{0,01}(1; 21 - 2) = F_{0,01}(1; 19) = 8,184947$ (ФРАСПОБР(0,01;1;19), рівень значущості $\alpha = 0,01$, надійність 99 %) та $F_{0,05}(1; 21 - 2) = F_{0,05}(1; 19) = 4,38075$ (ФРАСПОБР(0,05;1;19), $\alpha = 0,05$, надійність 95 %)).

F 0,01 (1,19)=	ФРАСПОБР(0,01;1;19)=	8,184947
F 0,05 (1,19)=	ФРАСПОБР(0,05;1;19)=	4,38075

Порівнюємо з F_R . Звідси маємо, що $F_{0,05} < F_R < F_{0,01}$ або $4,38075 < F_R < 8,184947$, тобто кореляційний зв'язок не є значущим.

- На базі отриманих значень коефіцієнтів b_0 та b_1 одержуємо значення лінії регресії y_p ($y_p = b_0 + b_1 \cdot x$ – парна лінія регресії Y на X) та заносимо їх у стовпець 3 (см. табл.1).

- Обчислюємо $\tilde{S}_{\bar{y}} = \frac{S_e}{\sqrt{n-2}} = \frac{7,247652}{\sqrt{19}} \approx 1,66$. Значення $S_e = 7,247652$ отримано у третьому рядку таблиці 2 за допомогою функцію "ЛИНЕЙН".

$\tilde{S}_{\bar{y}} =$	1,662725
-------------------------	----------

- За допомогою функції "ДИСПР" обчислюємо дисперсію для генеральної сукупності масиву випадкової величини x :

$S_x =$	4,578863
---------	----------

- Стовпець 4 містить значення $X = x - \bar{x}$. За цими значеннями обчислюємо $\tilde{S}_{y_p} = 2 \cdot \tilde{S}_{\bar{y}} \cdot \sqrt{1 + \frac{(x - \bar{x})^2}{S_x^2}}$ та містимо у стовпець 5 (см. табл.1).

- Для побудови гіперболи заповнюємо стовпці 6 та 7, вмістивши значення $y_p - \tilde{S}_{y_p}$ та $y_p + \tilde{S}_{y_p}$ відповідно (см. табл.1).

- Для побудови асимптот гіперболи одержимо їх значення за формулою $y_{ac} = \pm \frac{b}{a}(x - \bar{x}) + y_p$ (це 8 та 9 стовпці), де $b = \tilde{S}_{\bar{y}}$, $a = S_x$ (a , b – піввісі гіперболи $Y = \pm b \cdot \sqrt{1 + \frac{X^2}{a^2}}$, центр якої знаходиться у центрі вибіркової сукупності $(\bar{x}; \bar{y})$, $X = x - \bar{x}$, $Y = y_p$).

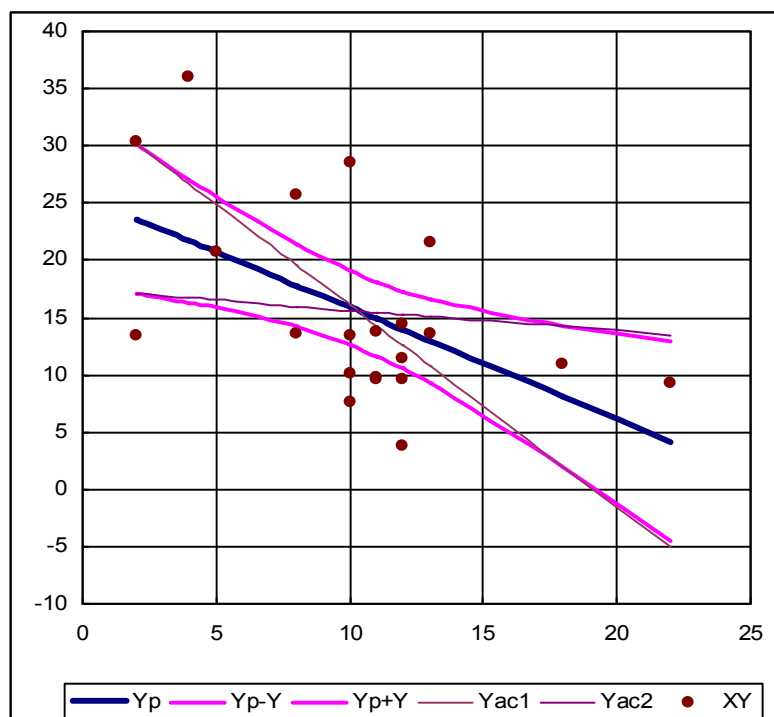
Отримаємо a , b (ці дані отримані були раніше):

$a = S_x =$	4,578863
$b = \tilde{S}_{\bar{y}} =$	1,662725

Заповнимо стовпці 8 та 9 (см. табл.1).

- Далі будуюмо всі необхідні лінії. Використовуємо *Мастер диаграмм* / точечная / точечная диаграмма со значениями, соединенными сглаживающими линиями / далее / ряд /
добавить: ряд 1, значения X – вміст 1-го стовпця, значения Y – вміст 3-го стовпця (теоретична лінія регресії);
добавить: ряд 2, значения X – вміст 1-го стовпця, значения Y – вміст 6-го стовпця (перша вітка гіперболи);
добавить: ряд 3 значения X – вміст 1-го стовпця, значения Y – вміст 7-го стовпця (друга вітка гіперболи);
добавить: ряд 4, значения X – вміст 1-го стовпця, значения Y – вміст 8-го стовпця (перша асимптота);
добавить: ряд 5, значения X – вміст 1-го стовпця, значения Y – вміст 9-го стовпця (друга асимптота)
добавить: ряд 6, значения X – вміст 1-го стовпця, значения Y – вміст 2-го стовпця (кореляційне поле)
далее / далее / готово.

Таким чином, побудуємо лінію регресії $y_p = b_0 + b_1 \cdot x$ з 95%-м довірчим інтервалом:



Наявність довірчої смуги дозволяє встановити межі використання моделі. Найбільш надійні результати (з малою довірчою похиб-

кою) одержуються в колі центра групування даних $(\bar{x} - S_x, \bar{x} + S_x)$, далі довірча похибка швидко зростає.

5. Звіт вирішення задач практичного заняття (зразковий):

5.1 Постанова завдання

5.2 Досліджувані дані

5.3 Методика рішення

5.4 Отримані результати

5.5 Висновки

6. Контрольні питання для захисту задач даного практичного заняття:

6.1 95%-ва довірча смуга для вибіркової лінії регресії.

6.2 Методи її побудови.

6.3 Поняття значущості лінії регресії.

6.4 Статистичні гіпотези перевірки значущості лінії регресії за допомогою критерію Фішера.

6.5 Значимість моделі. Розподіл Фішера. Критерій Фішера. Число ступенів свободи: чисельник, знаменник.

6.6 Середнє значення, дисперсія.

6.7 Що таке виправлена дисперсія?

6.8 Число ступенів свободи для похибки.

7. Варіанти для самостійної або індивідуальної роботи наведені в лабораторній роботі №8.

Питання до екзамену

з дисципліни «Теорія ймовірностей та математична статистика»

1. Випадкові події та операції над ними.
2. Означення ймовірності: класичне, геометричне, статистичне.
3. Аксиоматична побудова теорії ймовірностей.
4. Правила додавання та множення ймовірностей.
5. Формула повної ймовірності. Приклади застосування.
6. Формула Бейєса. Приклади застосування.
7. Повторні незалежні випробування. Формула Бернуллі. Приклади застосування.
8. Повторні незалежні випробування. Формула Пуассона. Приклади застосування.
9. Теореми Муавра-Лапласа: локальна та інтегральна. Приклади застосування.
10. Дискретні випадкові величини. Функція розподілу. Числові характеристики.
11. Властивості числових характеристик випадкової величини.
12. Біномний розподіл випадкової величини. Розподіл Пуассона.
13. Неперервні випадкові величини. Функція розподілу, щільність ймовірності. Властивості щільності ймовірності.
14. Числові характеристики неперервної випадкової величини.
15. Рівномірний та експоненціальний закони розподілу випадкової величини.
16. Нормальний закон розподілу випадкової величини.
17. Випадкові вектори.
18. Нормальний розподіл двовимірного випадкового вектора. Закон розподілу Релея.
19. Граничні теореми теорії ймовірностей.
20. Співвідношення теорії ймовірностей та математичної статистики. Поняття про генеральну сукупність та вибірку. Варіаційний та статистичний ряди, полігон частот, гістограма.

21. Постановка задачі статистичного оцінювання параметрів. Статистичні оцінки, їх основні властивості: поняття про незміщену, спроможну та ефективну оцінки.
22. Точкове оцінювання невідомих параметрів.
23. Інтервальне оцінювання невідомих параметрів.
24. Лінійна кореляція. Поняття про криволінійну кореляцію.
25. Метод найменших квадратів. Схема методу. Випадок лінійної моделі.
26. Метод найменших квадратів. Оцінка точності отриманих розв'язків у разі лінійної моделі.
27. Метод найменших квадратів. Випадок нелінійної моделі.
28. Основні типи гіпотез, що перевіряються у ході статистичної обробки даних. Загальна логічна схема статистичного критерію.
29. Перевірка відповідності обраної моделі розподілу вихідним даним. Критерій узгодження χ^2 . Схема застосування.
30. Поняття про непараметричні статистичні критерії. Критерії Колмогорова, Вілкоксона, Спірмена, Кендалла.
31. Метод рангової кореляції.