Міністерство освіти і науки України

Тернопільський національний технічний університет імені Івана Пулюя

Кафедра вищої математики

Конспект лекцій

з дисципліни

«Статистичні методи обробки результатів досліджень»

для здобувачів вищої освіти другого (магістерського) рівня за спеціальністю G13 «Харчові технології» денної та заочної форм навчання

Тернопіль, 2026

УДК 519.2(075.8)

Габрусєв Г. В. Конспект лекцій з дисципліни «Статистичні методи обробки результатів досліджень» для здобувачів вищої освіти другого (магістерського) рівня за спеціальністю G13 «Харчові технології» денної та заочної форм навчання / Григорій Габрусєв. — Тернопіль : СМП «Тайп», 2026. — 165 с.

Укладач:

Г. В. Габрусєв, кандидат фізико-математичних наук, доцент.

Рецензент:

Б. Г. Шелестовський, кандидат фізико-математичних наук, доцент.

Конспект лекцій розглянуто і затверджено на засіданні
кафедри вищої математики ТНТУ ім. Івана Пулюя
(Протокол № 1 від 27 серпня 2026 року).

Схвалено і рекомендовано до друку методичною комісією
факультету прикладних інформаційних технологій та електроінженерії
Тернопільського національного технічного університету імені Івана Пулюя
(Протокол № 1 від 28 серпня 2026 року).

© Габрусєв Г. В
© СМП "ТАЙП", 2026

Передмова

Конспект лекцій призначений для здобувачів вищої освіти другого (магістерського) рівня за спеціальністю G13 «Харчові технології» та забезпечує теоретичний супровід вивчення дисципліни «Статистичні методи обробки результатів досліджень». Його мета — сформувати цілісне розуміння ймовірнісних і статистичних методів, потрібних для аналізу експериментальних даних, оцінювання мінливості показників якості, порівняння технологічних режимів, перевірки наукових гіпотез і побудови моделей для інтерпретації та прогнозування технологічних процесів.

Матеріал організовано послідовно: від основ теорії ймовірностей і випадкових величин — до описової статистики, статистичного оцінювання, перевірки гіпотез, порівняння груп, дисперсійного, кореляційного та регресійного аналізу. Приклади й пояснення орієнтовано на дослідження у сфері харчових технологій: контроль якості, технологічні режими, вимірювання показників продукції, органолептичне оцінювання та прогнозування змін під час зберігання.

Конспект доцільно використовувати разом із практикумом з цієї дисципліни. Лекційний матеріал пояснює математичні засади та логіку вибору статистичного методу, а практичні роботи дають змогу відпрацювати алгоритми розрахунків і перевірити результати за допомогою Microsoft Excel та спеціалізованого статистичного програмного забезпечення. Використання програмних засобів не скасовує потреби розуміти припущення методу, обмеження висновків і зміст отриманих числових результатів.

Під час роботи з конспектом рекомендовано звертати увагу не лише на формули й обчислювальні процедури, а й на інтерпретацію результатів. Статистична значущість не тотожна практичній важливості; кореляція сама по собі не доводить причинності; висока якість підгонки моделі не гарантує точного прогнозу поза межами досліджених умов. Ці принципи є наскрізними для всього курсу.

Структура лекційного курсу

Навчальний матеріал згруповано у два послідовні логічні блоки.

Блок I. Імовірнісні основи, дані та статистичне оцінювання (лекції 1–6). Розглядаються роль статистики в дослідженнях харчових технологій, теорія ймовірностей, випадкові величини та закони розподілу, вибірковий метод, описова статистика й візуалізація даних, статистичне оцінювання параметрів і довірчі інтервали, підготовка даних та сучасне статистичне програмне забезпечення.

Блок II. Статистичне висновування та моделювання (лекції 7–14). Вивчаються методологія перевірки статистичних гіпотез, перевірка розподілів і статистичних припущень, порівняння двох технологічних процесів, непараметричні критерії, однофакторний дисперсійний аналіз ANOVA, кореляційний аналіз, парна лінійна регресія, множинна регресія та діагностика моделей.

Послідовність тем забезпечує перехід від опису даних до обґрунтованого статистичного висновку та побудови моделей. Для кожного дослідження необхідно узгоджувати статистичний метод із типом даних, дизайном експерименту, припущеннями методу та метою аналізу.

Умовні позначення

Позначення	Значення	Примітка
n, N	Обсяг вибірки; обсяг генеральної сукупності	Використання визначається контекстом
X, Y	Змінні або випадкові величини	У регресії зазвичай X — предиктор, Y — результативна змінна
$\bar{x}, \bar{y}$	Вибіркові середні	Оцінки середніх рівнів
μ, σ^2, σ	Середнє, дисперсія та стандартне відхилення генеральної сукупності	Параметри генеральної сукупності
s^2, s	Вибіркова дисперсія та стандартне відхилення	Описують мінливість у вибірці
SE	Стандартна похибка	Характеризує невизначеність оцінки
CI	Довірчий інтервал	Інтервал оцінювання параметра або середньої відповіді
H_0, H_1	Нульова та альтернативна гіпотези	Формулюються до вибору критерію
α	Рівень значущості	Наперед задана допустима ймовірність помилки першого роду
p	Імовірність або p-value	Точне значення визначається контекстом
t, F, χ^2	Статистики критеріїв	Застосовуються у відповідних процедурах перевірки гіпотез
r, ρ, ρ_s	Коефіцієнти кореляції Пірсона, генеральної кореляції та рангової кореляції Спірмена	Конкретне позначення уточнюється в темі
R^2	Коефіцієнт детермінації	Частка варіативності, описана моделлю в межах її визначення
e_i	Залишок регресійної моделі	Різниця між спостереженим і прогнозованим значеннями
$\hat{Y}, \hat{y}$	Оцінене або прогнозоване значення	Позначення залежить від контексту

Позначення в окремих лекціях можуть уточнюватися відповідно до конкретної статистичної моделі або критерію. Перед обчисленнями слід перевіряти, які величини є параметрами генеральної сукупності, а які — вибірковими оцінками.

Зміст

Розділ	Сторінка
Передмова	2
Структура лекційного курсу	3
Умовні позначення	4
Лекція 1. Статистичні методи в дослідженнях харчових технологій. Основи теорії ймовірностей	6
Лекція 2. Випадкові величини та їх моделювання у технологічних процесах	15
Лекція 3. Основні закони розподілу. Нормальний закон та його застосування у технологічних дослідженнях	26
Лекція 4. Вибірковий метод, описова статистика та візуалізація технологічних даних	35
Лекція 5. Статистичне оцінювання параметрів і довірчі інтервали	47
Лекція 6. Підготовка даних та сучасне програмне забезпечення для статистичного аналізу	57
Лекція 7. Методологія перевірки статистичних гіпотез	72
Лекція 8. Перевірка закону розподілу та статистичних припущень	82
Лекція 9. Порівняння двох технологічних процесів: F-критерій, t-критерій Стьюдента, Welch, довірчі інтервали та розмір ефекту	91
Лекція 10. Непараметричні критерії у технологічних дослідженнях: Mann–Whitney та Wilcoxon signed-rank	101
Лекція 11. Однофакторний дисперсійний аналіз (ANOVA) у дослідженнях харчових технологій	115
Лекція 12. Кореляційний аналіз технологічних даних	127
Лекція 13. Парна лінійна регресія: побудова, оцінювання та прогнозування	139
Лекція 14. Множинна регресія, діагностика моделі та технологічне прогнозування	150
Список літератури	164

Лекція 1

Статистичні методи в дослідженнях харчових технологій. Основи теорії ймовірностей

1. Мета лекції та результати навчання

Мета лекції — сформувати розуміння ролі статистичних методів у сучасних дослідженнях харчових технологій та повторити базові поняття теорії ймовірностей, які надалі використовуватимуться для аналізу вибірок, перевірки статистичних гіпотез, оцінювання технологічних ризиків, кореляційного та регресійного моделювання.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати роль випадковості й статистичної мінливості у виробництві харчової продукції;
- розрізняти випадковий експеримент, елементарну подію, випадкову подію та простір елементарних подій;
- обчислювати ймовірності за класичним і статистичним підходами;
- розрізняти сумісні й несумісні, залежні й незалежні події;
- застосовувати теореми додавання та множення ймовірностей;
- використовувати умовну ймовірність, формулу повної ймовірності та формулу Байєса;
- розпізнавати схему Бернуллі й обчислювати ймовірність появи дефектної продукції у серії незалежних випробувань;
- інтерпретувати ймовірнісний результат у термінах технологічного ризику, контролю якості та надійності.

2. План лекції

1. Роль статистичних методів у наукових та інженерно-технологічних дослідженнях.
2. Статистичні закономірності та джерела мінливості у харчовому виробництві.
3. Випадковий експеримент. Події та простір елементарних подій.
4. Класичне та статистичне означення ймовірності.
5. Сумісні, несумісні, залежні й незалежні події.
6. Теореми додавання та множення. Умовна ймовірність.
7. Формула повної ймовірності та формула Байєса.
8. Повторні незалежні випробування. Схема Бернуллі.
9. Застосування до контролю якості, технологічних ризиків і надійності.

3. Навіщо статистика харчовому технологу

Харчове виробництво працює з об'єктами, властивості яких не є абсолютно сталими. Сировина різних партій відрізняється за вологістю, білком, кислотністю, жирністю, цукристістю, мікробіологічним станом. Технологічне обладнання має похибки підтримання температури, тиску, швидкості подачі або маси дозування. Лабораторне вимірювання також містить

випадкову складову. Тому навіть за незмінного технологічного регламенту результати окремих спостережень не збігаються точно.

Статистика дає інструменти для відповіді на виробничі й наукові запитання:

- який типовий рівень показника якості та наскільки він мінливий;
- чи стабільний процес у часі;
- чи є різниця між двома рецептурами реальною чи випадковою;
- чи змінюється результат при зміні температури, концентрації або тривалості процесу;
- яка ймовірність отримати дефектну одиницю продукції;
- який обсяг вибірки потрібний для достатньо точного висновку;
- чи можна за даними прогнозувати показник якості або термін зберігання.

Ключова ідея. Статистика — не набір формул, а система прийняття рішень за умов неповної інформації та випадкової мінливості.

3.1. Три рівні роботи з даними

Рівень	Основне питання	Типові інструменти
Описовий	Що ми спостерігаємо у вибірці?	Середнє, медіана, стандартне відхилення, гістограма, boxplot
Інференційний	Що можна сказати про генеральну сукупність?	Довірчі інтервали, статистичні гіпотези, ANOVA
Модельний / прогнозний	Як пов'язані змінні і що буде за інших умов?	Кореляція, регресія, прогнозування

Теорія ймовірностей є математичним фундаментом інференційного й модельного рівнів.

4. Статистичні закономірності у виробництві

У реальному виробництві одночасно діють систематичні та випадкові чинники. Наприклад, середня температура пастеризації задається регламентом, але фактичні значення коливаються навколо цільового рівня. Маса упаковки має номінальне значення, але кожна окрема упаковка дещо відрізняється.

Такі явища називають **статистичними закономірностями**: окремий результат наперед точно невідомий, але для великої сукупності спостережень проявляється стійка структура — певна частота дефектів, середній рівень показника, форма розподілу, типова величина розсіювання.

4.1. Основні джерела мінливості

Джерело	Приклад
Сировина	Вологість зерна, жирність молока, цукристість фруктів
Обладнання	Коливання температури, швидкості конвеєра, дозування
Оператор	Відмінності у виконанні ручної операції
Метод вимірювання	Похибка ваг, титрування, сенсорного аналізу
Середовище	Температура і вологість повітря, сезонність
Неконтрольовані фактори	Сукупність малих випадкових впливів

Завдання статистичного аналізу — не «прибрати випадковість», а відокремити типову варіативність від істотних закономірностей і спеціальних причин.

5. Випадковий експеримент, подія та простір елементарних подій

5.1. Випадковий експеримент

Випадковим називають експеримент, який за однакових або близьких умов може завершуватися різними результатами. До початку експерименту відомий набір можливих результатів, але невідомо, який саме реалізується.

Приклади: випадковий відбір однієї упаковки з партії; контроль відповідності одиниці продукції нормативу; реєстрація кількості дефектних одиниць у вибірці; спостереження, чи відмовить вузол обладнання протягом заданого часу.

5.2. Елементарна подія

Окремий можливий результат випадкового експерименту називають **елементарною подією** і часто позначають ω.

5.3. Простір елементарних подій

Множину всіх можливих елементарних результатів позначають Ω. Для контролю однієї упаковки відносно дефекту:

$$\Omega = \{\text{придатна},\ \text{дефектна}\}.$$

Для двох упаковок:

$$\Omega = \{PP,\ PD,\ DP,\ DD\},$$

де P — придатна, D — дефектна.

5.4. Випадкова подія

Випадкова подія A — будь-яка підмножина Ω. Наприклад, подія «у двох перевірених упаковках є хоча б один дефект»:

$$A = \{PD,\ DP,\ DD\}.$$

5.5. Достовірна та неможлива події

$$P(\Omega) = 1, \qquad P(\varnothing) = 0.$$

6. Класичне означення ймовірності

Якщо випадковий експеримент має скінченну кількість **рівноможливих** елементарних результатів:

$$\boxed{P(A) = \frac{m}{n}},$$

де n — кількість усіх рівноможливих результатів, m — кількість результатів, сприятливих події A.

Приклад. Вибір дефектної упаковки

У партії 25 упаковок, із них 3 дефектні. За рівноймовірного випадкового вибору:

$$P(D) = \frac{3}{25} = 0.12.$$

Імовірність вибрати дефектну упаковку становить 12%.

6.1. Межі застосування

Класичне означення вимагає рівноможливості. Якщо дві виробничі лінії дають різні частки загального обсягу, події «продукт із лінії 1» і «продукт із лінії 2» не є рівноймовірними.

7. Статистичне означення ймовірності

У виробничих задачах імовірність часто оцінюють за повторними спостереженнями. Якщо подію A спостерігали m разів у n однотипних випробуваннях, її відносна частота:

$$\nu_n(A) = \frac{m}{n}.$$

За великої кількості стабільних однотипних випробувань:

$$\nu_n(A) \approx P(A).$$

Приклад. Емпіричний рівень дефектності

Перевірено 5000 упаковок, 82 дефектні:

$$\hat{p} = \frac{82}{5000} = 0.0164 = 1.64\%.$$

Це емпірична оцінка ймовірності дефекту, а не абсолютна стала «назавжди».

8. Основні властивості ймовірності

- $0 \le P(A) \le 1$;
- $P(\Omega) = 1$;
- $P(\varnothing) = 0$;
- якщо $A \subseteq B$, то $P(A) \le P(B)$;
- для протилежної події $\overline{A}$: $P(\overline{A}) = 1 - P(A)$.

8.1. Подія «хоча б одна»

У контролі якості часто простіше обчислити ймовірність відсутності дефектів, а потім перейти до події «хоча б один дефект»:

$$\boxed{P(X \ge 1) = 1 - P(X = 0)}.$$

9. Сумісні та несумісні події

Події A і B **несумісні**, якщо не можуть відбутися одночасно:

$$A \cap B = \varnothing.$$

Наприклад, відносно одного критерію одна конкретна упаковка не може одночасно бути «придатною» і «дефектною».

Події **сумісні**, якщо можуть відбутися одночасно. Наприклад, упаковка може одночасно мати недостатню масу та дефект маркування.

10. Теорема додавання ймовірностей

Для довільних двох подій:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}.$$

Для несумісних подій:

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

Приклад. Два типи дефектів

6% упаковок мають дефект маркування, 4% — дефект герметичності, 1% — обидва дефекти:

$$P(A \cup B) = 0.06 + 0.04 - 0.01 = 0.09.$$

Хоча б один із двох дефектів мають 9% упаковок.

11. Умовна ймовірність

Імовірність події може змінитися після отримання додаткової інформації:

$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}}, \qquad P(B) > 0.$$

Приклад. Дефектність у нічну зміну

20% продукції виготовляється у нічну зміну; 3% усієї продукції одночасно є нічною і дефектною:

$$P(D \mid N) = \frac{0.03}{0.20} = 0.15.$$

Серед продукції нічної зміни дефектність становить 15%.

12. Залежні та незалежні події

Події незалежні, якщо факт настання однієї не змінює ймовірності іншої:

$$P(A \mid B) = P(A).$$

Еквівалентна умова:

$$\boxed{P(A \cap B) = P(A)P(B)}.$$

Якщо ця умова не виконується, події залежні.

Важливо для вибіркового контролю. При послідовному відборі з невеликої партії **без повернення** склад партії змінюється після кожного відбору. Тому такі події зазвичай залежні.

13. Теорема множення ймовірностей

Для двох довільних подій:

$$\boxed{P(A \cap B) = P(A)P(B \mid A)}.$$

Для незалежних:

$$P(A \cap B) = P(A)P(B).$$

Приклад. Дві незалежні автоматизовані операції

Імовірність безпомилкового виконання першої операції — 0.99, другої — 0.98:

$$P = 0.99 \cdot 0.98 = 0.9702.$$

Імовірність успішного виконання обох операцій — 97.02%.

14. Формула повної ймовірності

Нехай $H_1, \dots, H_k$ — взаємовиключні гіпотези, що утворюють повну групу:

$$H_i \cap H_j = \varnothing \ (i \neq j), \qquad \bigcup_{i=1}^{k} H_i = \Omega.$$

Тоді для події A:

$$\boxed{P(A) = \sum_{i=1}^{k} P\,(H_i)P(A \mid H_i)}.$$

Приклад. Загальна дефектність двох ліній

Лінія 1 виготовляє 70% продукції з 2% дефектів, лінія 2 — 30% продукції з 5% дефектів:

$$P(D) = 0.70 \cdot 0.02 + 0.30 \cdot 0.05 = 0.029.$$

Загальний рівень дефектності становить 2.9%.

15. Формула Байєса

Формула повної ймовірності рухається від причин до наслідку. Формула Байєса дозволяє виконати зворотне міркування: якщо результат уже відомий, оцінити ймовірне джерело.

$$\boxed{P(H_i \mid A) = \frac{P(H_i)P(A \mid H_i)}{P(A)}}.$$

Продовження попереднього прикладу

Імовірність того, що знайдений дефектний продукт виготовлено на лінії 2:

$$P(H_2 \mid D) = \frac{0.30 \cdot 0.05}{0.029} \approx 0.5172.$$

Хоча лінія 2 виробляє лише 30% продукції, серед дефектних одиниць її апостеріорна частка перевищує 51%.

15.1. Апріорна й апостеріорна ймовірність

$P(H_i)$ — апріорна ймовірність до отримання нової інформації. $P(H_i \mid A)$ — апостеріорна ймовірність після врахування факту A.

Практичний зміст. Байєсівське оновлення дає змогу коректно поєднати «базову частоту» джерела з його специфічним ризиком дефекту.

16. Повторні незалежні випробування. Схема Бернуллі

Розглянемо n незалежних однотипних випробувань. У кожному подія A відбувається з однаковою ймовірністю p, а протилежна — з ймовірністю $q = 1 - p$. Ймовірність рівно k поява:

$$\boxed{P(X = k) = C_n^k p^k (1-p)^{n-k}}.$$

16.1. Умови схеми Бернуллі

1. Кількість випробувань n фіксована.
2. Кожне випробування має два результати відносно події: «подія відбулася / не відбулася».
3. Випробування незалежні.
4. Ймовірність p однакова в усіх випробуваннях.

16.2. Рівно k дефектів

Якщо $p = 0.03$, $n = 20$, то для двох дефектних:

$$P(X = 2) = C_{20}^2 (0.03)^2 (0.97)^{18}.$$

16.3. Жодного дефекту

$$P(X = 0) = (1-p)^n.$$

16.4. Хоча б один дефект

$$\boxed{P(X \geq 1) = 1 - (1-p)^n}.$$

Приклад. Ефективність контролю

Нехай $p = 0.04$, $n = 12$. Тоді:

$$P(X \geq 1) = 1 - 0.96^{12} \approx 0.3873.$$

Навіть при реальному рівні дефектності 4% контроль 12 незалежних одиниць виявить хоча б один дефект приблизно лише у 38.7% випадків.

17. Ймовірність і технологічний ризик

У технологічному контексті ризик часто пов'язують із імовірністю небажаної події та тяжкістю її наслідків. Теорія ймовірностей дає інструмент для оцінювання першої складової.

Приклади небажаних подій:

- порушення герметичності упаковки;
- перевищення допустимого мікробіологічного показника;

- відмова холодильного обладнання;
- недостатня теплова обробка;
- помилка дозування інгредієнта;
- потрапляння дефектної одиниці до споживача.

18. Типові помилки у ймовірнісному мисленні

1. **Ототожнення частоти й точної ймовірності.** Відносна частота є вибірковою оцінкою.
2. **Механічне припущення незалежності.** Події можуть бути пов'язані технологічно або способом відбору.
3. **Змішування $P(A \mid B)$ і $P(B \mid A)$.** Це різні величини.
4. **Ігнорування базової частоти.** У формулі Байєса важлива частка джерела $P(H_i)$.
5. **Хибне тлумачення «не виявлено дефектів».** Це не доводить відсутність дефектів у генеральній сукупності.
6. **Біноміальна модель для відбору без повернення.** Якщо партія невелика, незалежність порушується.

19. Зв'язок із практичною роботою №1

Практична робота №1 безпосередньо реалізує матеріал лекції через чотири послідовні блоки:

- комбінаторні моделі формування вибірки;
- вибірковий контроль кінцевої партії без повернення;
- повна ймовірність та Байєс для двох виробничих ліній;
- схема Бернуллі для повторного контролю незалежних одиниць.

Головна мета — навчитися **вибирати правильну модель**, а не лише підставляти числа у готову формулу.

20. Підсумок лекції

Логіка теми:

випадковий експеримент → подія → ймовірність → умовність/незалежність → повна ймовірність → Байєс

Теорія ймовірностей створює математичну основу, на якій надалі будуються випадкові величини, статистичне оцінювання, перевірка гіпотез та моделювання.

21. Контрольні питання

1. Чому в харчовому виробництві навіть за незмінного регламенту результати окремих вимірювань відрізняються?
2. Що називають випадковим експериментом, елементарною подією і простором елементарних подій?
3. За яких умов можна використовувати класичне означення ймовірності?
4. У чому полягає статистичний підхід до оцінювання ймовірності?
5. Чим сумісні події відрізняються від несумісних?
6. У чому різниця між незалежністю та несумісністю подій?

7. Сформулюйте теорему додавання ймовірностей.
8. Що таке умовна ймовірність?
9. Сформулюйте теорему множення.
10. Коли можна використовувати $P(A \cap B) = P(A)P(B)$?
11. У чому сенс формули повної ймовірності?
12. Яку задачу розв'язує формула Байєса?
13. Чим апріорна ймовірність відрізняється від апостеріорної?
14. Які умови схеми Бернуллі?
15. Чому подію «хоча б один дефект» зручно рахувати через протилежну?
16. Чому відсутність дефектів у невеликій вибірці не доводить відсутності дефектів у потоці?
17. Наведіть приклад технологічної задачі, у якій події є залежними.
18. Наведіть приклад застосування формули Байєса у контролі якості.

Лекція 2

Випадкові величини та їх моделювання у технологічних процесах

1. Мета лекції та результати навчання

Мета лекції — перейти від опису окремих випадкових подій до математичного моделювання числових результатів випадкових експериментів і показати, як випадкові величини використовуються для опису технологічних параметрів, тривалості операцій, часу безвідмовної роботи обладнання та інших виробничих показників.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати поняття випадкової величини як числової функції результату випадкового експерименту;
- розрізняти дискретні та неперервні випадкові величини;
- задавати закон розподілу дискретної випадкової величини;
- пояснювати зміст функції розподілу та густини ймовірності;
- обчислювати математичне сподівання, дисперсію та середнє квадратичне відхилення;
- розпізнавати рівномірну й показникову моделі та оцінювати межі їх застосування;
- інтерпретувати параметри розподілу у технологічному контексті;
- відрізняти математичну модель від фактичних виробничих даних і розуміти необхідність перевірки припущень.

2. План лекції

1. Поняття випадкової величини.
2. Дискретні та неперервні випадкові величини.
3. Закон розподілу дискретної випадкової величини.
4. Функція розподілу та її властивості.
5. Густина розподілу неперервної випадкової величини.
6. Математичне сподівання, дисперсія та стандартне відхилення.
7. Рівномірний розподіл.
8. Показниковий розподіл.
9. Технологічні приклади та межі застосування імовірнісної моделі.

3. Від випадкової події до випадкової величини

У попередній лекції випадковий результат описувався через події: «дефект є/немає», «обладнання відмовило/не відмовило». У багатьох задачах результат має числову форму: маса упаковки, кислотність, температура, кількість дефектів у вибірці, час до відмови, тривалість операції.

Для математичного опису таких результатів використовують **випадкову величину**.

3.1. Означення

Випадкова величина X — числова функція, визначена на просторі елементарних подій:

$$\boxed{X: \Omega \to \mathbb{R}}.$$

Кожному елементарному результату $\omega \in \Omega$ вона ставить у відповідність певне числове значення $X(\omega)$.

Приклад. Кількість дефектів у трьох упаковках

Нехай X — кількість дефектних упаковок серед трьох перевірених. Тоді:

$$X(PPP) = 0, \quad X(DPP) = 1, \quad X(DDP) = 2, \quad X(DDD) = 3.$$

Різні елементарні результати можуть давати одне й те саме числове значення випадкової величини.

4. Дискретні випадкові величини

Випадкову величину називають **дискретною**, якщо множина її можливих значень є скінченною або зліченною.

Приклади:

- кількість дефектних упаковок у вибірці;
- кількість відмов обладнання за зміну;
- кількість зразків, що не пройшли контроль;
- кількість повторних вимірювань до отримання прийнятного результату.

4.1. Закон розподілу дискретної випадкової величини

Закон розподілу задає всі можливі значення x_i та їх імовірності $p_i = P(X = x_i)$.

X	x_1	x_2	…	x_k
P	p_1	p_2	…	p_k

Обов'язково:

$$p_i \geq 0, \qquad \sum_i p_i = 1.$$

Приклад. Закон розподілу кількості дефектів

Перевіряють дві незалежні упаковки, ймовірність дефекту кожної $p = 0.1$. X — кількість дефектних упаковок.

$$P(X = 0) = 0.9^2 = 0.81.$$

$$P(X = 1) = 2 \cdot 0.1 \cdot 0.9 = 0.18.$$

$$P(X = 2) = 0.1^2 = 0.01.$$

x	**0**	**1**	**2**
P(X=x)	0.81	0.18	0.01

Перевірка: $0.81 + 0.18 + 0.01 = 1$.

5. Неперервні випадкові величини

Випадкову величину називають **неперервною**, якщо вона може набувати будь-якого значення в певному числовому інтервалі.

Приклади:

- маса упаковки;
- температура продукту;
- масова частка вологи;
- тривалість технологічної операції;
- час до відмови обладнання;
- концентрація речовини.

Для неперервної випадкової величини ймовірність одного точного значення дорівнює нулю:

$$P(X = x_0) = 0.$$

Тому зміст мають імовірності потрапляння в інтервали, наприклад $P(a < X < b)$.

6. Функція розподілу

Функція розподілу визначається для будь-якого дійсного x:

$$\boxed{F(x) = P(X \le x)}.$$

Вона показує ймовірність того, що випадкова величина набуде значення, не більшого за x.

6.1. Основні властивості

- $0 \le F(x) \le 1$;
- $F(x)$ не спадає;
- $\lim_{x \to -\infty} F(x) = 0$;
- $\lim_{x \to +\infty} F(x) = 1$;
- $P(a < X \le b) = F(b) - F(a)$.

6.2. Для дискретної величини

Функція розподілу має ступінчастий вигляд. У точці можливого значення x_i вона збільшується на $P(X = x_i)$.

7. Густина розподілу неперервної величини

Якщо функція розподілу диференційовна, її похідну називають густиною ймовірності:

$$\boxed{f(x) = F'(x)}.$$

І навпаки:

$$F(x) = \int_{-\infty}^{x} f\,(t)\,dt.$$

7.1. Властивості густини

- $f(x) \geq 0$;
- повна площа під кривою дорівнює 1;

$$\int_{-\infty}^{+\infty} f\,(x)\,dx = 1.$$

Імовірність потрапляння в інтервал:

$$\boxed{P(a < X < b) = \int_{a}^{b} f\,(x)\,dx}.$$

Для неперервної величини:

$$P(a < X < b) = P(a \leq X \leq b).$$

Типова помилка. Не можна трактувати значення густини $f(x_0)$ як ймовірність $P(X = x_0)$. Імовірність для неперервної величини — це площа під густиною на інтервалі.

8. Числові характеристики випадкової величини

Повний закон розподілу містить найбільше інформації, але для практичного опису часто потрібні компактні характеристики центрального рівня та розсіювання.

8.1. Математичне сподівання дискретної величини

$$\boxed{M(X) = E(X) = \sum_{i} x_i\,p_i}.$$

Це ймовірнісно зважене середнє можливих значень.

Приклад. Очікувана кількість дефектів

Для попереднього закону розподілу:

$$E(X) = 0 \cdot 0.81 + 1 \cdot 0.18 + 2 \cdot 0.01 = 0.20.$$

Це не означає, що в конкретній вибірці буде «0.2 дефекту». У довгій серії таких двоелементних перевірок середня кількість дефектів на одну перевірку наближатиметься до 0.2.

8.2. Для неперервної величини

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx},$$

якщо інтеграл існує.

8.3. Основні властивості

- $E(c) = c$;
- $E(cX) = cE(X)$;
- $E(X + Y) = E(X) + E(Y)$;
- для незалежних X і Y: $E(XY) = E(X)E(Y)$.

9. Дисперсія

Математичне сподівання характеризує центр, але не показує, наскільки сильно можливі значення відхиляються від нього. Для цього використовують дисперсію:

$$\boxed{D(X) = Var(X) = E\left[\left(X - E(X)\right)^2\right]}.$$

Еквівалентна обчислювальна формула:

$$\boxed{Var(X) = E(X^2) - [E(X)]^2}.$$

9.1. Для дискретної величини

$$Var(X) = \sum_i (x_i - \mu)^2 p_i, \qquad \mu = E(X).$$

Приклад

Для $X = 0,1,2$ з ймовірностями 0.81, 0.18, 0.01:

$$E(X) = 0.20.$$

$$E(X^2) = 0^2 \cdot 0.81 + 1^2 \cdot 0.18 + 2^2 \cdot 0.01 = 0.22.$$

$$Var(X) = 0.22 - 0.20^2 = 0.18.$$

10. Середнє квадратичне відхилення

$$\boxed{\sigma(X) = \sqrt{Var(X)}}.$$

Стандартне відхилення має ті самі одиниці вимірювання, що й сама випадкова величина, тому його простіше інтерпретувати предметно.

У попередньому прикладі:

$$\sigma = \sqrt{0.18} \approx 0.424.$$

10.1. Властивості дисперсії

- $Var(c) = 0$;
- $Var(cX) = c^2 Var(X)$;
- $Var(X + a) = Var(X)$;
- для незалежних X і Y: $Var(X + Y) = Var(X) + Var(Y)$.

11. Випадкова величина і фактична вибірка

Імовірнісна модель описує генеральний механізм, а вибірка є набором реалізацій цієї випадкової величини:

$$X_1, X_2, \dots, X_n \quad \rightarrow \quad x_1, x_2, \dots, x_n.$$

Наприклад, «маса упаковки» як випадкова величина має певний невідомий розподіл; значення 499.8, 500.3, 500.1 г — конкретні реалізації. На наступних лекціях ми вчитимемося оцінювати параметри генеральної моделі за такими реалізаціями.

12. Рівномірний розподіл

Неперервна випадкова величина має рівномірний розподіл на $[a, b]$, якщо однакові за довжиною підінтервали мають однакову ймовірність.

$$X \sim U(a, b).$$

12.1. Густина

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & x < a \text{ або } x > b. \end{cases}$$

12.2. Функція розподілу

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b. \end{cases}$$

12.3. Числові характеристики

$$\boxed{E(X) = \frac{a+b}{2}}.$$

$$\boxed{Var(X) = \frac{(b-a)^2}{12}}.$$

$$\boxed{\sigma = \frac{b-a}{\sqrt{12}}}.$$

12.4. Імовірність інтервалу

Якщо $[c, d] \subseteq [a, b]$, то:

$$P(c < X < d) = \frac{d-c}{b-a}.$$

Приклад. Момент запуску технологічної операції

Операція автоматично запускається в будь-який момент між 10.0 і 12.0 хв після попередньої стадії, і всі моменти в цьому інтервалі вважаються однаково можливими:

$$X \sim U(10{,}12).$$

Імовірність запуску між 10.5 і 11.0 хв:

$$P(10.5 < X < 11.0) = \frac{0.5}{2} = 0.25.$$

Математичне сподівання:

$$E(X) = \frac{10 + 12}{2} = 11 \text{ хв.}$$

12.5. Коли рівномірна модель доречна

- відомі жорсткі межі a і b;
- немає причин вважати якусь частину інтервалу більш імовірною;
- модель описує спеціальний механізм випадкового вибору або використовується як перше наближення.

Для більшості природних технологічних вимірювань рівномірність не слід припускати автоматично.

13. Показниковий розподіл

Показниковий розподіл часто використовують для моделювання часу очікування до випадкової події за умови приблизно сталої інтенсивності її виникнення.

$$X \sim Exp(\lambda), \qquad \lambda > 0.$$

13.1. Густина

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

13.2. Функція розподілу

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

13.3. Функція надійності

Імовірність того, що подія не відбудеться до часу t:

$$\boxed{P(X > t) = e^{-\lambda t}}.$$

У задачах надійності цю функцію часто позначають $R(t)$.

13.4. Числові характеристики

$$\boxed{E(X) = \frac{1}{\lambda}}.$$

$$\boxed{Var(X) = \frac{1}{\lambda^2}}.$$

$$\boxed{\sigma = \frac{1}{\lambda}}.$$

Для показникового розподілу середнє і стандартне відхилення чисельно однакові.

Приклад. Час безвідмовної роботи обладнання

Середній час до відмови вузла становить 200 год. За показникової моделі:

$$E(X) = 200 = \frac{1}{\lambda}, \qquad \lambda = 0.005 \text{ год}^{-1}.$$

Імовірність безвідмовної роботи понад 100 год:

$$P(X > 100) = e^{-0.005 \cdot 100} = e^{-0.5} \approx 0.6065.$$

Імовірність відмови протягом перших 100 год:

$$P(X \le 100) = 1 - e^{-0.5} \approx 0.3935.$$

13.5. Властивість відсутності пам'яті

$$\boxed{P(X > s + t \mid X > s) = P(X > t)}.$$

За цією моделлю обладнання, яке вже пропрацювало s годин, має такий самий розподіл додаткового часу до відмови, як новий об'єкт.

Це сильне припущення. Воно відповідає приблизно сталій інтенсивності відмов, але не описує виражене старіння або зношування, коли ризик відмови з часом зростає.

14. Моделювання тривалості технологічних операцій

Тривалість операції може мати різні розподіли залежно від фізичного механізму. Рівномірна модель доречна, якщо час гарантовано лежить у вузькому інтервалі і всі значення в ньому однаково можливі. Показникова — для часу очікування до випадкової події зі сталою інтенсивністю.

Реальні технологічні тривалості нерідко мають іншу форму. Тому вибір моделі має спиратися на предметний механізм і фактичні дані, а не лише на зручність формули.

15. Моделювання часу безвідмовної роботи

У теорії надійності випадковими величинами можуть бути час до першої відмови, кількість відмов за зміну, тривалість ремонту, час безвідмовної роботи після ремонту.

Показникова модель часто використовується як базова завдяки простоті й сталій інтенсивності відмов. Але для обладнання з ранніми відмовами або старінням можуть бути потрібні інші моделі, наприклад розподіл Вейбулла.

Методичний акцент. Важливо не запам'ятати максимальну кількість розподілів, а навчитися ставити питання: яке припущення про механізм процесу закладено у вибраній моделі?

16. Моделювання терміну зберігання продукції

Час до досягнення критичного рівня показника якості або безпеки також є випадковою величиною. Його розподіл залежить від температури, пакування, складу продукту, початкового мікробіологічного навантаження та багатьох інших факторів.

Показникова модель може слугувати лише спрощеною моделлю часу до випадкової події зі сталою інтенсивністю. Для реального shelf-life це припущення потребує перевірки.

У практичній роботі №7 буде використано інший підхід — регресійна модель зміни показника у часі та визначення моменту досягнення критичного рівня.

17. Поняття імовірнісної моделі

Імовірнісна модель — математичний опис випадкового механізму, який задає можливі значення результату та закон їх імовірностей.

Модель не є тотожною реальному процесу. Вона є спрощенням, яке виділяє суттєві риси для конкретної задачі.

17.1. Основні елементи побудови моделі

1. Визначити випадкову величину X.
2. Визначити множину її можливих значень.
3. Запропонувати закон розподілу.
4. Визначити або оцінити параметри розподілу.
5. Перевірити узгодженість моделі з даними й предметним механізмом.
6. Використати модель для розрахунку ймовірностей або прогнозування.

18. Як обирають модель у реальному дослідженні

Правильна послідовність:

предметний механізм → тип змінної → графік даних → кандидатна модель → перевірка припущень → вис

Для обґрунтування моделі використовують:

- знання технологічного механізму;
- область можливих значень;
- симетрію або асиметрію фактичних даних;
- гістограму й емпіричну функцію розподілу;
- Q–Q plot;
- формальні критерії узгодженості;
- аналіз залишків для регресійних моделей.

19. Межі застосування імовірнісної моделі

Будь-яка модель спирається на припущення. Помилкове припущення може дати математично правильний розрахунок, але технологічно хибний висновок.

19.1. Типові проблеми

1. **Нестабільність процесу.** Якщо параметри змінюються в часі, одна стаціонарна модель може бути непридатною.
2. **Змішування різних сукупностей.** Дані з різних ліній, змін або рецептур можуть утворювати суміш розподілів.
3. **Залежність спостережень.** Послідовні значення можуть бути автокорельованими.
4. **Цензурування.** Час до відмови може бути невідомим для об'єктів, які не відмовили до завершення спостереження.
5. **Викиди та помилки.** Нетипові точки можуть суттєво впливати на оцінювання параметрів.

6. **Неправильна форма розподілу.** Просту модель не слід вибирати лише через простоту формули.

20. Модель і дані: різні рівні опису

Поняття	Зміст
Випадкова величина X	Теоретичний числовий результат випадкового механізму
Закон розподілу	Теоретичне правило розподілу ймовірностей
Параметри μ, σ, λ тощо	Характеристики генеральної моделі
Вибірка $x_1,\ldots,x_n$	Фактично отримані реалізації
Статистики $\bar{x}$, s^2, s	Характеристики конкретної вибірки
Статистичне оцінювання	Перехід від вибірки до невідомих параметрів моделі

21. Зв'язок із наступними темами курсу

Математичне сподівання та дисперсія надалі переходять у вибіркове середнє та вибіркову дисперсію. Закон розподілу стає основою для перевірки нормальності, побудови довірчих інтервалів і вибору статистичного критерію.

На наступній лекції буде детально розглянуто **нормальний розподіл**, стандартизацію та правило трьох сигм. Тому нормальний закон тут свідомо не розгортається окремим блоком.

22. Типові помилки

1. **Плутати випадкову величину з її конкретним значенням.** X — модель, $x = 13.8$ — одна реалізація.
2. **Вважати $E(X)$ найімовірнішим значенням.** Сподівання — довгострокове середнє, а не обов'язково мода.
3. **Плутати дисперсію і стандартне відхилення.** Дисперсія має квадрат одиниць вимірювання.
4. **Вважати $f(x)$ ймовірністю точного значення.** Для неперервної величини $P(X = x) = 0$.
5. **Механічно підбирати розподіл за назвою задачі.** Потрібна перевірка припущень.
6. **Екстраполювати модель без обґрунтування.** За межами перевіреного діапазону поведінка процесу може змінитися.

23. Підсумок лекції

випадковий експеримент → X → закон розподілу → $F(x), f(x)$ → $E(X), Var(X), \sigma$ → конкретна модель →

Випадкова величина перетворює випадковий результат на числовий об'єкт, а закон розподілу описує структуру його можливих значень. Математичне сподівання характеризує центральний рівень, дисперсія — мінливість. Рівномірний і показниковий розподіли є корисними моделями лише тоді, коли їх припущення відповідають механізму процесу.

24. Контрольні питання

1. Що називають випадковою величиною?
2. Чим дискретна випадкова величина відрізняється від неперервної?
3. Як задають закон розподілу дискретної випадкової величини?
4. Які умови повинні виконувати ймовірності у дискретному законі розподілу?

5. Що показує функція розподілу $F(x)$?
6. Назвіть основні властивості функції розподілу.
7. Що таке густина ймовірності?
8. Чому для неперервної величини $P(X = x_0) = 0$?
9. Як обчислюють математичне сподівання дискретної випадкової величини?
10. Який предметний зміст математичного сподівання?
11. Що характеризує дисперсія?
12. Чому стандартне відхилення інтерпретувати простіше, ніж дисперсію?
13. Які параметри задають рівномірний розподіл?
14. За яких умов рівномірний розподіл може бути адекватною моделлю?
15. Запишіть густину та функцію розподілу показникової величини.
16. Що означає параметр λ?
17. У чому полягає властивість відсутності пам'яті?
18. Чому показникова модель не завжди придатна для часу безвідмовної роботи обладнання?
19. Чим імовірнісна модель відрізняється від фактичних виробничих даних?
20. Які джерела інформації слід використати для обґрунтування вибору закону розподілу?

Лекція 3

Основні закони розподілу. Нормальний закон та його застосування у технологічних дослідженнях

1. Мета лекції та результати навчання

Мета лекції — сформувати цілісне уявлення про роль законів розподілу у статистичному моделюванні технологічних показників та детально розглянути нормальний розподіл як одну з центральних моделей прикладної статистики.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати, що означає «закон розподілу» випадкової величини;
- розрізняти параметри розподілу і вибіркові оцінки цих параметрів;
- характеризувати рівномірний, показниковий і нормальний розподіли та межі їх застосування;
- записувати густину нормального розподілу та інтерпретувати параметри μ і σ;
- виконувати стандартизацію випадкової величини;
- використовувати стандартний нормальний розподіл для обчислення ймовірностей;
- пояснювати правило 1σ, 2σ, 3σ;
- відрізняти статистичні межі природної мінливості процесу від технологічних меж допуску;
- пояснювати роль центральної граничної теореми у статистичному аналізі;
- критично оцінювати припущення про нормальність виробничих даних.

2. План лекції

1. Закон розподілу як імовірнісна модель.
2. Коротке узагальнення рівномірного та показникового розподілів.
3. Нормальний розподіл: означення, параметри і властивості.
4. Стандартний нормальний розподіл і стандартизація.
5. Обчислення ймовірностей для нормального розподілу.
6. Правило однієї, двох і трьох сигм.
7. Нормальний розподіл і технологічні допуски.
8. Центральна гранична теорема: статистичний зміст.
9. Межі застосування нормальної моделі.

3. Закон розподілу як математична модель

У попередній лекції було введено випадкову величину та її основні числові характеристики. Наступний крок — описати не лише середній рівень і мінливість, а **повну структуру можливих значень**. Саме це робить закон розподілу.

Закон розподілу відповідає на запитання:

- які значення може набувати випадкова величина;
- наскільки ймовірними є різні області значень;
- де зосереджена основна маса результатів;
- наскільки ймовірні крайні або нетипові значення.

Важливо. Математичний розподіл є моделлю генеральної сукупності, а гістограма — графічним описом конкретної вибірки. Їх не слід ототожнювати.

4. Параметри розподілу та їх оцінювання

Більшість законів розподілу задається певними параметрами. Наприклад:

Розподіл	Основні параметри	Зміст
Рівномірний $U(a,b)$ a,b Нижня і в	ерхня межі	
Показниковий $Exp(\lambda)$ λ	Інтенсивність події	
Нормальний $N(\mu,\sigma^2)$ $,\si	gma^2$ Центральний рівень	і мінливість

У реальному дослідженні параметри генерального розподілу невідомі. Їх оцінюють за вибіркою:

$$\mu \approx \bar{x}, \qquad \sigma^2 \approx s^2, \qquad \sigma \approx s.$$

Саме цей перехід від невідомих генеральних параметрів до вибіркових оцінок буде детально розглянуто у наступних лекціях.

5. Коротке узагальнення рівномірного розподілу

Рівномірна модель використовується, коли випадкова величина обмежена інтервалом $[a,b]$ і немає підстав вважати одну частину інтервалу більш імовірною за іншу.

$$X \sim U(a,b).$$

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{інакше.} \end{cases}$$

Числові характеристики:

$$E(X) = \frac{a+b}{2}, \qquad Var(X) = \frac{(b-a)^2}{12}.$$

У харчовій технології рівномірний розподіл може використовуватися для моделювання випадкового моменту запуску операції в заданому інтервалі, випадкового положення об'єкта або похибки округлення.

6. Коротке узагальнення показникового розподілу

Показникова модель застосовується до часу очікування до випадкової події за умови приблизно сталої інтенсивності її виникнення.

$$X \sim Exp(\lambda).$$

$$f(x) = \lambda e^{-\lambda x}, \qquad x \ge 0.$$

$$F(x) = 1 - e^{-\lambda x}, \qquad x \ge 0.$$

$$E(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}.$$

Типові застосування: час до відмови вузла, час очікування випадкової події, інтервали між подіями.

7. Нормальний розподіл

Нормальний розподіл є однією з найважливіших моделей математичної статистики. Він часто виникає тоді, коли результат формується як сума великої кількості малих незалежних або слабко залежних факторів.

Позначення:

$$\boxed{X \sim N(\mu, \sigma^2)}.$$

Густина:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}.$$

7.1. Параметр μ

μ — математичне сподівання, яке задає положення центра нормального розподілу.

$$E(X) = \mu.$$

7.2. Параметр σ

σ — стандартне відхилення, яке задає ширину розподілу. Чим більше σ, тим сильніше розсіювання.

$$Var(X) = \sigma^2.$$

8. Основні властивості нормального розподілу

- Густина симетрична відносно $x = \mu$.
- Середнє, медіана та мода збігаються.
- Крива має дзвоноподібну форму.
- Теоретично можливі всі дійсні значення від $-\infty$ до $+\infty$.
- Площа під кривою дорівнює 1.
- Зі збільшенням відстані від μ густина швидко зменшується.

8.1. Вплив μ

Зміна μ зсуває криву вправо або вліво, не змінюючи її форми.

8.2. Вплив σ

Зменшення σ робить криву вужчою і вищою; збільшення — ширшою і нижчою. Загальна площа завжди залишається 1.

9. Чому нормальний розподіл часто зустрічається у виробництві

Багато технологічних показників формуються під впливом великої кількості малих факторів:

- варіація властивостей сировини;
- похибка дозатора;
- мікроколивання температури;
- похибка вимірювання;
- незначні відмінності в умовах середовища;
- дрібні неконтрольовані фактори.

Якщо жоден окремий фактор не домінує, сумарний результат часто набуває приблизно симетричного дзвоноподібного розподілу.

Але: «часто» не означає «завжди». Вологість, тривалість зберігання, мікробіологічні показники, концентрації та час до відмови можуть мати виражено ненормальні розподіли.

10. Стандартний нормальний розподіл

Особливий випадок нормального розподілу:

$$\boxed{Z \sim N(0,1)}.$$

Його математичне сподівання дорівнює 0, а стандартне відхилення — 1.

Стандартний нормальний розподіл використовується як універсальна шкала для порівняння нормальних величин із різними одиницями вимірювання.

11. Стандартизація

Нехай:

$$X \sim N(\mu, \sigma^2).$$

Тоді стандартизована величина:

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

має стандартний нормальний розподіл:

$$Z \sim N(0,1).$$

Значення z показує, на скільки стандартних відхилень конкретне значення віддалене від середнього.

Приклад. Стандартизація маси упаковки

Нехай маса упаковки має приблизно нормальний розподіл із:

$$\mu = 500 \text{ г}, \qquad \sigma = 1.5 \text{ г}.$$

Для упаковки масою 503 г:

$$z = \frac{503 - 500}{1.5} = 2.$$

Маса 503 г знаходиться на дві стандартні сигми вище середнього.

12. Функція стандартного нормального розподілу

Для стандартної нормальної величини використовують функцію:

$$\Phi(z) = P(Z \le z).$$

Вона показує накопичену ймовірність ліворуч від точки z.

Через симетрію:

$$\Phi(-z) = 1 - \Phi(z).$$

13. Обчислення ймовірностей

Для довільної нормальної величини:

$$P(X \le x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Для інтервалу:

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Приклад. Ймовірність перевищення порогового значення

Нехай $X \sim N(500, 1.5^2)$. Знайдемо ймовірність маси більше 503 г.

$$z = \frac{503 - 500}{1.5} = 2.$$

$$P(X > 503) = P(Z > 2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228.$$

Отже, приблизно 2.28% упаковок за такої моделі матимуть масу понад 503 г.

14. Правило однієї, двох і трьох сигм

Для нормального розподілу:

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0.6827.$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.9545.$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973.$$

Тобто приблизно:

Інтервал	Частка значень	Поза інтервалом
$\mu \pm \sigma$ 68.27%	31.73%	
$\mu \pm 2\sigma$ 95.45%	4.55%	
$\mu \pm 3\sigma$ 99.73%	0.27%	

14.1. Практичний зміст правила трьох сигм

Якщо процес стабільний і приблизно нормальний, значення за межами $\mu \pm 3\sigma$ є дуже рідкісними. Саме ця ідея лежить в основі класичних контрольних меж Шухарта.

Зв'язок із ПР2. Контрольні межі на I-карті фактично оцінюють діапазон приблизно $\mu \pm 3\sigma$, але σ оцінюється за рухомими розмахами.

15. Нормальний розподіл і технологічні допуски

Потрібно чітко розрізняти два типи меж:

- **статистичні межі**, які визначаються властивостями процесу;
- **межі специфікації**, які задаються нормативом або технологічним регламентом.

Позначення:

$$LSL = \text{нижня межа специфікації}, \qquad USL = \text{верхня межа специфікації}.$$

Статистично стабільний процес може мати занадто велике σ і виробляти значну частку продукції поза специфікацією. І навпаки, процес може поки що давати продукцію в межах допуску, але мати сигнал нестабільності.

$$\boxed{\text{стабільність процесу} \neq \text{відповідність продукції нормативу}}.$$

16. Приклад: частка продукції поза допуском

Нехай маса продукту:

$$X \sim N(500, 1.5^2),$$

а нормативний діапазон:

$$497 \leq X \leq 503.$$

Межі відповідають:

$$z_1 = \frac{497 - 500}{1.5} = -2, \qquad z_2 = \frac{503 - 500}{1.5} = 2.$$

Отже:

$$P(497 \leq X \leq 503) = P(-2 \leq Z \leq 2) \approx 0.9545.$$

Поза нормативним діапазоном теоретично буде:

$$1 - 0.9545 = 0.0455 = 4.55\%.$$

Навіть стабільний нормальний процес може бути технологічно неприйнятним, якщо природне розсіювання надто велике відносно допуску.

17. Центральна гранична теорема: основна ідея

Одна з причин фундаментальної ролі нормального розподілу — центральна гранична теорема (ЦГТ).

У спрощеній формі: якщо $X_1, \dots, X_n$ — незалежні однаково розподілені випадкові величини з математичним сподіванням μ і скінченною дисперсією σ^2, то при достатньо великому n розподіл вибіркового середнього наближається до нормального:

$$\boxed{\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)}.$$

17.1. Стандартна похибка середнього

Стандартне відхилення розподілу вибіркового середнього:

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}.$$

У статистичній практиці при невідомому σ його оцінюють за вибіркою:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}.$$

Це безпосередньо підводить до теми довірчих інтервалів.

17.2. Що ЦГТ не означає

ЦГТ не означає, що будь-які сирі дані стають нормальними при збільшенні вибірки. Нормальним наближається розподіл вибіркового середнього, а не обов'язково розподіл самих спостережень.

18. Коли припущення про нормальність є важливим

Нормальність особливо важлива для:

- класичних довірчих інтервалів для дисперсії;
- F-критерію рівності дисперсій;
- t-критеріїв у малих вибірках;
- ANOVA;
- аналізу залишків регресійної моделі.

При великих вибірках деякі процедури стають стійкішими до помірних відхилень від нормальності, але це не скасовує потреби аналізувати форму розподілу.

19. Як перевіряють нормальність на практиці

Доцільно поєднувати кілька інструментів:

- гістограму;
- boxplot;
- Q–Q plot;
- Shapiro–Wilk test;
- предметний аналіз причин асиметрії та викидів.

У цьому курсі формальна перевірка Shapiro–Wilk детально застосовуватиметься в ПР4.

19.1. Інтерпретація Shapiro–Wilk

Типова нульова гіпотеза:

$$H_0\text{:дані не суперечать нормальній моделі.}$$

Якщо $p > 0.05$, це означає лише, що немає достатніх підстав відхилити нормальність. Це не є доказом того, що розподіл «точно нормальний».

20. Типові помилки

1. **«Якщо гістограма схожа на дзвін — розподіл нормальний».** Гістограма залежить від обсягу вибірки та вибору інтервалів.
2. **«p>0.05 доводить нормальність».** Ні, лише не дає підстав її відхилити.
3. **Ототожнення $\mu \pm 3\sigma$ з нормативним допуском.** Це статистичні межі, а не специфікація.
4. **«Усі вимірювання у виробництві нормальні».** Час, концентрації, дефектність, мікробіологічні показники можуть мати іншу форму.
5. **Неправильна інтерпретація стандартизації.** z-score не змінює порядок спостережень, а лише переводить їх у шкалу стандартних відхилень.
6. **Хибне трактування ЦГТ.** Вона стосується розподілу середнього, а не автоматичної нормалізації сирих даних.

21. Зв'язок із практичною роботою №2

ПР2 не вимагає формально доводити нормальність, але матеріал цієї лекції є фундаментальним для розуміння:

- форми гістограми;
- центра та розсіювання;
- значення стандартного відхилення;
- правила трьох сигм;
- походження контрольних меж Шухарта;
- різниці між природною статистичною мінливістю і нормативним допуском.

22. Підсумок лекції

$$\text{закон розподілу} \to N(\mu, \sigma^2) \to Z = \frac{X - \mu}{\sigma} \to \Phi(z) \to 1\sigma, 2\sigma, 3\sigma \to \text{статистичні межі} \to \text{контроль процесу}.$$

Нормальний розподіл є базовою, але не універсальною моделлю. Його сила полягає у простій інтерпретації параметрів, стандартизації, зв'язку з правилом трьох сигм і центральною граничною теоремою.

23. Контрольні питання

1. Що означає закон розподілу випадкової величини?
2. Які параметри задають нормальний розподіл?
3. Який зміст мають μ та σ?

4. Як змінюється крива нормального розподілу при збільшенні σ?
5. Що таке стандартний нормальний розподіл?
6. Для чого використовується стандартизація?
7. Що означає $z = 2$ для конкретного спостереження?
8. Як через $\Phi(z)$ знайти ймовірність потрапляння в інтервал?
9. Яка частка значень нормального розподілу знаходиться в межах $\mu \pm \sigma$?
10. Яка частка знаходиться в межах $\mu \pm 3\sigma$?
11. Чому правило трьох сигм важливе для контрольних карт?
12. У чому різниця між контрольними межами та межами специфікації?
13. Чи може стабільний процес виготовляти продукцію поза нормативом?
14. У чому полягає основна ідея центральної граничної теореми?
15. Чи означає ЦГТ, що сирі дані стають нормальними при великому n?
16. Які інструменти використовують для оцінювання нормальності даних?
17. Як правильно трактувати результат Shapiro–Wilk з $p > 0.05$?
18. Наведіть приклад технологічного показника, для якого нормальна модель може бути непридатною.

Лекція 4

Вибірковий метод, описова статистика та візуалізація технологічних даних

1. Мета лекції та результати навчання

Мета лекції — сформувати практичне розуміння вибіркового методу, основних описових статистик і способів графічного представлення даних, а також показати, як статистичне спостереження переходить у контроль стабільності технологічного процесу.

Після опрацювання лекції здобувач повинен уміти:

- розрізняти генеральну сукупність і вибірку;
- пояснювати поняття репрезентативності та випадкового відбору;
- формувати варіаційний та інтервальний ряди;
- обчислювати середнє, медіану, моду, розмах, дисперсію, стандартне відхилення, коефіцієнт варіації, квартилі та IQR;
- вибирати адекватний спосіб візуалізації даних;
- інтерпретувати гістограму, boxplot і часовий графік;
- виявляти потенційні викиди за правилом $1.5IQR$ без автоматичного їх вилучення;
- пояснювати ідею статистичного управління процесами;
- будувати й інтерпретувати I–MR контрольні карти Шухарта;
- розрізняти статистичну стабільність процесу та відповідність продукції технологічним нормативам.

2. План лекції

1. Генеральна сукупність і вибірка.
2. Репрезентативність та способи відбору.
3. Варіаційний і частотний ряди.
4. Основні характеристики положення та розсіювання.
5. Квартилі, IQR та потенційні викиди.
6. Гістограма, полігон, boxplot, scatterplot і часовий графік.
7. Основи Statistical Process Control.
8. Контрольні карти Шухарта.
9. I–MR карти для одиничних послідовних вимірювань.
10. Статистична стабільність і технологічна відповідність.

3. Генеральна сукупність і вибірка

Генеральна сукупність — множина всіх об'єктів або результатів, щодо яких потрібно зробити статистичний висновок.

Приклади:

- усі упаковки, вироблені протягом зміни;
- усі партії певного продукту за сезон;
- усі можливі вимірювання вологості при заданому технологічному режимі;
- усі споживачі певної цільової групи.

Вибірка — частина генеральної сукупності, фактично відібрана для дослідження:

$$x_1, x_2, \dots, x_n.$$

Число n називають обсягом вибірки.

Головна ідея вибіркового методу. Ми досліджуємо не всю генеральну сукупність, а обмежену вибірку, але прагнемо зробити обґрунтований висновок про ширшу сукупність.

4. Чому повне спостереження часто неможливе

У харчових технологіях суцільне дослідження може бути:

- надто дорогим;
- тривалим;
- фізично неможливим;
- руйнівним, якщо лабораторний аналіз знищує зразок;
- беззмістовним для безперервного виробничого потоку.

Тому якість статистичного висновку визначається не лише кількістю вимірювань, а й тим, **як саме сформовано вибірку**.

5. Репрезентативність вибірки

Вибірка є репрезентативною, якщо вона достатньо добре відображає структуру генеральної сукупності відносно досліджуваної ознаки.

Репрезентативність може порушуватися через:

- відбір лише на початку зміни;
- виключення незручних або підозрілих зразків;
- відбір тільки з однієї з кількох ліній;
- систематичний відбір лише «хорошої» продукції;
- змішування різних режимів без фіксації групи;
- самовідбір учасників у сенсорному дослідженні.

Велика вибірка не компенсує систематичне зміщення відбору. Тисяча неправильно відібраних спостережень може давати гірший висновок, ніж сто коректно відібраних.

6. Основні способи відбору

Спосіб	Ідея	Приклад
Простий випадковий	Кожний об'єкт має однакову ймовірність потрапити до вибірки	Випадковий вибір упаковок із партії
Систематичний	Відбір через фіксований інтервал після випадкового старту	Кожна 50-та упаковка
Стратифікований	Окремий відбір із важливих підгруп	Проби з кожної виробничої зміни
Кластерний	Вибираються цілі групи об'єктів	Кілька випадково вибраних партій

7. Варіаційний ряд

Початкову вибірку можна впорядкувати за зростанням:

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}.$$

Отримуємо **варіаційний ряд**. Він дозволяє побачити:

- мінімум і максимум;
- центральні значення;
- повтори;
- прогалини;
- крайні спостереження.

Для контрольних карт часовий порядок, навпаки, зберігають. Тому важливо розуміти мету впорядкування.

8. Частоти і відносні частоти

Якщо значення x_j зустрічається n_j разів, число n_j називають абсолютною частотою.

Відносна частота:

$$w_j = \frac{n_j}{n}.$$

Для повної таблиці частот:

$$\sum_j n_j = n, \qquad \sum_j w_j = 1.$$

9. Інтервальний варіаційний ряд

Для неперервних даних або великої кількості різних значень вибірку групують в інтервали.

Орієнтовну кількість інтервалів можна визначити за формулою Стерджеса:

$$\boxed{k = 1 + 3.322 \lg n}.$$

Розмах:

$$R = x_{\max} - x_{\min}.$$

Орієнтовна ширина інтервалу:

$$h \approx \frac{R}{k}.$$

9.1. Формула Стерджеса — лише орієнтир

Кількість інтервалів не є «єдино правильною». Надто мало інтервалів приховує структуру, надто багато — робить гістограму нестабільною та шумною.

10. Вибіркове середнє

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}.$$

Середнє враховує всі значення і є основною характеристикою центрального рівня для кількісних даних.

Приклад

Нехай результати вимірювання маси: 499.8; 500.1; 500.4; 499.9; 500.3 г.

$$\bar{x} = \frac{499.8 + 500.1 + 500.4 + 499.9 + 500.3}{5} = 500.1 \text{ г}.$$

10.1. Чутливість середнього до крайніх значень

Якщо одне значення 500.3 замінити на 510.3 г, середнє суттєво зміниться. Тому його потрібно інтерпретувати разом із медіаною, розсіюванням і графіком.

11. Медіана

Медіана ділить впорядковану вибірку на дві частини.

Для непарного n:

$$Me = x_{\left(\frac{n+1}{2}\right)}.$$

Для парного n:

$$Me = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}.$$

Медіана стійкіша до крайніх значень і асиметрії.

12. Мода

Мода — значення, яке трапляється найчастіше. Розподіл може бути:

- одномодальним;
- багатомодальним;
- без вираженої моди.

Дві виражені моди можуть свідчити про змішування різних технологічних режимів або груп.

13. Мінімум, максимум і розмах

$$x_{\min}, \qquad x_{\max}.$$

$$\boxed{R = x_{\max} - x_{\min}}.$$

Розмах простий, але залежить лише від двох крайніх значень і є нестійкою характеристикою мінливості.

14. Вибіркова дисперсія

Виправлена вибіркова дисперсія:

$$\boxed{s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

Знаменник $n-1$ пов'язаний із втратою одного ступеня свободи при оцінюванні невідомого генерального середнього за тією самою вибіркою.

15. Вибіркове стандартне відхилення

$$\boxed{s = \sqrt{s^2}}.$$

Стандартне відхилення має ті самі одиниці вимірювання, що й вихідна величина, і характеризує типовий масштаб розсіювання навколо середнього.

16. Коефіцієнт варіації

$$\boxed{V = \frac{s}{\bar{x}} \cdot 100\%}.$$

Коефіцієнт варіації характеризує відносну мінливість і дає змогу порівнювати розсіювання показників, виміряних у різних масштабах.

16.1. Обмеження коефіцієнта варіації

Його не слід механічно застосовувати, якщо:

- середнє близьке до нуля;
- можливі від'ємні значення без природного абсолютного нуля;
- шкала є умовною.

17. Квартилі

Квартилі ділять впорядковану вибірку на чотири частини:

- Q_1 — приблизно 25% значень не перевищують його;
- $Q_2 = Me$ — медіана;
- Q_3 — приблизно 75% значень не перевищують його.

18. Міжквартильний розмах

$$\boxed{IQR = Q_3 - Q_1}.$$

IQR характеризує розсіювання центральних 50% спостережень і є значно стійкішим до крайніх значень, ніж звичайний розмах.

19. Потенційні викиди: правило 1.5IQR

Нижня та верхня «огорожі»:

$$L_f = Q_1 - 1.5IQR,$$

$$U_f = Q_3 + 1.5IQR.$$

Спостереження поза цими межами позначають як **потенційні викиди**.

Потенційний викид ≠ помилка. Спочатку потрібно встановити його причину. Це може бути помилка введення, несправність приладу, реальна спеціальна причина або закономірне значення з хвоста розподілу.

19.1. Чому не можна автоматично вилучати викиди

Автоматичне вилучення може приховати саме той технологічний сигнал, який потрібно знайти: поломку дозатора, зміну сировини, порушення температури або інший реальний процес.

20. Гістограма

Гістограма — графічне представлення частот або щільності спостережень за інтервалами.

Вона допомагає оцінити:

- центр;
- розсіювання;
- симетрію або асиметрію;
- мультимодальність;
- можливі крайні значення.

20.1. Чого гістограма не доводить

Гістограма не є формальним тестом нормальності. Її вигляд залежить від обсягу вибірки та числа інтервалів.

21. Полігон частот

Полігон частот будують, з'єднуючи точки, що відповідають серединам інтервалів та їх частотам. Він зручний для порівняння кількох розподілів на одному графіку.

22. Boxplot

Boxplot компактно відображає:

- Q_1;
- медіану;
- Q_3;
- IQR;
- межі «вусів»;
- потенційні викиди.

Він особливо зручний для порівняння декількох груп або технологічних режимів.

23. Scatterplot

Діаграма розсіювання використовується для парних кількісних даних (x_i, y_i). Вона дозволяє оцінити:

- напрям зв'язку;
- лінійність або нелінійність;
- силу розсіювання;
- кластери;
- викиди.

Scatterplot стане основним графічним інструментом у темах кореляції та регресії.

24. Часовий графік

Якщо спостереження отримано послідовно в часі, їх порядок не можна втрачати. Часовий графік може показати:

- тренд;
- стрибок рівня;
- періодичність;
- серії підвищених або знижених значень;
- поодинокі аномалії.

Саме часовий порядок є основою контрольних карт.

25. Описова статистика не замінює графіку

Два набори даних можуть мати однакові середні та стандартні відхилення, але дуже різну внутрішню структуру. Тому якісний аналіз повинен поєднувати:

$$\boxed{\text{числові характеристики} + \text{графічне представлення}}.$$

26. Statistical Process Control — статистичне управління процесами

SPC — система статистичних методів для спостереження за процесом у часі та виявлення нетипових змін.

Основна ідея: не кожне відхилення потребує втручання. Потрібно розрізняти два типи причин мінливості.

26.1. Звичайні причини

Малі постійні джерела варіативності, властиві процесу: мікроколивання сировини, температури, дозування, вимірювання.

26.2. Спеціальні причини

Нетипові фактори:

- поломка обладнання;
- помилка налаштування;

- різка зміна сировини;
- помилка оператора;
- порушення температурного режиму;
- збій засобу вимірювання.

27. Статистично керований процес

Процес називають статистично керованим або стабільним, якщо його мінливість пояснюється переважно звичайними причинами і статистичні характеристики не демонструють сигналів структурної зміни.

Стабільність не означає ідеальність. Стабільний процес може мати велике розсіювання або бути зміщеним відносно цільового значення.

28. Контрольна карта Шухарта

Контрольна карта — часовий графік із трьома основними лініями:

- CL — центральна лінія;
- UCL — верхня контрольна межа;
- LCL — нижня контрольна межа.

Ідеологічно межі пов'язані з діапазоном приблизно $\mu \pm 3\sigma$ стабільного процесу.

29. Яку карту обрати

Тип контрольної карти залежить від структури даних.

Структура спостережень	Тип карти
Одне кількісне значення в кожний момент часу	I–MR
Невеликі підгрупи кількісних спостережень	$\bar{X}$–R або $\bar{X}$–S
Частка дефектної продукції	p-chart
Кількість дефектів	c- або u-chart

У ПР2 використовується I–MR, оскільки студент має послідовність одиничних вимірювань маси упаковок.

30. I-карта індивідуальних значень

Нехай маємо часову послідовність:

$$x_1, x_2, \dots, x_n.$$

Центральна лінія:

$$\boxed{CL_I = \bar{x}}.$$

31. Рухомий розмах

Оскільки в кожний момент маємо лише одне спостереження, мінливість оцінюють за різницею сусідніх значень:

$$\boxed{MR_i = |x_i - x_{i-1}|}, \qquad i = 2, \dots, n.$$

Середній рухомий розмах:

$$\boxed{\overline{MR} = \frac{1}{n-1}\sum_{i=2}^{n} M\,R_i}.$$

32. Оцінка стандартного відхилення процесу

Для рухомого розмаху довжини 2 використовується константа:

$$d_2 = 1.128.$$

Тому:

$$\boxed{\hat{\sigma} = \frac{\overline{MR}}{1.128}}.$$

33. Контрольні межі I-карти

Загальна форма:

$$UCL_I = \bar{x} + 3\hat{\sigma}, \qquad LCL_I = \bar{x} - 3\hat{\sigma}.$$

Після підстановки оцінки σ:

$$\boxed{UCL_I = \bar{x} + 2.66\overline{MR}}.$$

$$\boxed{LCL_I = \bar{x} - 2.66\overline{MR}}.$$

34. MR-карта

Центральна лінія:

$$\boxed{CL_{MR} = \overline{MR}}.$$

Для рухомого розмаху довжини 2:

$$D_3 = 0, \qquad D_4 = 3.267.$$

Отже:

$$\boxed{UCL_{MR} = 3.267\overline{MR}}.$$

$$\boxed{LCL_{MR} = 0}.$$

35. Інтерпретація I–MR карти

I-карта відповідає переважно на питання: чи виникло нетипове значення рівня процесу?

MR-карта: чи виникла нетипово велика зміна між сусідніми спостереженнями?

Вони аналізуються разом.

36. Найпростіший сигнал спеціальної причини

Основний сигнал у ПР2:

$$x_i > UCL \quad \text{або} \quad x_i < LCL.$$

Таке спостереження малоймовірне для стабільного процесу і потребує предметного розслідування.

37. Інші можливі сигнали на контрольній карті

У повній практиці SPC також аналізують:

- довгі серії точок по один бік CL;
- монотонні тренди;
- циклічність;
- надмірне групування біля центральної лінії;
- послідовності точок поблизу контрольних меж.

У ПР2 ці правила не перевантажують ручний розрахунок, але студент повинен розуміти, що вихід за UCL/LCL — не єдиний можливий сигнал.

38. Контрольні межі та межі специфікації

Це одна з найважливіших відмінностей у SPC.

Межі	Звідки походять	На яке питання відповідають
LCL, UCL Статистичні влас	тивості фактичного процесу Чи стабільний пр	оцес?
LSL, USL Норматив, станда	рт, технічна специфікація Чи відповідає пр	одукція вимогам?

статистична стабільність ≠ технологічна відповідність.

39. Чотири принципові ситуації

Стабільність	Відповідність нормативу	Інтерпретація
Так	Так	Бажаний стан
Так	Ні	Процес передбачуваний, але його центр/розсіювання неприйнятні
Ні	Поки так	Продукція може бути в допуску, але процес непередбачуваний
Ні	Ні	Потрібне термінове технологічне втручання

40. Приклад логіки аналізу технологічної вибірки

Нехай отримано 20 послідовних вимірювань маси упаковок.

Послідовність аналізу:

дані → описові статистики → гістограма/boxplot → IQR → часовий порядок → I–MR → висновок.

40.1. Чому потрібні і гістограма, і I–MR карта

Гістограма ігнорує часовий порядок і показує форму розподілу. Контрольна карта зберігає порядок і показує динаміку процесу. Вони відповідають на різні питання.

41. Описова статистика і статистичний висновок — не одне й те саме

Описова статистика характеризує лише конкретну вибірку. Наприклад:

$$\bar{x} = 500.3 \text{ г}, \qquad s = 0.8 \text{ г}$$

є характеристиками саме отриманого набору даних. Щоб робити формальні висновки про генеральне середнє, дисперсію або різницю між процесами, потрібні методи статистичного оцінювання та перевірки гіпотез.

42. Типові помилки

1. **Вважати вибірку репрезентативною лише через великий n.**
2. **Плутати середнє з «типовим» значенням при сильній асиметрії.**
3. **Використовувати лише середнє без оцінювання розсіювання.**
4. **Автоматично вилучати потенційні викиди.**
5. **Робити висновок про нормальність лише за гістограмою.**
6. **Сортувати дані перед побудовою контрольної карти.** Це знищує часову інформацію.
7. **Ототожнювати UCL/LCL з технологічним допуском.**
8. **Інтерпретувати одну точку поза UCL як доказ конкретної причини.** Карта виявляє сигнал, але причина потребує технологічного розслідування.
9. **Робити висновок про стабільність лише за малим коефіцієнтом варіації.**

43. Зв'язок із практичною роботою №2

У ПР2 студент послідовно застосовує весь основний матеріал лекції:

- аналізує вибірку послідовних вимірювань маси;
- будує варіаційний ряд;
- обчислює описові статистики;
- визначає квартилі, IQR і потенційні викиди;
- будує гістограму і boxplot;
- повертає дані у часовий порядок;
- обчислює рухомі розмахи;
- будує I–MR карти;
- формулює окремо статистичний і технологічний висновки.

44. Підсумок лекції

генеральна сукупність → вибірка → репрезентативність → описові статистики → візуалізація → викиди →

Якісний статистичний аналіз починається не з критерію і не з p-value, а з правильно сформованої вибірки, перевірки структури даних, описових статистик та графічного аналізу. Контрольні карти додають до цього часовий вимір і дозволяють оцінювати стабільність процесу.

45. Контрольні питання

1. Що називають генеральною сукупністю та вибіркою?
2. Що означає репрезентативність вибірки?
3. Чому велика вибірка не гарантує репрезентативності?
4. У чому різниця між простим випадковим і систематичним відбором?
5. Що таке варіаційний ряд?
6. Коли доцільно будувати інтервальний ряд?
7. Який зміст має вибіркове середнє?
8. Чому медіана стійкіша до викидів?
9. Що характеризують дисперсія та стандартне відхилення?
10. Для чого використовують коефіцієнт варіації?
11. Що таке квартилі та IQR?
12. Як визначають потенційні викиди за правилом 1.5IQR?
13. Чому потенційний викид не можна автоматично вилучати?
14. Яку інформацію дає гістограма?
15. Чим boxplot зручний для порівняння груп?
16. Чим часовий графік відрізняється від гістограми?
17. Що таке звичайні та спеціальні причини варіативності?
18. Що означає статистично керований процес?
19. Чому для одиничних послідовних спостережень використовують I–MR карту?
20. Як обчислюється рухомий розмах?
21. Як оцінюється σ через $\overline{MR}$?
22. Запишіть контрольні межі I-карти.
23. Запишіть контрольні межі MR-карти.
24. Що означає точка за UCL або LCL?
25. У чому різниця між UCL/LCL і USL/LSL?
26. Чи може статистично стабільний процес бути технологічно неприйнятним?
27. Чому дані не можна сортувати перед побудовою контрольної карти?

Лекція 5

Статистичне оцінювання параметрів і довірчі інтервали

1. Мета лекції та результати навчання

Мета лекції — сформувати розуміння логіки статистичного оцінювання: як за обмеженою вибіркою отримувати обґрунтовані оцінки невідомих характеристик генеральної сукупності та як кількісно описувати невизначеність цих оцінок.

Після опрацювання лекції здобувач повинен уміти:

- розрізняти генеральні параметри та вибіркові статистики;
- пояснювати різницю між точковою та інтервальною оцінкою;
- характеризувати основні властивості статистичних оцінок;
- обчислювати стандартну похибку середнього;
- пояснювати різницю між стандартним відхиленням s та стандартною похибкою SE;
- будувати довірчий інтервал для генерального середнього за невідомої генеральної дисперсії;
- будувати довірчі інтервали для генеральної дисперсії та стандартного відхилення;
- правильно інтерпретувати рівень довіри та межі довірчого інтервалу;
- пояснювати, як на ширину довірчого інтервалу впливають обсяг вибірки, розсіювання та рівень довіри;
- розрізняти статистичну точність оцінки та природну мінливість індивідуальних спостережень.

2. План лекції

1. Параметри генеральної сукупності та вибіркові статистики.
2. Точкове статистичне оцінювання.
3. Властивості оцінок: незміщеність, спроможність, ефективність.
4. Стандартна похибка середнього.
5. Інтервальне оцінювання та рівень довіри.
6. t-розподіл Стьюдента.
7. Довірчий інтервал для генерального середнього.
8. Довірчий інтервал для генеральної дисперсії та стандартного відхилення.
9. Чинники, що визначають ширину інтервалу.
10. Правильна статистична інтерпретація довірчих інтервалів.

3. Генеральні параметри і вибіркові статистики

У попередніх лекціях було введено генеральну сукупність і вибірку. Тепер центральне питання змінюється: як за вибіркою зробити висновок про невідомі характеристики всієї генеральної сукупності?

Генеральна характеристика	Позначення	Типова вибіркова оцінка
Генеральне середнє	μ	$\bar{x}$
Генеральна дисперсія	σ^2	s^2
Генеральне стандартне відхилення	σ	s
Генеральна частка	p	$\hat{p}$

Параметри генеральної сукупності зазвичай невідомі, а статистики обчислюються безпосередньо з фактичних даних.

Ключова ідея. Вибіркова статистика є випадковою: якби ми сформували іншу вибірку з тієї самої генеральної сукупності, отримали б трохи інше значення.

4. Точкова оцінка

Точкова оцінка — одне числове значення, використане для оцінювання невідомого параметра.

Наприклад:

$$\hat{\mu} = \bar{x}.$$

Якщо у вибірці з 12 зразків середня масова частка вологи становить 13.92%, то 13.92% є точковою оцінкою невідомого генерального середнього.

4.1. Чому точкової оцінки недостатньо

Значення 13.92% саме по собі не показує, наскільки точним є оцінювання. Результат може бути отриманий:

- за 5 або за 500 спостереженнями;
- при малому або великому розсіюванні;
- за стабільного або нестабільного процесу.

Тому поряд із точковою оцінкою потрібна характеристика її невизначеності.

5. Властивості статистичних оцінок

5.1. Незміщеність

Оцінка $\hat{\theta}$ параметра θ є незміщеною, якщо:

$$E(\hat{\theta}) = \theta.$$

Інтуїтивно: при багаторазовому повторенні вибіркового експерименту оцінка в середньому не завищує і не занижує параметр.

5.2. Спроможність

Оцінка є спроможною, якщо зі збільшенням обсягу вибірки вона наближається до істинного параметра.

$$\hat{\theta}_n \to \theta \quad \text{при} \quad n \to \infty.$$

5.3. Ефективність

Серед кількох незміщених оцінок ефективнішою є та, що має меншу дисперсію, тобто меншу вибіркову невизначеність.

6. Чому у вибірковій дисперсії використовується $n-1$

Вибіркова дисперсія:

$$\boxed{s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

Знаменник $n-1$ пов'язаний із тим, що середнє $\bar{x}$ вже оцінено за тією самою вибіркою. Після фіксації середнього відхилення $x_i - \bar{x}$ не є повністю незалежними: їх сума дорівнює нулю.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

Тому число ступенів свободи:

$$\nu = n - 1.$$

7. Розподіл вибіркового середнього

Нехай з однієї генеральної сукупності багаторазово формують незалежні вибірки однакового обсягу n. Для кожної обчислюють середнє.

Вибіркове середнє також є випадковою величиною:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Для незалежних однаково розподілених спостережень:

$$E(\bar{X}) = \mu.$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Стандартне відхилення розподілу вибіркового середнього:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

8. Стандартна похибка середнього

Оскільки генеральне σ зазвичай невідоме, його оцінюють за вибірковим s:

$$\boxed{SE_{\bar{x}} = \frac{s}{\sqrt{n}}}.$$

Стандартна похибка середнього характеризує не розсіювання окремих спостережень, а невизначеність оцінки генерального середнього.

8.1. Принципова різниця між s і SE

Характеристика	Що вимірює	Залежність від n
s Мінливість індивідуаль	них спостережень Не зменшується автомат	ично при збільшенні n
$SE = s/\sqrt{n}$ Невизначені	сть оцінки середнього Зменшується	приблизно як $1/\sqrt{n}$

Приклад

Нехай $s = 0.18$, $n = 9$.

$$SE = \frac{0.18}{\sqrt{9}} = 0.06.$$

Якщо при тому самому s збільшити обсяг до $n = 36$:

$$SE = \frac{0.18}{\sqrt{36}} = 0.03.$$

Збільшення обсягу вибірки у 4 рази зменшило стандартну похибку у 2 рази.

9. Інтервальне оцінювання

Інтервальна оцінка задає не одне значення, а діапазон можливих значень параметра:

$$\theta_L < \theta < \theta_U.$$

Інтервал (θ_L, θ_U) називають **довірчим інтервалом**.

10. Рівень довіри та рівень значущості

Рівень довіри:

$$\gamma = 1 - \alpha.$$

Найчастіше:

$$\gamma = 0.95, \qquad \alpha = 0.05.$$

Це означає, що використовується процедура побудови 95% довірчих інтервалів.

10.1. Правильна інтерпретація

У класичному частотному підході після побудови конкретного інтервалу параметр не є випадковою величиною. Правильний зміст такий:

Якби багато разів повторювали процедуру формування вибірки та побудови інтервалу одним і тим самим методом, приблизно 95% таких інтервалів містили б істинне значення параметра.

10.2. Практична мова

У прикладних роботах допустимо писати: «з довірою 95% генеральне середнє оцінюється інтервалом...», але потрібно розуміти статистичний зміст цієї фрази.

11. Чому не можна використовувати лише нормальний розподіл при невідомому σ

Якби генеральне стандартне відхилення σ було відоме, стандартизована статистика:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

за нормальності генеральної сукупності мала б стандартний нормальний розподіл.

На практиці σ зазвичай невідоме і замінюється на s. Це додає невизначеність, особливо для малих вибірок. Тому використовується t-розподіл Стьюдента.

12. t-розподіл Стьюдента

Статистика:

$$\boxed{T = \frac{\bar{X} - \mu}{S/\sqrt{n}}}$$

для нормально розподіленої генеральної сукупності має t-розподіл із:

$$\nu = n - 1$$

ступенями свободи.

12.1. Основні властивості t-розподілу

- симетричний відносно 0;
- подібний до стандартного нормального;
- має важчі хвости;
- форма залежить від числа ступенів свободи;
- при збільшенні ν наближається до $N(0,1)$.

12.2. Чому хвости важчі

Використання вибіркового s замість відомого σ створює додаткову невизначеність. Саме тому критичне t для малої вибірки більше за відповідне z.

13. Довірчий інтервал для генерального середнього

Якщо генеральна дисперсія невідома:

$$\boxed{\mu \in \bar{x} \pm t_{1-\alpha/2;\, n-1} \frac{s}{\sqrt{n}}}.$$

Або:

$$\mu \in \bar{x} \pm t_{\text{crit}} SE_{\bar{x}}.$$

Гранична похибка:

$$\boxed{\Delta = t_{\text{crit}} SE_{\bar{x}}}.$$

14. Умови використання t-інтервалу для середнього

- дані кількісні;
- спостереження незалежні;
- вибірка сформована коректно;
- для малих n генеральний розподіл приблизно нормальний;
- відсутні необґрунтовані грубі помилки.

Для великих вибірок t-інтервал часто досить стійкий до помірних відхилень від нормальності завдяки центральній граничній теоремі.

15. Покроковий приклад довірчого інтервалу для середнього

Нехай за $n = 12$ вимірюваннями отримано:

$$\bar{x} = 13.9167, \qquad s = 0.17495.$$

Стандартна похибка:

$$SE = \frac{0.17495}{\sqrt{12}} \approx 0.05050.$$

Для $\alpha = 0.05$ і $\nu = 11$:

$$t_{0.975;11} \approx 2.201.$$

Гранична похибка:

$$\Delta = 2.201 \cdot 0.05050 \approx 0.1112.$$

Тоді:

$$13.9167 - 0.1112 < \mu < 13.9167 + 0.1112.$$

$$\boxed{95\%\, CI_{\mu} = [13.8055;\ 14.0279]}.$$

16. Що означає цей інтервал технологічно

Точкова оцінка середньої вологості — 13.92%, але через обмеженість вибірки справжнє генеральне середнє оцінюється з невизначеністю.

Інтервал показує діапазон значень генерального середнього, сумісних із даними та вибраною процедурою оцінювання.

Не плутати. 95% довірчий інтервал для середнього не означає, що 95% окремих зразків мають вологість у цьому інтервалі.

17. Довірчий інтервал для генеральної дисперсії

Для нормально розподіленої генеральної сукупності:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Звідси:

$$\boxed{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2;\,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2;\,n-1}}}.$$

18. Розподіл χ^2

Chi-square розподіл:

- набуває лише невід'ємних значень;

- несиметричний;
- його форма залежить від числа ступенів свободи;
- використовується для задач, пов'язаних із дисперсією, а також у критеріях узгодженості.

19. Чому інтервал для дисперсії несиметричний

Через асиметрію χ^2-розподілу відстань від s^2 до нижньої й верхньої межі різна. Це принципова відмінність від симетричного t-інтервалу для середнього.

20. Приклад інтервалу для дисперсії

Нехай:

$$n = 12, \qquad s^2 = 0.030606, \qquad \nu = 11.$$

Для 95% рівня довіри:

$$\chi^2_{0.025;11} \approx 3.8157, \qquad \chi^2_{0.975;11} \approx 21.9200.$$

Чисельник:

$$(n-1)s^2 = 11 \cdot 0.030606 \approx 0.33667.$$

Нижня межа:

$$\sigma_L^2 = \frac{0.33667}{21.9200} \approx 0.01536.$$

Верхня:

$$\sigma_U^2 = \frac{0.33667}{3.8157} \approx 0.08823.$$

$$\boxed{95\%\, CI_{\sigma^2} = [0.01536;\ 0.08823]}.$$

21. Довірчий інтервал для стандартного відхилення

Беремо квадратні корені з меж інтервалу для дисперсії:

$$\boxed{\sigma_L = \sqrt{\sigma_L^2}, \qquad \sigma_U = \sqrt{\sigma_U^2}}.$$

У прикладі:

$$\sigma_L \approx 0.1239, \qquad \sigma_U \approx 0.2970.$$

$$\boxed{95\%\, CI_{\sigma} = [0.1239;\ 0.2970]}.$$

22. Умови застосування χ^2-інтервалу

Класичний довірчий інтервал для дисперсії є суттєво чутливим до порушення нормальності. Тому для явно ненормальних даних його не слід використовувати механічно.

23. Що впливає на ширину довірчого інтервалу

23.1. Обсяг вибірки

Для середнього:

$$SE = \frac{s}{\sqrt{n}}.$$

Зі збільшенням n інтервал звужується.

23.2. Розсіювання

Чим більше s, тим більша невизначеність середнього та ширший інтервал.

23.3. Рівень довіри

99% інтервал ширший за 95%, а 95% ширший за 90%.

Вища надійність інтервальної процедури потребує ширшого діапазону.

24. Точність і достовірність — компроміс

Для фіксованого набору даних неможливо одночасно підвищити рівень довіри і зробити інтервал вужчим без додаткової інформації. Один зі способів покращити точність — збільшити обсяг вибірки.

25. Відносна точність оцінки

Іноді граничну похибку порівнюють із оціненим середнім:

$$\delta = \frac{\Delta}{\bar{x}} \cdot 100\%.$$

Це дозволяє оцінити відносну точність інтервального оцінювання.

26. Довірчий інтервал і технологічний норматив

Наявність нормативного значення дає додатковий рівень інтерпретації. Наприклад, якщо весь 95% CI генерального середнього лежить нижче критичної межі, це сильніше підтримує технологічний висновок, ніж одна лише точкова оцінка.

Водночас інтервал середнього не описує розсіювання індивідуальної продукції. Для оцінювання частки одиниць поза допуском потрібна інша постановка задачі.

27. Чому статистично точне середнє не означає низької варіативності

За великого n стандартна похибка може бути дуже малою навіть тоді, коли s велике.

$$SE = \frac{s}{\sqrt{n}}.$$

Тому:

точно оцінене середнє ≠ малий розкид окремих значень.

28. Довірчий інтервал і статистична гіпотеза

Довірчі інтервали тісно пов'язані з перевіркою гіпотез.

Наприклад, для двосторонньої перевірки:

$$H_0: \mu = \mu_0$$

на рівні $\alpha = 0.05$ значення μ_0 буде відхилене тоді, коли воно не входить у відповідний 95% довірчий інтервал.

Цей зв'язок буде використано в ПР4.

29. Типові помилки

1. **Плутати s і SE.**
2. **Писати «95% спостережень лежать у 95% CI середнього».**
3. **Трактувати 95% CI як 95% ймовірність конкретного фіксованого параметра після побудови інтервалу в класичному підході.**
4. **Ігнорувати припущення про нормальність для χ^2-інтервалу дисперсії.**
5. **Вважати вузький CI доказом малого технологічного розсіювання.**
6. **Порівнювати лише середні без оцінки невизначеності.**
7. **Збільшувати рівень довіри й очікувати одночасного звуження інтервалу.**

30. Зв'язок із практичною роботою №3

ПР3 реалізує матеріал лекції у повному циклі:

- обчислення описових статистик;
- визначення s^2, s і SE;
- визначення t_{crit};
- побудова 95% CI для μ;
- визначення квантилів χ^2;
- побудова 95% CI для σ^2 та σ;
- перевірка результатів у Microsoft Excel та спеціалізованому ПЗ.

31. Підсумок лекції

$$\boxed{\text{вибірка} \rightarrow \bar{x}, s^2, s \rightarrow SE \rightarrow t\text{-розподіл} \rightarrow CI_\mu \rightarrow \chi^2\text{-розподіл} \rightarrow CI_{\sigma^2}, CI_\sigma \rightarrow \text{інтерпретація}}.$$

Статистичне оцінювання переходить від одного вибіркового числа до кількісного опису його невизначеності. Саме довірчі інтервали роблять статистичний висновок інформативнішим за просте повідомлення середнього.

32. Контрольні питання

1. Чим параметр генеральної сукупності відрізняється від вибіркової статистики?
2. Що таке точкова оцінка?
3. Що означає незміщеність оцінки?

4. Що означає спроможність оцінки?
5. Чому у вибірковій дисперсії використовується $n-1$?
6. Що характеризує стандартне відхилення s?
7. Що характеризує стандартна похибка SE?
8. Як зміниться SE, якщо обсяг вибірки збільшити у 4 рази?
9. Що таке довірчий інтервал?
10. Як пов'язані γ та α?
11. Як правильно трактувати 95% довірчий інтервал?
12. Чому при невідомому σ використовується t-розподіл?
13. Як змінюється t-розподіл при збільшенні числа ступенів свободи?
14. Запишіть формулу 95% CI для генерального середнього.
15. Що таке гранична похибка Δ?
16. Від чого залежить ширина довірчого інтервалу?
17. Який розподіл використовується для побудови класичного CI дисперсії?
18. Чому інтервал для дисперсії несиметричний?
19. Як отримати CI для стандартного відхилення?
20. Чому нормальність особливо важлива для χ^2-інтервалу?
21. Чому вузький CI середнього не означає, що окремі спостереження мало відрізняються?
22. Як пов'язані довірчий інтервал і двостороння перевірка статистичної гіпотези?

Лекція 6

Підготовка даних та сучасне програмне забезпечення для статистичного аналізу

1. Мета лекції та результати навчання

Мета лекції — сформувати практичні навички правильної організації дослідницьких даних перед статистичним аналізом та ознайомити здобувачів із сучасними інструментами, які можуть використовуватися для обчислень, перевірки результатів, візуалізації та відтворюваного аналізу.

Після опрацювання лекції здобувач повинен уміти:

- розрізняти об'єкт спостереження, змінну та значення змінної;
- визначати тип і шкалу вимірювання змінної;
- організовувати дані за принципом «один рядок — одне спостереження, один стовпець — одна змінна»;
- розрізняти сирі, очищені та агреговані дані;
- виявляти пропущені значення, дублікати, помилки введення та підозрілі значення;
- правильно кодувати категоріальні змінні та групувальні фактори;
- готувати таблицю для імпорту у статистичне програмне забезпечення;
- пояснювати роль Microsoft Excel, Analysis ToolPak, Real Statistics, Jamovi, JASP, PSPP та R/RStudio;
- обирати програмне середовище відповідно до мети аналізу;
- пояснювати принцип відтворюваності статистичного дослідження;
- виконувати схему «ручний розрахунок → Excel → спеціалізоване ПЗ → зіставлення → висновок».

2. План лекції

1. Дані як основа статистичного дослідження.
2. Об'єкт спостереження, змінна та фактор.
3. Типи змінних і шкали вимірювання.
4. Правильна структура таблиці даних.
5. Сирі й агреговані дані.
6. Пропущені значення, дублікати та помилки.
7. Потенційні викиди й перевірка якості даних.
8. Кодування категоріальних змінних.
9. Microsoft Excel та Analysis ToolPak.
10. Real Statistics Resource Pack.
11. Jamovi, JASP, PSPP, R/RStudio.
12. Відтворюваність і документування статистичного аналізу.

3. Чому підготовка даних є частиною статистичного аналізу

Статистичний метод працює не з абстрактними «числами», а з даними, отриманими за певним експериментальним дизайном. Помилка у структурі таблиці, одиницях вимірювання або кодуванні груп може зіпсувати результат ще до застосування будь-якої формули.

Практичний принцип. Якість статистичного висновку не може бути вищою за якість вихідних даних і спосіб їх отримання.

До статистичного аналізу потрібно відповісти щонайменше на такі питання:

- що є об'єктом спостереження;
- які змінні вимірюються;
- які одиниці вимірювання;
- які змінні є кількісними, а які категоріальними;
- яка змінна є результативною;
- які змінні визначають групи або умови експерименту;
- чи є спостереження незалежними або парними;
- чи є пропуски, дублікати, очевидні помилки;
- чи збережені сирі дані.

4. Об'єкт спостереження, змінна та значення

4.1. Об'єкт спостереження

Об'єкт спостереження — одиниця, для якої фіксуються характеристики.

У харчових технологіях це може бути:

- окрема упаковка;
- зразок продукту;
- партія;
- окремий технологічний дослід;
- дегустатор;
- окреме вимірювання у певний момент часу.

4.2. Змінна

Змінна — характеристика, яка може набувати різних значень для різних об'єктів.

Наприклад:

- маса упаковки;
- вологість;
- температура процесу;
- рецептура;
- тип пакування;
- органолептична оцінка;

- час зберігання.

4.3. Значення змінної

Конкретне число або категорія, зафіксована для конкретного об'єкта.

Приклад

Об'єкт	Температура, °C	Вихід продукту, %	Режим
Дослід 1	75	93.2	B
Дослід 2	80	94.1	C

Тут один рядок — один дослід; «Температура», «Вихід продукту» та «Режим» — змінні.

5. Основні типи змінних

5.1. Кількісні

Значення виражаються числом і мають зміст кількості або вимірювання.

- маса;
- температура;
- вологість;
- кислотність;
- тривалість;
- концентрація.

5.2. Категоріальні

Значення позначають належність до групи або класу.

- тип рецептури: A/B/C;
- лінія виробництва: 1/2;
- вид пакування;
- наявність дефекту: так/ні.

5.3. Дискретні кількісні

Кількість об'єктів або подій:

- число дефектних одиниць;
- кількість колоній;
- кількість відмов.

5.4. Неперервні кількісні

Теоретично можуть набувати будь-якого значення в інтервалі: маса, температура, вологість, час.

6. Шкали вимірювання

Тип статистичного методу залежить не лише від того, чи записано значення числом, а й від змісту шкали.

Шкала	Властивість	Приклад

Шкала	Властивість	Приклад
Номінальна	Категорії без природного порядку	Тип рецептури, вид упаковки
Порядкова	Категорії мають порядок	Органолептична оцінка, ранг
Інтервальна	Різниці мають зміст, абсолютний нуль умовний	Температура у °C
Відношень	Є змістовний абсолютний нуль	Маса, час, концентрація

6.1. Чому органолептичний бал — особливий випадок

Оцінка 8 балів краща за 7, але не обов'язково можна стверджувати, що різниця 8−7 еквівалентна різниці 5−4. Тому для таких даних часто використовують рангові непараметричні методи.

7. Фактор, рівень фактора та відгук

У технологічному експерименті часто використовують поняття:

- **фактор** — змінна, вплив якої досліджують;
- **рівень фактора** — конкретне значення або категорія фактора;
- **відгук** — результативна змінна.

Фактор — температура сушіння.

Рівні — 70, 75, 80 і 85 °C.

Відгук — вихід готового продукту, %.

8. Основний принцип правильної таблиці

один рядок = одне спостереження, один стовпець = одна змінна.

Такий формат часто називають «довгим» або tidy-форматом. Він особливо зручний для статистичного ПЗ.

8.1. Правильний приклад

ID	Температура	Вихід
1	70	91.8
2	70	92.2
3	75	93.0
4	75	93.4

8.2. Широкий формат

Для ручного ANOVA в Excel інколи зручний широкий формат:

70 °C	75 °C	80 °C	85 °C
91.8	93.0	94.2	92.7
92.2	93.4	94.5	93.0

Обидва формати можуть бути корисними, але потрібно розуміти їх призначення і вміти переходити між ними.

9. Сирі дані

Сирі дані — дані в максимально близькому до первинного спостереження вигляді, до статистичного агрегування.

Наприклад, 24 окремі результати експериментів краще зберігати, ніж лише чотири групові середні.

Правило. Сирі дані необхідно зберігати окремо та не перезаписувати під час очищення або перетворень.

10. Агреговані дані

Агреговані дані — середні, суми, частки, медіани або інші підсумкові характеристики.

Наприклад:

Режим	Середній вихід, %
70 °C	91.88
75 °C	93.08

Вони зручні для звіту, але часто недостатні для повноцінного статистичного аналізу, оскільки приховують внутрішньогрупове розсіювання.

11. Чому не слід зберігати лише середні

Дві групи можуть мати однакове середнє, але різне розсіювання.

$$\bar{x}_1 = \bar{x}_2$$

не означає:

$$s_1 = s_2.$$

Без сирих спостережень неможливо коректно оцінити дисперсію, викиди, форму розподілу та багато статистичних критеріїв.

12. Ідентифікатор спостереження

Доцільно мати окремий ID:

- номер зразка;
- номер упаковки;
- номер дегустатора;
- код експерименту.

ID не є вимірюваною характеристикою і не повинен випадково потрапляти в статистичний аналіз як кількісна змінна.

13. Назви змінних

Добра назва повинна бути:

- однозначною;
- короткою;
- стабільною в усьому наборі даних;
- бажано без зайвих службових символів.

Одиниці вимірювання краще фіксувати в окремому словнику змінних або чітко в заголовку.

14. Одиниці вимірювання

Одна змінна в одному наборі даних повинна мати одну систему одиниць.

Не можна змішувати, наприклад:

- г і кг в одному стовпці;
- хв і с;
- частку 0.05 і відсоток 5 без явного перетворення.

15. Категоріальне кодування

Категорії можуть бути записані текстом:

$$\text{A, B, C}$$

або кодами:

$$1,2,3.$$

Якщо використовуються числові коди, потрібно мати таблицю відповідності.

Код	Режим
1	Контроль
2	Рецептура A
3	Рецептура B

Числовий код не робить категорію кількісною. Різниця між кодами 3 і 1 не означає «у два рази більший режим».

16. Пропущені значення

Пропуск може виникнути через:

- відсутність вимірювання;
- поломку приладу;
- втрату зразка;
- відмову учасника;
- неможливість визначити показник.

16.1. Як позначати пропуски

Бажано використовувати системний missing value, який правильно розпізнає програмне забезпечення.

Не слід бездумно кодувати пропуск як:

$$0, \quad -1, \quad 999.$$

Такі числа можуть бути сприйняті як реальні дані.

16.2. Види механізмів пропусків — концептуально

Причина пропуску важлива. Якщо значення відсутні випадково, наслідки одні. Якщо пропуски систематично виникають саме при дуже високих або низьких результатах, статистичне зміщення може бути суттєвим.

17. Дублікати

Дублікат — повторний запис того самого спостереження. Він може виникати через:

- повторне копіювання рядка;
- об'єднання файлів;
- помилку експорту.

Перш ніж видаляти дублікати, потрібно перевірити ID та зрозуміти, чи це справді повторний запис, а не два однакові за значеннями реальні спостереження.

18. Помилки введення

Типові помилки:

- 13.8 → 138;
- 0.52 → 52;
- переплутані десяткові роздільники;
- помилкова одиниця;
- текст у числовому стовпці;
- пробіли або службові символи.

18.1. Перевірка допустимого діапазону

Для багатьох показників можна сформулювати фізично або технологічно допустимий інтервал:

$$x_{\min}^{\text{phys}} \le x_i \le x_{\max}^{\text{phys}}.$$

Значення поза ним потребує перевірки первинного джерела.

19. Потенційні викиди

Викид — не автоматично помилка. Він може бути:

- помилкою введення;
- помилкою вимірювання;
- реальним спеціальним технологічним випадком;
- закономірним крайнім значенням розподілу.

Для первинного пошуку використовують boxplot, правило 1.5IQR, z-score, графіки залишків та предметний контроль.

$$L_f = Q_1 - 1.5IQR, \qquad U_f = Q_3 + 1.5IQR.$$

Вилучення даних повинно мати документовану причину. Не можна видаляти спостереження лише тому, що воно «псує» бажаний статистичний результат.

20. Логічні перевірки між змінними

Якість даних перевіряють не лише по одному стовпцю.

Приклади логічних правил:

- кінцева дата не може передувати початковій;
- час зберігання не може бути від'ємним;
- температура режиму повинна відповідати коду групи;
- парні вимірювання «до/після» повинні належати тому самому ID.

21. Підготовка даних до аналізу: рекомендований алгоритм

імпорт → перевірка структури → типи змінних → пропуски → дублікати → діапазони → викиди → графік

22. Словник змінних

Для дослідницького набору корисно створити data dictionary.

Змінна	Опис	Тип	Одиниці / коди
ID	Номер зразка	Ідентифікатор	—
moisture	Масова частка вологи	Кількісна	%
regime	Режим сушіння	Номінальна	A/B
time	Час зберігання	Кількісна	доба

Такий словник підвищує відтворюваність і знижує ризик помилкової інтерпретації змінних.

23. Microsoft Excel як навчальне середовище

Excel є зручним для:

- введення та первинної перевірки даних;
- ручного відтворення статистичних формул;
- обчислення проміжних величин;
- побудови графіків;
- демонстрації логіки статистичного алгоритму.

У межах курсу Excel використовується не лише як «калькулятор», а як інструмент для прозорого переходу від формули до числового результату.

24. Основні статистичні функції Excel

Задача	Функція
Середнє	AVERAGE
Медіана	MEDIAN
Мода	MODE.SNGL
Вибіркова дисперсія	VAR.S
Вибіркове стандартне відхилення	STDEV.S
Квартилі	QUARTILE.INC
Кореляція	CORREL, PEARSON
Критичне t	T.INV.2T
Квантіль χ^2	CHISQ.INV

25. Analysis ToolPak

Надбудова Analysis ToolPak дозволяє виконувати стандартні процедури через графічний інтерфейс:

- Descriptive Statistics;
- Histogram;
- F-Test Two-Sample for Variances;
- t-Test;
- ANOVA: Single Factor;
- Correlation;
- Regression.

Її перевага — швидкість; недолік — студент може отримати велику таблицю результатів, не розуміючи, звідки вони взялися. Саме тому в курсі ключові етапи спочатку виконуються вручну.

26. Real Statistics Resource Pack

Real Statistics — надбудова для Excel, яка розширює його статистичні можливості. Для нашого курсу вона корисна насамперед для процедур, яких немає у стандартному ToolPak або які незручно реалізовувати вручну.

Приклади:

- Shapiro–Wilk;
- Levene test;
- Mann–Whitney;
- Wilcoxon signed-rank;
- Tukey HSD;
- додаткові статистичні таблиці та критичні значення.

26.1. Методична роль Real Statistics

У цьому курсі надбудова не замінює ручний алгоритм, а використовується як:

$$\boxed{\text{перевірка} + \text{розширення стандартного Excel}}.$$

27. Jamovi

Jamovi — спеціалізоване статистичне середовище з графічним інтерфейсом. Воно корисне для студентів, які лише переходять від Excel до професійного статистичного аналізу.

Типові можливості:

- описова статистика;
- гістограми, boxplot, Q–Q plots;
- Shapiro–Wilk;
- t-тести;
- непараметричні критерії;

- ANOVA та post-hoc;
- кореляція;
- лінійна регресія.

27.1. Перевага Jamovi у навчанні

Вихідні таблиці оновлюються при зміні налаштувань аналізу, тому студент може порівнювати моделі та параметри без написання коду.

28. JASP

JASP — ще одне графічне статистичне середовище, орієнтоване на зручний аналіз і підготовку статистичних результатів. Воно підтримує класичні методи та додатково відоме широкими можливостями байєсівського аналізу.

Для нашого курсу JASP можна використовувати як альтернативу Jamovi для:

- описової статистики;
- t-тестів;
- ANOVA;
- кореляції;
- регресії.

29. PSPP

PSPP — вільне статистичне програмне забезпечення, концептуально близьке до класичних пакетів табличного статистичного аналізу.

Його можна розглядати як альтернативу для базових процедур:

- описової статистики;
- частот;
- порівняння груп;
- кореляції;
- регресії.

30. R та RStudio

R — мова та середовище статистичних обчислень. **RStudio** — популярне інтегроване середовище для роботи з R.

Переваги:

- дуже широкий набір статистичних методів;
- повна автоматизація аналізу;
- відтворюваність через код;
- гнучка візуалізація;
- можливість обробляти великі набори даних.

30.1. Чому R не є основним інструментом цього курсу

Основна мета дисципліни — статистичне мислення та інтерпретація, а не програмування. Тому R демонструється як професійний напрям подальшого розвитку, тоді як базові практичні роботи виконуються у Excel із перевіркою в Jamovi/Real Statistics.

31. Який інструмент для якої задачі

Задача	Основний навчальний інструмент	Додаткова перевірка
Описова статистика	Excel	Jamovi / JASP
Довірчі інтервали	Ручний розрахунок + Excel	Jamovi / Real Statistics
Shapiro–Wilk	—	Real Statistics / Jamovi
F-test, t-test	Ручний розрахунок + Excel	Jamovi
Mann–Whitney, Wilcoxon	Ручні ранги + Excel	Real Statistics / Jamovi
ANOVA	Ручний розрахунок + Excel	Jamovi / Real Statistics
Кореляція	Excel	Jamovi
Регресія	Excel	Jamovi / R

32. Чому потрібно використовувати більше одного способу перевірки

На навчальному етапі корисна схема:

ручний алгоритм → Excel → спеціалізоване ПЗ → зіставлення.

Якщо результати не збігаються, це сигнал перевірити:

- формулу;
- діапазон клітинок;
- тип тесту;
- односторонній або двосторонній варіант;
- припущення про рівність дисперсій;
- обробку ties і нульових різниць;
- точний або асимптотичний алгоритм.

33. Програмне забезпечення не обирає метод замість дослідника

Програма може швидко обчислити статистику, але не знає автоматично:

- чи незалежні вибірки;
- чи правильно сформована генеральна сукупність;
- чи має сенс середнє для порядкової шкали;
- чи можна трактувати кореляцію причинно;
- чи є викид помилкою або технологічним сигналом;
- чи має отримана різниця практичне значення.

Статистичне ПЗ автоматизує розрахунок, але не звільняє від методичного вибору та предметної інтерпретації.

34. Відтворюваність дослідження

Відтворюваний аналіз — аналіз, який інший дослідник може повторити, використовуючи ті самі вихідні дані та описані процедури.

Для цього потрібно зберігати:

- сирі дані;
- очищений набір даних;
- опис змінних;
- правила очищення;
- версію або назву використаного ПЗ;
- налаштування статистичних тестів;
- вихідні таблиці та графіки;
- остаточний аналітичний файл.

35. Сирі дані не повинні перезаписуватися

Рекомендована структура:

- `raw_data.xlsx` — первинні дані, лише для читання;
- `clean_data.xlsx` — очищена версія;
- `analysis.xlsx` — розрахунки;
- окремий файл із описом змін і рішень.

Це дозволяє повернутися до первинного джерела, якщо виникне питання щодо результату.

36. Документування змін

Якщо спостереження вилучено, одиниці перераховано або значення виправлено, потрібно зафіксувати:

- що змінено;
- чому змінено;
- на підставі якого джерела;
- хто і коли виконав зміну.

37. Приклад підготовки даних для ПРЗ

Початкові дані вологості:

ID	Волога, %
1	13.7
2	13.9
3	14.1
4	13.8

Перед аналізом потрібно перевірити:

1. чи всі значення числові;
2. чи немає пропусків;

3. чи одиниця всюди однакова — %;
4. чи немає дубльованих ID;
5. чи всі значення фізично правдоподібні;
6. чи кожен рядок відповідає незалежному зразку.

38. Робочий workflow ПРЗ

вихідні дані → перевірка структури → описова статистика → s, SE → CI_{μ} → $CI_{\sigma^2}, CI_{\sigma}$ → перевірка у ПЗ →

39. Звітність статистичного аналізу

Якісний звіт не повинен складатися лише зі скриншота програмного виводу.

Потрібно подати:

- опис даних;
- обсяг вибірки;
- використаний метод;
- основні статистики;
- довірчі інтервали;
- графіки, якщо вони потрібні;
- статистичний висновок;
- технологічну інтерпретацію.

40. Надмірна точність чисел

Статистичне ПЗ може повертати 8–12 десяткових знаків, але це не означає, що всі вони мають предметний зміст.

Кількість знаків у звіті повинна відповідати:

- точності вимірювання;
- зручності інтерпретації;
- вимогам методики або журналу.

Якщо вологість вимірюється до 0.1%, звіт:

$$\bar{x} = 13.91666667\%$$

зазвичай недоречний. Для розрахунків можна зберігати повну точність, а в тексті подавати, наприклад, 13.92%.

41. Типові помилки підготовки даних

1. Зберігати різні показники в одному стовпці.
2. Використовувати об'єднані клітинки всередині робочої таблиці даних.
3. Змішувати числові дані з текстовими коментарями.
4. Використовувати колір клітинки як єдиний спосіб кодування групи.
5. Кодувати missing як 0 або 999 без спеціальної обробки.

6. Зберігати лише групові середні, видаляючи сирі спостереження.
7. Перезаписувати первинний файл після очищення.
8. Видаляти викиди без фіксації причини.
9. Плутати ID із кількісною змінною.
10. Копіювати p-value з програми без перевірки обраного тесту.

42. Типові помилки використання програмного забезпечення

1. **«Програма видала результат — отже метод правильний».**
2. **Неправильно заданий тип змінної.** Наприклад, код групи 1/2/3 трактують як неперервну величину.
3. **Неправильно обраний paired/independent test.**
4. **Ігнорування припущень методу.**
5. **Порівняння p-value, отриманих різними алгоритмами, без розуміння поправок.**
6. **Надмірне округлення проміжних обчислень.**
7. **Відсутність резервної копії сирих даних.**

43. Мінімальний чек-лист перед статистичним аналізом

Питання	Так / Ні
Один рядок відповідає одному спостереженню?	□
Один стовпець відповідає одній змінній?	□
Тип кожної змінної визначено?	□
Одиниці вимірювання єдині?	□
Пропуски позначені коректно?	□
Дублікати перевірені?	□
Підозрілі значення звірені з первинним джерелом?	□
Сирий файл збережений окремо?	□
Визначено статистичний метод до натискання кнопки «Analyze»?	□

44. Зв'язок із практичною роботою №3

Лекція 6 завершує підготовку до ПР3. У практичній роботі студент:

- працює з правильно структурованою кількісною вибіркою;
- обчислює описові статистики вручну та в Excel;
- розрізняє s і SE;
- будує довірчі інтервали;
- використовує Excel як основне навчальне середовище;
- використовує Real Statistics або Jamovi для незалежної перевірки;
- порівнює результати різних способів обчислення;
- формулює статистичний та технологічний висновки.

45. Підсумок лекції

дизайн → сирі дані → структура → очищення → типи змінних → Excel → спеціалізоване ПЗ → перевірка →

Статистичний аналіз починається задовго до обчислення p-value. Правильна структура даних, збереження первинної інформації, прозора процедура очищення і незалежна перевірка результатів є невід'ємними частинами наукової доброчесності та якості дослідження.

46. Контрольні питання

1. Що є об'єктом спостереження?
2. Що називають змінною?
3. Чим кількісна змінна відрізняється від категоріальної?
4. Наведіть приклади номінальної та порядкової змінних у харчових технологіях.
5. Що означає принцип «один рядок — одне спостереження»?
6. У чому різниця між довгим і широким форматом даних?
7. Що називають сирими даними?
8. Чому не слід зберігати лише групові середні?
9. Для чого потрібний ID?
10. Чому ID не повинен аналізуватися як звичайна кількісна змінна?
11. Як слід позначати пропущені значення?
12. Чому небезпечно кодувати пропуск числом 0 або 999?
13. Як перевіряти дублікати?
14. Чому два однакові числові значення не обов'язково є дубльованими записами?
15. Що слід робити з потенційним викидом?
16. Чому не можна вилучати дані лише через небажаний статистичний результат?
17. Для чого потрібний словник змінних?
18. Яка роль Excel у цьому курсі?
19. Які можливості дає Analysis ToolPak?
20. Для чого використовується Real Statistics?
21. Яка методична роль Jamovi?
22. Чому R/RStudio є важливим для професійної статистики, але не є основним середовищем цього курсу?
23. Що означає відтворюваність статистичного аналізу?
24. Чому сирі дані потрібно зберігати окремо?
25. Чому програмне забезпечення не може самостійно визначити коректний статистичний метод?
26. Що потрібно перевірити, якщо ручний розрахунок та програмний результат не збігаються?

Лекція 7

Методологія перевірки статистичних гіпотез

1. Мета лекції та результати навчання

Мета лекції — сформувати системне розуміння логіки перевірки статистичних гіпотез як інструменту прийняття рішень у дослідженнях харчових технологій.

Після опрацювання лекції здобувач повинен уміти:

- формулювати нульову та альтернативну статистичні гіпотези;
- розрізняти односторонні та двосторонні гіпотези;
- пояснювати зміст рівня значущості α;
- розрізняти помилки I та II роду;
- пояснювати зміст статистичної потужності;
- правильно інтерпретувати p-value;
- відрізняти статистичну значущість від практичної або технологічної значущості;
- пояснювати роль розміру ефекту;
- розрізняти параметричні та непараметричні критерії;
- вибирати загальну послідовність статистичної перевірки.

2. План лекції

1. Що таке статистична гіпотеза.
2. Нульова та альтернативна гіпотези.
3. Односторонні та двосторонні перевірки.
4. Статистика критерію та її розподіл.
5. Рівень значущості.
6. Помилки I та II роду.
7. Статистична потужність.
8. p-value та його правильна інтерпретація.
9. Статистична й практична значущість.
10. Розмір ефекту.
11. Параметричні та непараметричні критерії.
12. Загальний алгоритм перевірки гіпотез.

3. Від опису даних до статистичного рішення

Описова статистика відповідає на питання: що спостерігаємо у конкретній вибірці? Наприклад, можна отримати:

$$\bar{x}_A = 12.60, \qquad \bar{x}_B = 12.35.$$

Різниця вибіркових середніх:

$$\bar{x}_A - \bar{x}_B = 0.25.$$

Але цього недостатньо, щоб стверджувати, що два технологічні режими відрізняються у генеральних сукупностях. Різниця могла виникнути через випадкову вибіркову мінливість.

Статистична перевірка гіпотез — формалізована процедура, яка дозволяє оцінити, наскільки спостережувані дані сумісні з певним припущенням про генеральну сукупність.

4. Статистична гіпотеза

Статистична гіпотеза — твердження про параметр, розподіл або іншу властивість генеральної сукупності, яке перевіряють за вибірковими даними.

Приклади:

- середня вологість після двох режимів однакова;
- генеральні дисперсії двох процесів рівні;
- дані узгоджуються з нормальним розподілом;
- кореляція між температурою і виходом продукту дорівнює нулю;
- усі чотири середні у ANOVA однакові.

5. Нульова гіпотеза

Нульова гіпотеза H_0 зазвичай описує відсутність ефекту, відмінності або зв'язку.

Для порівняння двох середніх:

$$\boxed{H_0: \mu_A = \mu_B}.$$

Або:

$$H_0: \mu_A - \mu_B = 0.$$

6. Альтернативна гіпотеза

Альтернативна гіпотеза H_1 формулює варіант, який розглядають замість H_0.

6.1. Двостороння альтернатива

$$\boxed{H_1: \mu_A \neq \mu_B}.$$

Вона перевіряє наявність будь-якої різниці незалежно від її напряму.

6.2. Правостороння альтернатива

$$H_1: \mu_A > \mu_B.$$

6.3. Лівостороння альтернатива

$$H_1: \mu_A < \mu_B.$$

7. Одностороння чи двостороння перевірка

Напрям альтернативи потрібно визначати **до аналізу даних**, виходячи з дослідницького питання.

Двостороння перевірка доречна, якщо важлива будь-яка відмінність.

Одностороння — лише якщо протилежний напрям не має сенсу або не становить дослідницького інтересу.

Не можна спочатку подивитися на результат, а потім вибрати односторонню гіпотезу лише для отримання меншого p-value.

8. Статистика критерію

Для перевірки гіпотези з даних обчислюють певну величину — **статистику критерію**.

Наприклад, у t-тесті:

$$t = \frac{\text{спостережувана різниця}}{\text{стандартна похибка різниці}}.$$

Статистика показує, наскільки далеко спостережуваний результат віддалений від очікуваного за H_0 у одиницях його випадкової мінливості.

9. Розподіл статистики за умови H_0

Кожний статистичний критерій спирається на відомий або наближено відомий розподіл статистики за умови, що нульова гіпотеза істинна.

Задача	Статистика	Типовий розподіл
Порівняння середніх	t	t-розподіл
Порівняння дисперсій	F	F-розподіл
ANOVA	F	F-розподіл
Узгодженість розподілу	χ^2	χ^2-розподіл

10. Рівень значущості α

α — заздалегідь задана ймовірність помилки I роду, яку дослідник готовий прийняти.

Найчастіше:

$$\boxed{\alpha = 0.05}.$$

Також можуть використовуватися 0.01 або 0.10 залежно від задачі.

11. Помилка I роду

Помилка I роду — відхилення істинної нульової гіпотези.

$$\boxed{P(\text{відхилити } H_0 \mid H_0 \text{ істинна}) = \alpha}.$$

У технологічному контексті це може означати висновок про відмінність режимів, якої насправді немає.

12. Помилка II роду

Помилка II роду — невідхилення H_0, коли насправді альтернативна гіпотеза істинна.

$$\boxed{P(\text{не відхилити } H_0 \mid H_1 \text{ істинна}) = \beta}.$$

Наприклад, реальний технологічний ефект існує, але вибірка надто мала, щоб його надійно виявити.

13. Статистична потужність

Потужність критерію:

$$\boxed{Power = 1 - \beta}.$$

Це ймовірність виявити реальний ефект заданої величини.

13.1. Від чого залежить потужність

- обсяг вибірки;
- реальна величина ефекту;
- розсіювання даних;
- рівень α;
- тип статистичного критерію;
- якість дизайну експерименту.

14. Вплив обсягу вибірки

Зі збільшенням n стандартні похибки зменшуються, тому навіть відносно невеликий реальний ефект легше виявити.

$$SE \propto \frac{1}{\sqrt{n}}.$$

15. p-value

p-value — ймовірність за умови істинності H_0 отримати статистику не менш екстремальну, ніж фактично спостережувана.

$$\boxed{p = P(\text{настільки ж або більш екстремальні дані} \mid H_0)}.$$

15.1. Правило рішення

Якщо:

$$p < \alpha,$$

нульову гіпотезу відхиляють.

Якщо:

$$p \geq \alpha,$$

немає достатніх підстав відхилити H_0.

16. Що p-value НЕ означає

1. p не є ймовірністю того, що H_0 істинна.
2. $1 - p$ не є ймовірністю того, що H_1 істинна.
3. Малий p не означає великий ефект.
4. Великий p не доводить відсутність ефекту.
5. $p = 0.049$ і $p = 0.051$ не описують принципово різні реальності.

17. Приклад інтерпретації p-value

Нехай у порівнянні двох технологічних режимів:

$$p = 0.0042.$$

За $\alpha = 0.05$:

$$p < \alpha.$$

Отже, дані погано узгоджуються з припущенням про рівність генеральних середніх, і H_0 відхиляється.

Але для повного висновку потрібно також оцінити величину різниці, довірчий інтервал і розмір ефекту.

18. Статистична значущість

Результат називають статистично значущим, якщо відповідно до обраного критерію:

$$p < \alpha.$$

Це означає, що спостережуваний результат малоймовірний за нульової гіпотези.

19. Практична або технологічна значущість

Практична значущість відповідає на інше питання:

чи має величина ефекту реальне технологічне, економічне або якісне значення?

Наприклад, різниця середніх 0.02% при дуже великій вибірці може бути статистично значущою, але технологічно несуттєвою.

$$\boxed{\text{статистична значущість} \neq \text{практична значущість}}.$$

20. Розмір ефекту

Розмір ефекту кількісно описує силу відмінності або зв'язку.

У цьому курсі використовуються:

- Cohen's d для різниці двох середніх;
- rank-biserial correlation для Mann–Whitney та Wilcoxon;
- η^2 для ANOVA;
- r для кореляції;

- R^2 для регресії.

21. Cohen's d — концептуально

Для двох груп:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}}.$$

Він показує різницю середніх у одиницях типового внутрішньогрупового стандартного відхилення.

21.1. Орієнтовна інтерпретація

$\lvert d \rvert$ Умовна і	нтерпретація
≈0.2	малий ефект
≈0.5	середній
≈0.8 і більше	великий

Ці межі не є універсальними технологічними стандартами; предметна інтерпретація важливіша.

22. Довірчий інтервал як доповнення до p-value

p-value відповідає переважно на питання про сумісність даних з H_0. Довірчий інтервал додатково показує діапазон правдоподібних величин ефекту.

Наприклад:

$$\Delta\mu = 0.25, \qquad 95\%\, CI = [0.09;\ 0.41].$$

Такий результат значно інформативніший за саме $p = 0.004$.

23. Параметричні критерії

Параметричні критерії спираються на модель розподілу і параметри генеральної сукупності.

Приклади:

- t-критерій;
- F-критерій;
- ANOVA;
- кореляція Пірсона;
- лінійна регресія.

23.1. Типові припущення

- незалежність;
- приблизна нормальність;
- однорідність дисперсій — для деяких процедур;
- кількісна шкала.

24. Непараметричні критерії

Непараметричні методи мають слабші припущення щодо конкретної форми розподілу і часто використовують ранги.

У курсі:

- Mann–Whitney;
- Wilcoxon signed-rank;
- Spearman correlation.

25. «Непараметричний» не означає «без припущень»

Навіть непараметричні критерії вимагають:

- правильного дизайну;
- незалежності або коректної парності;
- відповідного типу шкали;
- коректної інтерпретації.

26. Загальний алгоритм перевірки статистичної гіпотези

дослідницьке питання → H_0, H_1 → перевірка припущень → вибір критерію → статистика → p → CI + effec

26.1. Покроково

1. Визначити дизайн і тип змінних.
2. Сформулювати H_0 та H_1.
3. Задати α.
4. Перевірити статистичні припущення.
5. Обрати критерій.
6. Обчислити статистику та p-value.
7. Прийняти статистичне рішення.
8. Оцінити довірчий інтервал.
9. Оцінити розмір ефекту.
10. Сформулювати технологічний висновок.

27. «Прийняти H_0» чи «не відхилити H_0»?

У класичній перевірці коректніше говорити:

немає достатніх підстав відхилити H_0.

Великий p-value не доводить істинність H_0; він може бути наслідком малого обсягу вибірки або низької потужності.

28. Вплив великої вибірки

При дуже великому n статистичний тест може виявити дуже малі ефекти.

Тому для великих вибірок особливо важливо звітувати:

- величину ефекту;
- довірчий інтервал;
- предметну інтерпретацію.

29. Вплив малої вибірки

При малому n:

- оцінки менш точні;
- довірчі інтервали ширші;
- потужність нижча;
- припущення про розподіл стають важливішими;
- окремі викиди сильніше впливають на результат.

30. Попереднє планування обсягу вибірки

У хорошому дослідженні бажано визначати n до збору даних, виходячи з:

- очікуваного ефекту;
- допустимого α;
- бажаної потужності;
- очікуваної варіативності.

Концептуально:

$$\boxed{n = f(\alpha, \beta, \text{effect size}, \sigma)}.$$

31. Множинні перевірки і накопичення помилки I роду

Якщо виконувати багато незалежних тестів на рівні $\alpha = 0.05$, загальна ймовірність хоча б одного хибнопозитивного результату зростає.

Саме тому для порівняння 3+ середніх використовується ANOVA, а не велика серія окремих t-тестів.

32. Типові помилки

1. Формулювати H_1 після перегляду даних.
2. Вважати p ймовірністю істинності H_0.
3. Вважати $p > 0.05$ доказом відсутності ефекту.
4. Ототожнювати статистичну і технологічну значущість.
5. Не подавати effect size.
6. Ігнорувати припущення критерію.
7. Застосовувати односторонній тест без попереднього обґрунтування.
8. Робити багато парних тестів без контролю множинності.
9. Повідомляти лише «значуще/незначуще» без числового результату.

33. Зв'язок із ПР4

ПР4 реалізує цю методологію у конкретному кейсі порівняння двох технологічних процесів:

- формулювання H_0 та H_1;
- перевірка нормальності;
- перевірка рівності дисперсій;
- вибір Student або Welch;
- обчислення t і p-value;
- 95% CI різниці середніх;
- Cohen's d;
- статистичний та технологічний висновки.

34. Підсумок лекції

$H_0/H_1 \rightarrow \alpha \rightarrow$ припущення $\rightarrow$ критерій $\rightarrow$ статистика $\rightarrow p \rightarrow CI \rightarrow$ effect size $\rightarrow$ технологічний висновок.

Перевірка гіпотез — це не процедура «пошуку p<0.05», а частина системного статистичного рішення, яке має включати припущення, невизначеність і величину ефекту.

35. Контрольні питання

1. Що називають статистичною гіпотезою?
2. Чим H_0 відрізняється від H_1?
3. Коли використовують двосторонню альтернативу?
4. Чому напрям гіпотези потрібно визначати до аналізу даних?
5. Що таке статистика критерію?
6. Що означає рівень значущості α?
7. Що називають помилкою I роду?
8. Що називають помилкою II роду?
9. Що таке статистична потужність?
10. Від чого залежить потужність?
11. Що означає p-value?
12. Чи є p-value ймовірністю істинності H_0?
13. Що означає $p > 0.05$?
14. У чому різниця між статистичною та практичною значущістю?
15. Що таке розмір ефекту?
16. Які effect size використовуються у цьому курсі?
17. Чому довірчий інтервал доповнює p-value?
18. Чим параметричні критерії відрізняються від непараметричних?
19. Чи означає «непараметричний» повну відсутність припущень?
20. Чому коректніше говорити «не відхиляємо H_0», а не «приймаємо H_0»?

21. Чому при великому n потрібно особливо враховувати effect size?
22. Чому при малому n зростає ризик помилки II роду?

Лекція 8

Перевірка закону розподілу та статистичних припущень

1. Мета лекції та результати навчання

Мета лекції — навчити оцінювати, чи виконуються припущення, на яких ґрунтуються класичні параметричні статистичні методи, та правильно поєднувати графічні й формальні способи діагностики даних.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати, навіщо перед параметричним тестом перевіряти статистичні припущення;
- оцінювати форму розподілу за гістограмою, boxplot та Q–Q plot;
- формулювати гіпотези для Shapiro–Wilk test;
- правильно інтерпретувати p-value тесту нормальності;
- пояснювати вплив обсягу вибірки на формальні тести нормальності;
- розрізняти потенційний викид, помилку та реальне нетипове спостереження;
- пояснювати припущення про однорідність дисперсій;
- застосовувати класичний F-критерій і розуміти його чутливість до ненормальності;
- пояснювати роль критерію Levene;
- обирати між класичним параметричним, робастним або непараметричним підходом.

2. План лекції

1. Статистичні припущення як частина методу.
2. Візуальна оцінка розподілу.
3. Гістограма та boxplot.
4. Q–Q plot.
5. Shapiro–Wilk test.
6. Вплив обсягу вибірки на перевірку нормальності.
7. Викиди та їх вплив.
8. Однорідність дисперсій.
9. F-критерій рівності дисперсій.
10. Levene test.
11. Алгоритм вибору подальшого методу.

3. Чому статистичні припущення важливі

Формула статистичного критерію не є універсальною. Її теоретичний розподіл виведено за певних умов. Якщо вони суттєво порушені, p-value, критичні значення та довірчі інтервали можуть бути некоректними.

Принцип. Перед застосуванням методу потрібно перевірити не лише «які дані маємо», а й «за яких умов цей метод працює коректно».

4. Типові припущення параметричних методів

- незалежність спостережень;
- кількісний характер результативної змінної;
- приблизна нормальність у відповідних групах або нормальність залишків;
- однорідність дисперсій — якщо її вимагає конкретний метод;
- відсутність грубих помилок, що визначають результат.

5. Незалежність — припущення, яке не можна «перевірити кнопкою»

Незалежність визначається дизайном дослідження.

Наприклад, якщо одну й ту саму партію виміряли двічі, ці спостереження не є незалежними лише тому, що записані у двох рядках.

Так само оцінки одного дегустатора до і після зміни рецептури є парними.

6. Графічна діагностика перед формальним тестом

Доцільно починати з графіків, оскільки вони можуть показати структуру, яку один p-value не відображає.

Базовий набір:

- гістограма;
- boxplot;
- Q–Q plot;
- для часових даних — графік у порядку спостереження.

7. Гістограма і нормальність

Для приблизно нормальних даних гістограма часто має:

- один виражений центр;
- приблизну симетрію;
- поступове зменшення частот до країв;
- відсутність сильних окремих піків або довгого хвоста.

Але гістограма сама по собі не доводить нормальність. За малого n вона може бути дуже нестабільною, а її вигляд суттєво залежить від числа інтервалів.

8. Boxplot як діагностичний графік

Boxplot допомагає побачити:

- медіану;
- центральні 50% даних;
- асиметрію;

- потенційні викиди;
- відмінності розсіювання між групами.

Для двох процесів boxplot особливо корисний для початкового порівняння центрів та варіативності.

9. Q–Q plot

Q–Q plot порівнює емпіричні квантілі вибірки з теоретичними квантилями заданого розподілу, найчастіше нормального.

Якщо дані приблизно нормальні, точки розташовуються поблизу прямої.

9.1. Типові відхилення

- S-подібна форма — можливі відмінності у хвостах;
- систематичне викривлення — асиметрія;
- одна-дві точки далеко від лінії — можливі викиди;
- кластерна структура — можлива суміш різних груп.

10. Формальна перевірка нормальності

Одним із найпоширеніших критеріїв є Shapiro–Wilk test.

Нульова гіпотеза:

$$\boxed{H_0\text{:вибірка узгоджується з нормальним розподілом}}.$$

Альтернативна:

$$H_1\text{:розподіл відрізняється від нормального.}$$

11. Статистика Shapiro–Wilk

Концептуально тест порівнює впорядковані вибіркові значення з тим, чого очікували б від нормального розподілу.

Статистика позначається W і лежить поблизу 1 для даних, добре узгоджених із нормальністю.

Повний ручний розрахунок W у цьому курсі не виконується; використовується спеціалізоване ПЗ.

12. Інтерпретація Shapiro–Wilk

За $\alpha = 0.05$:

- якщо $p < 0.05$ — є статистичні підстави відхилити нормальність;
- якщо $p \geq 0.05$ — немає достатніх підстав відхилити нормальність.

$$\boxed{p > 0.05 \neq \text{«нормальність доведена»}}.$$

13. Чому p-value тесту нормальності залежить від n

За малого n тест може мати низьку потужність і не виявити помітне відхилення.

За дуже великого n навіть невелике й практично неважливе відхилення від ідеальної нормальності може дати мале p-value.

13.1. Тому рішення має спиратися на комбінацію

$$\boxed{\text{графік} + \text{Shapiro–Wilk} + \text{обсяг вибірки} + \text{стійкість обраного методу}}.$$

14. Приклад інтерпретації

Для двох вибірок отримано:

$$W_A = 0.984, \quad p_A = 0.982;$$

$$W_B = 0.966, \quad p_B = 0.849.$$

Для обох груп $p > 0.05$. Отже, дані не дають підстав відхиляти нормальну модель.

15. Викиди та нормальність

Один нетиповий результат може:

- погіршити Shapiro–Wilk;
- суттєво змінити середнє;
- збільшити дисперсію;
- змінити результат F- і t-тесту.

16. Що робити з підозрілим значенням

1. Перевірити первинний запис.
2. Перевірити одиниці вимірювання.
3. З'ясувати, чи була технологічна подія.
4. Оцінити графічно.
5. Зафіксувати рішення.
6. За потреби провести аналіз із точкою та без неї як sensitivity analysis.

Не можна вилучати значення лише тому, що після його вилучення p-value стає «кращим».

17. Однорідність дисперсій

Для класичного Student t-test із pooled variance припускається:

$$\boxed{\sigma_1^2 = \sigma_2^2}.$$

Це називають гомоскедастичністю або однорідністю дисперсій.

18. Навіщо перевіряти дисперсії

Якщо групи мають суттєво різне розсіювання, pooled Student t-test використовує невідповідну модель стандартної похибки.

У такій ситуації кращим вибором є Welch t-test.

19. Класичний F-критерій рівності дисперсій

Для двох незалежних нормальних вибірок:

$$H_0: \sigma_1^2 = \sigma_2^2.$$

Статистика:

$$\boxed{F = \frac{s_1^2}{s_2^2}}.$$

Часто більшу вибіркову дисперсію ставлять у чисельник, щоб:

$$F \geq 1.$$

20. Ступені свободи F-розподілу

$$\nu_1 = n_1 - 1, \qquad \nu_2 = n_2 - 1.$$

F-розподіл асиметричний і залежить від двох чисел ступенів свободи.

21. Рішення за F-критерієм

Можна використовувати критичну область або p-value.

За двостороннього тесту та $\alpha = 0.05$:

- $p_F < 0.05$ — відхиляємо рівність дисперсій;
- $p_F \geq 0.05$ — немає достатніх підстав відхиляти її.

22. Приклад F-тесту

Нехай:

$$s_A^2 = 0.03333, \qquad s_B^2 = 0.02500.$$

Тоді:

$$F = \frac{0.03333}{0.02500} = 1.3333.$$

Для $n_A = n_B = 10$:

$$\nu_1 = \nu_2 = 9.$$

Якщо двосторонній $p_F \approx 0.675$, немає підстав відхиляти рівність дисперсій.

23. Чутливість F-критерію до ненормальності

Класичний F-test є чутливим до:

- асиметрії;
- важких хвостів;
- викидів.

Тому формальний висновок про дисперсії слід робити після аналізу нормальності.

24. Критерій Levene

Levene test — більш робастний метод перевірки однорідності дисперсій.

Гіпотези:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2.$$

У двогруповому випадку — аналогічно рівність двох дисперсій.

24.1. Ідея Levene

Замість прямого аналізу відношення дисперсій він працює з абсолютними відхиленнями спостережень від центру групи.

У робастних модифікаціях часто використовують медіану.

25. F або Levene?

Критерій	Перевага	Обмеження
F-test	Класичний, простий ручний розрахунок	Чутливий до ненормальності
Levene	Стійкіший до відхилення від нормальності	Зазвичай виконується програмно

У ПР4 студент вручну виконує F-критерій, а Levene використовується як сучасна програмна перевірка.

26. Незначущий Levene не доводить ідентичність дисперсій

Як і в інших тестах:

$$p > 0.05$$

означає лише відсутність достатніх доказів проти рівності. Особливо при малих вибірках тест може мати низьку потужність.

27. Практична оцінка співвідношення дисперсій

Окрім формального тесту, корисно порівнювати:

$$\frac{s^2_{\max}}{s^2_{\min}}$$

та boxplot груп. Це допомагає оцінити практичну величину різниці у варіативності.

28. Коли переходити до Welch t-test

Welch доцільний, якщо:

- дисперсії суттєво відрізняються;
- обсяги груп різні;
- немає сильного предметного обґрунтування рівності дисперсій.

Сучасна практика. Welch t-test часто розглядають як безпечний default для двох незалежних середніх, оскільки при рівних дисперсіях він майже не втрачає ефективності, а при нерівних є коректнішим.

29. Коли параметричний t-test усе ще може бути робастним

t-тести досить стійкі до помірного відхилення від нормальності, якщо:

- немає сильних викидів;
- розподіли не дуже асиметричні;
- обсяги вибірок достатні;
- групи мають близькі форми.

30. Коли варто перейти до непараметричного підходу

Для двох незалежних груп Mann–Whitney може бути доречним, якщо:

- дані порядкові;
- сильна асиметрія;
- виражені викиди;
- дуже мала вибірка й нормальність не підтримується;
- інтерпретація рангової різниці відповідає дослідницькому питанню.

31. Важливе уточнення про Mann–Whitney

Mann–Whitney не є буквально «тестом медіан» за будь-яких умов. Інтерпретація як зсув центрального положення найбільш природна, коли форми розподілів груп подібні.

32. Алгоритм перевірки припущень для ПР4

$$\boxed{\text{дані} \rightarrow \text{boxplot/Q-Q} \rightarrow \text{Shapiro–Wilk} \rightarrow \text{F/Levene} \rightarrow \text{Student або Welch} \rightarrow \text{t-test}}.$$

32.1. Розгорнуто

1. Обчислити описові статистики.
2. Побудувати boxplot.
3. Оцінити Q–Q plot.
4. Виконати Shapiro–Wilk у кожній групі.
5. Якщо грубих порушень немає — перевірити дисперсії.
6. Виконати F-test вручну.
7. За можливості перевірити Levene програмно.
8. При однорідних дисперсіях — Student pooled.
9. При неоднорідних — Welch.
10. При сильній ненормальності — розглянути Mann–Whitney.

33. Типові помилки

1. Перевіряти лише Shapiro–Wilk без графіків.
2. Трактувати $p > 0.05$ як доказ нормальності.
3. Ігнорувати вплив n на тест нормальності.
4. Видаляти викиди автоматично.

5. Виконувати F-test при явній сильній ненормальності без застережень.
6. Використовувати pooled Student після значущого тесту неоднорідності.
7. Вважати Levene «доказом однакових дисперсій» при $p > 0.05$.
8. Переходити до Mann–Whitney лише тому, що Shapiro–Wilk дав $p = 0.049$, ігноруючи графіки та робастність t-тесту.

34. Зв'язок із ПР4

У ПР4 студент повинен не механічно виконати три тести, а побудувати логічний маршрут:

- Shapiro–Wilk — чи є підстави сумніватися в нормальності;
- F-test / Levene — чи припустимо об'єднувати дисперсії;
- Student або Welch — який t-test відповідає структурі даних;
- при сильних порушеннях — перехід до непараметричної логіки ПР5.

35. Підсумок лекції

припущення → графіки → нормальність → викиди → однорідність дисперсій → вибір тесту.

Перевірка статистичних припущень — це не формальна «галочка» перед t-тестом, а частина обґрунтування того, чому конкретний статистичний висновок можна вважати коректним.

36. Контрольні питання

1. Навіщо перевіряти припущення статистичного критерію?
2. Які припущення типові для параметричних методів?
3. Чому незалежність визначається дизайном, а не Shapiro–Wilk?
4. Яку інформацію дає гістограма?
5. Що дозволяє побачити Q–Q plot?
6. Які гіпотези перевіряє Shapiro–Wilk?
7. Що означає $p > 0.05$ у Shapiro–Wilk?
8. Чому результат тесту нормальності залежить від n?
9. Як викид може вплинути на t-test?
10. Чому не можна автоматично вилучати викиди?
11. Що означає однорідність дисперсій?
12. Запишіть статистику F-критерію.
13. Які ступені свободи має F-test?
14. Чому F-test чутливий до ненормальності?
15. Для чого використовується Levene test?
16. Чим Levene відрізняється від класичного F-test?
17. Коли потрібно використовувати Welch t-test?
18. Чому Welch часто вважають безпечним вибором?
19. Коли доцільно переходити до Mann–Whitney?

20. Чому Mann–Whitney не завжди коректно називати «тестом медіан»?

Лекція 9

Порівняння двох технологічних процесів: F-критерій, t-критерій Стьюдента, Welch, довірчі інтервали та розмір ефекту

1. Мета лекції та результати навчання

Мета лекції — сформувати цілісний алгоритм статистичного порівняння двох незалежних технологічних процесів із урахуванням нормальності, однорідності дисперсій, вибору між Student та Welch, оцінювання довірчого інтервалу різниці середніх і розміру ефекту.

Після опрацювання лекції здобувач повинен уміти:

- розрізняти незалежні та парні вибірки;
- формулювати гіпотези щодо двох генеральних середніх;
- пояснювати логіку pooled Student t-test;
- обчислювати pooled variance, стандартну похибку різниці та t-статистику;
- виконувати Welch t-test за нерівних дисперсій;
- розуміти наближені ступені свободи Welch–Satterthwaite;
- будувати 95% довірчий інтервал для різниці середніх;
- обчислювати Cohen's d та інтерпретувати величину ефекту;
- відрізняти статистичну значущість від технологічної;
- формулювати повний висновок порівняльного аналізу.

2. План лекції

1. Дизайн двох незалежних вибірок.
2. Описова частина аналізу.
3. Гіпотези щодо середніх.
4. Класичний Student t-test при рівних дисперсіях.
5. Pooled variance і стандартна похибка.
6. Welch t-test при нерівних дисперсіях.
7. Довірчий інтервал різниці середніх.
8. Cohen's d.
9. Статистична і технологічна значущість.
10. Повний алгоритм ПР4.

3. Незалежні та залежні вибірки

Перед вибором t-тесту потрібно визначити структуру даних.

3.1. Незалежні вибірки

Спостереження в одній групі не пов'язані зі спостереженнями в іншій.

Приклад: дві окремі партії продукту, виготовлені за режимами A та B.

3.2. Парні вибірки

Кожному спостереженню в одній умові відповідає конкретне спостереження в іншій.

Приклади: той самий дегустатор до/після; той самий зразок до/після обробки.

Неправильне визначення незалежності або парності — помилка дизайну, а не лише обчислення.

4. Постановка задачі для двох незалежних процесів

Маємо дві вибірки:

$$X_1, \dots, X_{n_1}$$

та:

$$Y_1, \dots, Y_{n_2}.$$

Потрібно перевірити, чи відрізняються генеральні середні:

$$\mu_1, \qquad \mu_2.$$

5. Гіпотези

Для двостороннього порівняння:

$$\boxed{H_0: \mu_1 = \mu_2}$$

$$\boxed{H_1: \mu_1 \neq \mu_2}.$$

Еквівалентно:

$$H_0: \mu_1 - \mu_2 = 0.$$

6. Попередній опис даних

До t-тесту потрібно обчислити для кожної групи:

- обсяг n;
- середнє $\bar{x}$;
- медіану;
- дисперсію s^2;
- стандартне відхилення s;
- мінімум і максимум.

Також доцільно побудувати boxplot і Q–Q plots.

7. Спостережувана різниця середніх

Основна ефектна величина:

$$\boxed{\Delta\bar{x} = \bar{x}_1 - \bar{x}_2}.$$

Це вибіркова оцінка генеральної різниці:

$$\mu_1 - \mu_2.$$

8. Student t-test при рівних дисперсіях

Класичний двовибірковий t-тест Стьюдента з pooled variance використовує припущення:

$$\sigma_1^2 = \sigma_2^2.$$

8.1. Об'єднана дисперсія

$$\boxed{s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

8.2. Об'єднане стандартне відхилення

$$s_p = \sqrt{s_p^2}.$$

8.3. Стандартна похибка різниці середніх

$$\boxed{SE_\Delta = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

8.4. t-статистика

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{SE_\Delta}}.$$

8.5. Ступені свободи

$$\boxed{\nu = n_1 + n_2 - 2}.$$

9. Приклад Student t-test

Нехай:

$$n_A = n_B = 10,$$

$$\bar{x}_A = 12.600, \qquad \bar{x}_B = 12.350,$$

$$s_A^2 = 0.033333, \qquad s_B^2 = 0.025000.$$

9.1. Різниця середніх

$$\Delta\bar{x} = 12.600 - 12.350 = 0.250.$$

9.2. Pooled variance

$$s_p^2 = \frac{9 \cdot 0.033333 + 9 \cdot 0.025000}{18} = 0.0291667.$$

9.3. Pooled SD

$$s_p = \sqrt{0.0291667} \approx 0.17078.$$

9.4. Standard error

$$SE_\Delta = 0.17078 \sqrt{\frac{1}{10} + \frac{1}{10}} \approx 0.07638.$$

9.5. t-статистика

$$t = \frac{0.250}{0.07638} \approx 3.273.$$

9.6. Ступені свободи

$$\nu = 10 + 10 - 2 = 18.$$

Для двостороннього тесту:

$$p \approx 0.0042.$$

Отже, при $\alpha = 0.05$ нульова гіпотеза відхиляється.

10. Welch t-test

Якщо дисперсії не можна вважати рівними, використовується Welch t-test.

Його t-статистика:

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}.$$

10.1. Стандартна похибка Welch

$$\boxed{SE_{\Delta,W} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

10.2. Ступені свободи Welch–Satterthwaite

$$\boxed{\nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}}.$$

Число ступенів свободи може бути нецілим.

11. Чому Welch коректніший при нерівних дисперсіях

Welch не об'єднує дві дисперсії в одну pooled estimate, а враховує їх окремо в стандартній похибці.

Це особливо важливо, коли:

- s_1^2 і s_2^2 суттєво відрізняються;
- n_1 і n_2 різні;
- група з більшою дисперсією має менший обсяг.

12. Student чи Welch?

Ситуація	Рекомендований підхід
Нормальність прийнятна, дисперсії близькі	Student pooled або Welch
Нормальність прийнятна, дисперсії різні	Welch
Різні n і сумнів щодо рівності дисперсій	Welch
Сильна ненормальність / порядкова шкала	Розглянути Mann–Whitney

Практична рекомендація. У сучасній статистичній практиці Welch часто є надійним вибором за замовчуванням для двох незалежних середніх.

13. Критична область і p-value

Для двостороннього t-тесту:

$$|t_{\text{emp}}| > t_{1-\alpha/2;\nu}$$

еквівалентно статистично значущому результату.

Або використовують правило:

$$p < \alpha.$$

14. Довірчий інтервал різниці середніх

Для Student pooled:

$$\boxed{(\mu_1 - \mu_2) \in (\bar{x}_1 - \bar{x}_2) \pm t_{\text{crit}} SE_{\Delta}}.$$

Для Welch — та сама структура, але використовується Welch standard error та Welch df.

15. Приклад 95% CI різниці

Для попереднього прикладу:

$$\Delta\bar{x} = 0.250, \qquad SE_{\Delta} = 0.07638.$$

Для $\nu = 18$:

$$t_{0.975;18} \approx 2.101.$$

Гранична похибка:

$$\Delta_{CI} = 2.101 \cdot 0.07638 \approx 0.16046.$$

Тоді:

$$\boxed{95\%\, CI = [0.0895;\ 0.4105]}.$$

16. Зв'язок CI і t-тесту

Для двосторонньої перевірки $H_0: \mu_1 - \mu_2 = 0$:

- якщо 95% CI не містить 0 — результат значущий при $\alpha = 0.05$;
- якщо містить 0 — немає підстав відхиляти H_0.

17. Чому CI інформативніший за саме p-value

p-value не показує величину різниці. CI дає:

- напрям ефекту;
- оцінку його величини;
- невизначеність;
- діапазон значень, сумісних із даними.

18. Cohen's d

Для двох незалежних груп при використанні pooled standard deviation:

$$\boxed{d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}}.$$

19. Приклад Cohen's d

Для:

$$\bar{x}_A - \bar{x}_B = 0.250, \qquad s_p = 0.17078,$$

отримаємо:

$$d = \frac{0.250}{0.17078} \approx 1.464.$$

Це великий стандартизований ефект.

20. Орієнтири Cohen's d

$\lvert d \rvert$	Умовний орієнтир
0.2	малий
0.5	середній
0.8	великий

Але у харчовій технології практичне значення залежить від одиниць, нормативів, безпеки, витрат і технологічного контексту.

21. Знак Cohen's d

Знак залежить від порядку віднімання:

$$d = \frac{\bar{x}_A - \bar{x}_B}{s_p}.$$

Позитивний d означає $\bar{x}_A > \bar{x}_B$, від'ємний — навпаки.

Для величини ефекту часто розглядають $|d|$, але напрям потрібно зберігати в технологічній інтерпретації.

22. Статистична значущість проти технологічної

Припустимо:

$$p < 0.001$$

але різниця становить лише 0.01 процентного пункту і не має технологічного значення. Тоді статистична значущість є, але практичний ефект мінімальний.

Інша ситуація: різниця технологічно важлива, але $p = 0.08$ при малому n. Це може означати недостатню потужність, а не доведену відсутність ефекту.

23. Що повинен містити повний статистичний висновок

Для двох процесів бажано повідомляти:

- n_1, n_2;
- $\bar{x}_1, \bar{x}_2$;
- s_1, s_2;
- результат перевірки припущень;
- тип t-тесту;
- t, df, p-value;
- різницю середніх;
- 95% CI;
- Cohen's d;
- технологічний зміст.

24. Приклад повного статистичного формулювання

Для режиму A середня кінцева вологість становила 12.60% ($s = 0.183$), для режиму B — 12.35% ($s = 0.158$). Дані обох груп не суперечили нормальності, а F-test не виявив значущої різниці дисперсій. Student t-test показав статистично значущу різницю середніх: $t(18) = 3.27$, $p = 0.0042$. Середня різниця становила 0.25 процентного пункту, 95% CI [0.09; 0.41]. Cohen's $d = 1.46$, що відповідає великому стандартизованому ефекту.

25. Приклад технологічного формулювання

Режим B забезпечує нижчу кінцеву вологість порівняно з режимом A. Різниця не лише статистично значуща, але й велика відносно внутрішньогрупового розсіювання. Остаточна технологічна оцінка має врахувати цільовий діапазон вологості, енерговитрати та інші показники якості.

26. Чому не слід писати «режим B кращий» лише за t-test

Статистика може підтвердити різницю за конкретною змінною, але «кращий» режим може залежати також від:

- смакових характеристик;
- кольору;
- енерговитрат;
- виходу продукту;
- терміну зберігання;
- безпеки;
- вартості.

27. Незначущий результат: як звітувати

Не слід писати «процеси однакові», якщо $p > 0.05$. Коректніше:

статистично значущої різниці за наявними даними не виявлено.

Потрібно також дивитися на CI та effect size.

28. Приклад незнайденої значущості

Нехай:

$$\Delta\bar{x} = 0.18, \qquad 95\%\, CI = [-0.05;\ 0.41], \qquad p = 0.12.$$

Інтервал містить і нуль, і потенційно суттєву позитивну різницю. Це означає високу невизначеність, а не доказ еквівалентності.

29. Еквівалентність — окрема статистична задача

Щоб довести практичну еквівалентність процесів, потрібно заздалегідь визначити допустиму межу:

$$-\Delta_E < \mu_1 - \mu_2 < \Delta_E$$

і використовувати спеціальний equivalence testing, а не звичайний t-test з $p > 0.05$.

30. Парний t-test — коротко для розрізнення

Для парних даних аналізують не дві незалежні вибірки, а різниці:

$$d_i = y_i - x_i.$$

Тоді:

$$H_0: \mu_d = 0.$$

Це інша постановка задачі. У ПР4 використовується саме незалежний двовибірковий аналіз.

31. Excel: типові інструменти

Для ПР4 можуть використовуватися:

- `AVERAGE`;
- `VAR.S`;
- `STDEV.S`;

- `F.TEST`;
- `T.TEST`;
- Data Analysis → F-Test Two-Sample for Variances;
- Data Analysis → t-Test: Two-Sample Assuming Equal Variances;
- Data Analysis → t-Test: Two-Sample Assuming Unequal Variances.

32. Важливе застереження щодо T.TEST

У функції потрібно правильно задати:

- кількість хвостів;
- тип тесту;
- парність або незалежність;
- припущення про дисперсії.

Інакше Excel дасть числовий результат для неправильної статистичної моделі.

33. Real Statistics / Jamovi у ПР4

Спеціалізоване ПЗ доцільно використати для:

- Shapiro–Wilk;
- Q–Q plots;
- Levene test;
- Student / Welch t-test;
- довірчих інтервалів;
- effect size.

34. Повний алгоритм ПР4

дві вибірки → descriptives → boxplot/Q-Q → Shapiro–Wilk → F/Levene → Student/Welch → p → CI → d → т

35. Алгоритм вибору тесту

Крок	Рішення
1	Переконатися, що вибірки незалежні
2	Оцінити розподіли й викиди
3	Перевірити Shapiro–Wilk
4	При прийнятній нормальності — F/Levene
5	Рівні дисперсії → Student pooled
6	Нерівні дисперсії → Welch
7	Сильна ненормальність → розглянути Mann–Whitney

36. Типові помилки

1. Виконувати t-test без перевірки структури вибірок.
2. Використовувати pooled Student за нерівних дисперсій.
3. Ігнорувати Welch.
4. Звітувати лише p-value.

5. Не подавати різницю середніх та CI.
6. Ототожнювати великий $|t|$ з великою практичною різницею.
7. Ототожнювати $p > 0.05$ з еквівалентністю.
8. Вважати Cohen's d універсальним критерієм технологічної значущості.
9. Робити висновок «кращий режим» лише за одним показником.
10. Плутати незалежний t-test із парним.

37. Підсумок лекції

дизайн → припущення → Student/Welch → t, p → $CI_{\mu_1-\mu_2}$ → d → статистичний висновок → технологічна і

Порівняння двох процесів має завершуватися не фразою «p<0.05», а кількісною оцінкою різниці, її невизначеності, стандартизованої величини ефекту та предметного значення.

38. Контрольні питання

1. Чим незалежні вибірки відрізняються від парних?
2. Які гіпотези формулюють для двох незалежних середніх?
3. Яке припущення щодо дисперсій використовує pooled Student t-test?
4. Запишіть формулу pooled variance.
5. Як обчислюється pooled standard error різниці середніх?
6. Запишіть t-статистику Student.
7. Які df має pooled Student?
8. Коли потрібно застосовувати Welch t-test?
9. Як обчислюється standard error у Welch?
10. Чому df Welch можуть бути нецілими?
11. Чим Welch відрізняється від pooled Student?
12. Як будується CI різниці середніх?
13. Який зв'язок між 95% CI і двостороннім t-test при $\alpha = 0.05$?
14. Чому CI інформативніший за саме p-value?
15. Запишіть формулу Cohen's d.
16. Що означає $d = 1.0$?
17. Чому орієнтири 0.2/0.5/0.8 не слід трактувати як абсолютні технологічні нормативи?
18. Як звітувати статистично незначущий результат?
19. Чому $p > 0.05$ не доводить еквівалентність?
20. Чому висновок «режим кращий» не можна робити лише на основі t-test одного показника?
21. Які елементи повинен містити повний статистичний висновок ПР4?

Лекція 10

Непараметричні критерії у технологічних дослідженнях: Mann–Whitney та Wilcoxon signed-rank

1. Мета лекції та результати навчання

Мета лекції — сформувати розуміння логіки непараметричних методів порівняння двох вибірок та навчити правильно застосовувати рангові критерії у задачах органолептичного й технологічного аналізу.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати, чим непараметричні критерії відрізняються від параметричних;
- визначати, коли доцільно використовувати рангові методи;
- правильно формувати ранги та обробляти однакові значення (ties);
- розрізняти незалежні та парні вибірки;
- виконувати критерій Mann–Whitney для двох незалежних груп;
- виконувати критерій знакових рангів Wilcoxon для парних спостережень;
- правильно інтерпретувати статистики U, T_+, T_- та p-value;
- обчислювати рангові розміри ефекту;
- пояснювати обмеження інтерпретації Mann–Whitney як «тесту медіан»;
- формулювати статистичний і технологічний висновки для органолептичних даних.

2. План лекції

1. Навіщо потрібні непараметричні методи.
2. Параметричні та непараметричні критерії.
3. Ранг як основа непараметричного аналізу.
4. Однакові значення та середні ранги.
5. Mann–Whitney test для двох незалежних вибірок.
6. Інтерпретація статистики U.
7. Wilcoxon signed-rank test для парних даних.
8. Нульові різниці та ties у Wilcoxon.
9. Ранговий розмір ефекту.
10. Exact та asymptotic p-value.
11. Органолептичні дані та межі інтерпретації.
12. Алгоритм вибору критерію.

3. Чому параметричні методи не завжди підходять

У попередніх лекціях розглядалися t-тести, які порівнюють генеральні середні та спираються на кількісну шкалу й певні припущення щодо розподілу.

У харчових технологіях часто зустрічаються дані, для яких ці припущення є сумнівними або сама шкала має порядковий характер.

Приклади:

- дегустаційні оцінки за 5-, 7- або 9-бальною шкалою;
- рангове оцінювання привабливості продукту;
- невеликі вибірки з вираженою асиметрією;
- дані з викидами, які не є помилками;
- показники, для яких середнє не є найкращою характеристикою центру.

Непараметричні методи часто працюють не з самими числовими значеннями, а з їх порядком — рангами.

4. Параметричні та непараметричні критерії

Ознака	Параметричні	Непараметричні
Тип даних	Переважно кількісні	Кількісні або порядкові
Основний об'єкт	Середні, дисперсії, параметри	Ранги, положення розподілів
Нормальність	Часто важлива	Не вимагається у класичному вигляді
Чутливість до викидів	Вища	Зазвичай нижча
Інтерпретація	Часто проста для середніх	Потребує уважнішої предметної інтерпретації

Непараметричний ≠ «без припущень». Такі критерії все одно вимагають правильного дизайну, незалежності або парності та відповідного типу даних.

5. Ранг

Ранг — порядковий номер спостереження у впорядкованій сукупності.

Наприклад, для значень:

$$4,\ 7,\ 9,\ 12$$

ранги:

$$1,\ 2,\ 3,\ 4.$$

Важливим є не абсолютний розрив між значеннями, а їх взаємний порядок.

6. Однакові значення — ties

У дегустаційних оцінках однакові бали трапляються дуже часто. Якщо кілька значень займають декілька послідовних рангів, усім їм присвоюють **середній ранг**.

Приклад

Впорядковані значення:

$$5,\ 5,\ 5,\ 6,\ 6,\ 7.$$

Три значення 5 займають ранги 1, 2, 3, тому:

$$R(5) = \frac{1+2+3}{3} = 2.$$

Два значення 6 займають ранги 4 і 5:

$$R(6) = \frac{4+5}{2} = 4.5.$$

Для 7:

$$R(7) = 6.$$

7. Чому ties важливі

Якщо однаковим значенням присвоїти різні ранги довільно, результат критерію залежатиме від випадкового порядку запису, що є некоректним.

однакові значення → однаковий середній ранг.

8. Mann–Whitney test: призначення

Критерій Mann–Whitney використовується для порівняння **двох незалежних вибірок**.

Приклади:

- дві незалежні групи дегустаторів оцінюють два продукти;
- дві незалежні партії виготовлено за різними режимами;
- порівнюються дві різні групи споживачів.

9. Гіпотези Mann–Whitney

У загальному формулюванні:

H_0:розподіли двох груп не мають систематичного зсуву.

H_1:одна група має тенденцію до більших або менших значень.

Для двостороннього тесту перевіряється будь-який систематичний зсув.

10. Чи є Mann–Whitney тестом медіан?

Часто критерій спрощено називають «непараметричним тестом різниці медіан». Це не завжди коректно.

Важливо. Інтерпретація Mann–Whitney як тесту зсуву центрального положення або медіан найбільш природна тоді, коли форми розподілів у двох групах приблизно однакові.

Якщо форми суттєво різняться, значущий Mann–Whitney може відображати не лише різницю медіан, а загальну відмінність у розташуванні розподілів.

11. Алгоритм Mann–Whitney

1. Об'єднати дві вибірки.
2. Впорядкувати всі спостереження.
3. Присвоїти ранги.
4. Для ties використати середні ранги.
5. Повернути кожному рангу мітку групи.

6. Обчислити суму рангів кожної групи.
7. Обчислити U_A та U_B.
8. Вибрати менше значення U.
9. Визначити p-value або критичне значення.
10. Оцінити розмір ефекту.

12. Суми рангів

Нехай:

$$R_A = \sum R_i \quad \text{для групи A},$$

$$R_B = \sum R_i \quad \text{для групи B}.$$

Контрольна перевірка:

$$\boxed{R_A + R_B = \frac{N(N+1)}{2}},$$

де:

$$N = n_A + n_B.$$

13. Статистика Mann–Whitney

$$\boxed{U_A = n_A n_B + \frac{n_A(n_A+1)}{2} - R_A}.$$

$$\boxed{U_B = n_A n_B + \frac{n_B(n_B+1)}{2} - R_B}.$$

Виконується:

$$U_A + U_B = n_A n_B.$$

Для двостороннього ручного аналізу часто використовують:

$$\boxed{U = \min(U_A, U_B)}.$$

14. Демонстраційний приклад Mann–Whitney

Нехай дві незалежні групи дегустаторів оцінюють продукт за 9-бальною шкалою:

Продукт A	Продукт B
7, 8, 6, 9, 7, 8, 7, 6, 8, 7	5, 6, 6, 7, 5, 6, 7, 5, 6, 6

14.1. Частоти та ранги

Значення	Кількість	Ранги	Середній ранг
5	3	1–3	2
6	7	4–10	7
7	6	11–16	13.5
8	3	17–19	18
9	1	20	20

Суми рангів:

$$R_A = 142, \qquad R_B = 68.$$

Перевірка:

$$142 + 68 = 210 = \frac{20 \cdot 21}{2}.$$

14.2. Обчислення U

$$U_A = 10 \cdot 10 + \frac{10 \cdot 11}{2} - 142 = 13.$$

$$U_B = 10 \cdot 10 + \frac{10 \cdot 11}{2} - 68 = 87.$$

$$\boxed{U = 13}.$$

14.3. Висновок

Для $n_A = n_B = 10$ двостороннє критичне значення при $\alpha = 0.05$ становить приблизно 23.

$$U = 13 < 23.$$

Отже, нульову гіпотезу відхиляють. Продукт А має систематично вищі оцінки.

15. p-value для Mann–Whitney

У спеціалізованому ПЗ для наведеного прикладу p-value становить приблизно:

$$p \approx 0.004.$$

Точне значення може дещо залежати від:

- exact або asymptotic алгоритму;
- корекції на ties;
- continuity correction.

16. Exact та asymptotic підходи

16.1. Exact p-value

Обчислюється на основі точного розподілу статистики за нульової гіпотези.

Особливо корисний для малих вибірок.

16.2. Asymptotic p-value

Використовує наближення розподілу статистики нормальним при достатньо великих вибірках.

Різні програми можуть повертати трохи різні p-value через різні алгоритми обробки ties та поправки.

17. Ранговий розмір ефекту Mann–Whitney

Одним із зручних показників є rank-biserial correlation.

Для абсолютної величини:

$$\boxed{|r_{rb}| = 1 - \frac{2U}{n_A n_B}}.$$

Для прикладу:

$$|r_{rb}| = 1 - \frac{2 \cdot 13}{10 \cdot 10} = 0.74.$$

Це великий ранговий ефект.

18. Напрям effect size

Знак залежить від того, яку групу прийнято за першу та як визначена формула. Тому в звіті потрібно чітко вказувати напрям:

А має вищі оцінки, ніж В.

а не лише подавати абсолютний коефіцієнт.

19. Що бажано звітувати для Mann–Whitney

- n_A, n_B;
- медіани та, бажано, квартилі;
- U;
- p-value;
- effect size;
- напрям відмінності;
- предметний висновок.

20. Wilcoxon signed-rank: призначення

Критерій Wilcoxon signed-rank використовується для **парних** або залежних даних.

Приклади:

- ті самі дегустатори оцінюють продукт до і після зміни рецептури;
- той самий зразок вимірюють до і після обробки;
- та сама технологічна одиниця досліджується у двох умовах.

21. Що аналізує Wilcoxon

Спочатку для кожної пари обчислюють різницю:

$$\boxed{d_i = y_i - x_i}.$$

Далі аналізуються:

- знак різниці;
- абсолютна величина різниці;
- ранг абсолютної різниці.

22. Гіпотези Wilcoxon

Для симетричного розподілу різниць критерій можна інтерпретувати як перевірку:

$$H_0\text{:центральний зсув дорівнює } 0.$$

Для двосторонньої альтернативи:

$$H_1\text{:центральний зсув відрізняється від } 0.$$

23. Важливе припущення Wilcoxon

Для інтерпретації саме як тесту зсуву бажано, щоб розподіл парних різниць був приблизно симетричним.

Якщо різниці сильно асиметричні, інтерпретація Wilcoxon потребує обережності.

24. Алгоритм Wilcoxon signed-rank

1. Обчислити $d_i = y_i - x_i$.
2. Вилучити пари з $d_i = 0$.
3. Обчислити $|d_i|$.
4. Впорядкувати абсолютні різниці.
5. Присвоїти ранги.
6. Для ties використати середні ранги.
7. Повернути знаки початкових d_i.
8. Обчислити суму позитивних рангів T_+.
9. Обчислити суму негативних рангів T_-.
10. Визначити статистику:

$$\boxed{T = \min(T_+, T_-)}.$$

25. Нульові різниці

Якщо:

$$d_i = 0,$$

така пара не дає інформації про напрям зміни і вилучається **лише з ранжування**.

Ефективний обсяг:

$$n_{\text{eff}} = n - n_0,$$

де n_0 — кількість нульових різниць.

26. Демонстраційний приклад Wilcoxon

12 дегустаторів оцінюють продукт до і після зміни рецептури:

Дегустатор	До	Після
1	6	7
2	7	9

Дегустатор	До	Після
3	5	6
4	6	6
5	7	8
6	6	5
7	5	7
8	7	8
9	6	6
10	5	6
11	7	8
12	6	8

26.1. Різниці

$$d = (+1, +2, +1, 0, +1, -1, +2, +1, 0, +1, +1, +2).$$

Дві нульові різниці вилучаються з ранжування:

$$n_{\text{eff}} = 10.$$

26.2. Абсолютні різниці та ранги

Маємо 7 значень $|d| = 1$ та 3 значення $|d| = 2$.

Для $|d| = 1$ займаються ранги 1–7:

$$R(|d| = 1) = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7}{7} = 4.$$

Для $|d| = 2$ займаються ранги 8–10:

$$R(|d| = 2) = \frac{8 + 9 + 10}{3} = 9.$$

26.3. Суми знакових рангів

Позитивні ранги:

$$T_+ = 51.$$

Негативні:

$$T_- = 4.$$

Перевірка:

$$T_+ + T_- = \frac{10 \cdot 11}{2} = 55.$$

Статистика:

$$\boxed{T = 4}.$$

27. p-value Wilcoxon

У спеціалізованому ПЗ для цього набору:

$$p \approx 0.015$$

залежно від exact/asymptotic алгоритму та поправок.

При $\alpha = 0.05$:

$$p < 0.05.$$

Отже, є статистичні підстави вважати, що оцінки після зміни рецептури систематично відрізняються від оцінок до зміни.

28. Rank-biserial effect size для Wilcoxon

Один із варіантів:

$$\boxed{r_{rb} = \frac{T_+ - T_-}{T_+ + T_-}}.$$

Для прикладу:

$$r_{rb} = \frac{51 - 4}{55} \approx 0.855.$$

Це великий позитивний ефект у напрямі покращення після зміни рецептури.

29. Медіани у Wilcoxon

Для опису результату корисно подати:

$$Me_{\text{до}}, \qquad Me_{\text{після}}, \qquad Me_d.$$

У демонстраційному прикладі:

$$Me_{\text{до}} = 6, \qquad Me_{\text{після}} = 7, \qquad Me_d = 1.$$

Це робить статистичний результат предметно зрозумілішим.

30. Mann–Whitney і Wilcoxon: принципова різниця

Ознака	Mann–Whitney	Wilcoxon signed-rank
Структура даних	Дві незалежні групи	Парні спостереження
Основний об'єкт	Ранги об'єднаних спостережень	Ранги абсолютних парних різниць
Нульові різниці	Не застосовується	Вилучаються з ранжування
Статистика	U T	
Типовий приклад	Дві незалежні дегустаційні групи	Ті самі дегустатори до/після

31. Не плутати Wilcoxon signed-rank із Wilcoxon rank-sum

У літературі назва Wilcoxon може використовуватися для двох різних процедур:

- **Wilcoxon rank-sum** — фактично еквівалент Mann–Whitney для незалежних вибірок;
- **Wilcoxon signed-rank** — для парних даних.

У ПР5 під словом «Wilcoxon» використовується саме **signed-rank test**.

32. Органолептичні дані як порядкові

Дегустаційні бали часто мають порядкову природу. Наприклад, 8 балів означає вищу оцінку, ніж 7, але не обов'язково різниця:

$$8 - 7$$

має абсолютно той самий психологічний зміст, що й:

$$5 - 4.$$

Саме тому рангові методи є природним інструментом для таких даних.

33. Чи можна все ж обчислювати середнє для дегустаційних балів?

У прикладних роботах середні бали часто наводяться для опису, особливо коли шкала має багато рівнів і поводиться приблизно як інтервальна.

Але для формальної перевірки при сумнівній інтервальності шкали ранговий критерій часто є методично безпечнішим.

34. Важливість незалежності дегустаторів

Для Mann–Whitney оцінки двох груп повинні бути незалежними.

Якщо ті самі дегустатори оцінюють обидва продукти, дані вже парні і Mann–Whitney втрачає інформацію про відповідність пар.

У такій ситуації потрібний Wilcoxon signed-rank.

35. Вплив ties на статистику

Велика кількість однакових балів є типовою для органолептики.

Спеціалізоване програмне забезпечення зазвичай автоматично враховує ties у:

- розрахунку середніх рангів;
- дисперсії статистики;
- асимптотичному p-value.

36. Continuity correction

При нормальному наближенні дискретної статистики інколи використовується поправка неперервності.

Через це, наприклад, Excel-надбудова і Jamovi можуть дати p-value, які відрізняються в останніх знаках.

У звіті важливо вказувати програму/алгоритм, якщо точний p-value суттєво залежить від налаштувань.

37. Коли непараметричний тест не є автоматично кращим

Не слід переходити до Mann–Whitney лише через незначне відхилення від нормальності.

Параметричний t-test часто:

- має вищу потужність;
- дає простішу інтерпретацію середніх;
- є робастним до помірних відхилень за відсутності сильних викидів.

Метод вибирають за сукупністю:

$$\boxed{\text{дизайн} + \text{шкала} + \text{форма розподілу} + \text{викиди} + \text{дослідницьке питання}}.$$

38. Коли Mann–Whitney є природним вибором

- дві незалежні групи;
- порядкова шкала;
- виражена ненормальність;
- сильні викиди;
- малий обсяг вибірки;
- аналіз центрального положення в ранговій шкалі.

39. Коли Wilcoxon signed-rank є природним вибором

- парні дані;
- порядкова або кількісна шкала;
- немає підстав використовувати paired t-test;
- різниці допускають змістовне впорядкування;
- для інтерпретації зсуву — різниці приблизно симетричні.

40. Signed-rank vs sign test

Wilcoxon signed-rank використовує і знак, і величину рангу різниці.

Sign test використовує лише знак:

$$d_i > 0 \quad \text{або} \quad d_i < 0.$$

Тому sign test має слабші припущення, але зазвичай меншу потужність.

41. Effect size і p-value

Малий p-value не показує величину ефекту.

Наприклад:

$$p = 0.003$$

може відповідати як невеликому ефекту при великому n, так і великому ефекту при малому n.

Тому:

$$\boxed{p\text{-value} + \text{effect size} + \text{описові статистики}}.$$

42. Статистичний висновок для Mann–Whitney

Приклад структури:

Медіана оцінок продукту A становила 7 балів, продукту B — 6 балів. Критерій Mann–Whitney показав статистично значущу відмінність між незалежними групами: $U = 13$, $p \approx 0.004$. Rank-biserial effect size становив приблизно $|r_{rb}| = 0.74$, що свідчить про великий ефект.

43. Технологічний висновок Mann–Whitney

Продукт A отримує систематично вищі дегустаційні оцінки, ніж продукт B. Різниця є не лише статистично значущою, але й великою за ранговим effect size.

Остаточний технологічний висновок має враховувати дизайн дегустації, склад панелі та інші показники якості.

44. Статистичний висновок для Wilcoxon

Медіана оцінки до зміни рецептури становила 6 балів, після — 7 балів. Wilcoxon signed-rank test виявив статистично значущу зміну: $T = 4$, $p \approx 0.015$. Rank-biserial effect size $r_{rb} \approx 0.855$ свідчить про великий позитивний ефект.

45. Технологічний висновок Wilcoxon

Після зміни рецептури оцінки тих самих дегустаторів переважно підвищилися, що підтримує висновок про покращення органолептичного сприйняття продукту.

46. Типові помилки

1. Використовувати Mann–Whitney для парних даних.
2. Використовувати Wilcoxon signed-rank для незалежних груп.
3. Не враховувати ties.
4. Не вилучати нульові різниці з ранжування Wilcoxon.
5. Називати Mann–Whitney універсальним «тестом медіан».
6. Звітувати лише p-value без медіан і effect size.
7. Плутати Wilcoxon signed-rank і rank-sum.
8. Переходити до непараметричного тесту автоматично при $p = 0.049$ у Shapiro–Wilk.
9. Ігнорувати симетрію різниць при інтерпретації Wilcoxon як тесту зсуву.
10. Розглядати ранговий effect size без напряму ефекту.

47. Алгоритм вибору для двох груп

две групи → незалежні чи парні? → шкала/розподіл → параметричний чи ранговий метод.

Структура	Параметричний варіант	Непараметричний варіант
Дві незалежні групи	Student / Welch t-test	Mann–Whitney
Парні спостереження	Paired t-test	Wilcoxon signed-rank

48. Зв'язок із практичною роботою №5

ПР5 складається з двох взаємопов'язаних частин:

- **Частина A.** Дві незалежні дегустаційні групи → Mann–Whitney.
- **Частина B.** Ті самі дегустатори до/після → Wilcoxon signed-rank.

У кожній частині студент:

1. описує дані;
2. визначає структуру вибірок;

3. формує ранги;
4. враховує ties;
5. обчислює статистику;
6. визначає p-value;
7. обчислює effect size;
8. формулює статистичний та технологічний висновки.

49. Підсумок лекції

$$\text{дизайн} \rightarrow \text{ранги} \rightarrow \text{ties} \rightarrow \begin{cases}\text{Mann–Whitney,} & \text{незалежні групи,}\\ \text{Wilcoxon signed-rank,} & \text{парні дані}\end{cases} \rightarrow p \rightarrow \text{rank effect size} \rightarrow \text{технологічний}$$

Непараметричні критерії дозволяють аналізувати порядкові або ненормальні дані без жорсткої прив'язки до нормального розподілу. Їх сила — у роботі з порядком спостережень, але коректність висновку все одно залежить від правильного дизайну, обробки ties і предметної інтерпретації.

50. Контрольні питання

1. Чим непараметричні критерії відрізняються від параметричних?
2. Що таке ранг?
3. Як присвоюються ранги однаковим значенням?
4. Чому ties особливо характерні для органолептичних даних?
5. Для яких вибірок використовується Mann–Whitney?
6. Які гіпотези перевіряє Mann–Whitney?
7. Чому Mann–Whitney не завжди коректно називати тестом медіан?
8. Як обчислюються U_A та U_B?
9. Яка контрольна рівність виконується для U_A та U_B?
10. Що означає мале значення U?
11. Чим exact p-value відрізняється від asymptotic?
12. Як обчислюється rank-biserial effect size для Mann–Whitney?
13. Для яких даних використовується Wilcoxon signed-rank?
14. Яку величину d_i обчислюють у Wilcoxon?
15. Що робити з нульовими різницями?
16. Як ранжуються абсолютні різниці?
17. Що таке T_+ та T_-?
18. Як визначається статистика T?
19. Яке припущення щодо форми розподілу різниць бажане для інтерпретації Wilcoxon як тесту зсуву?
20. Як обчислюється rank-biserial effect size для Wilcoxon?
21. Чим Wilcoxon signed-rank відрізняється від Wilcoxon rank-sum?

22. Чому одна й та сама група дегустаторів до/після не повинна аналізуватися Mann–Whitney?
23. Чому p-value потрібно доповнювати effect size?
24. Коли параметричний t-test може бути кращим за непараметричний?
25. Який непараметричний аналог незалежного t-test?
26. Який непараметричний аналог paired t-test?

Лекція 11

Однофакторний дисперсійний аналіз (ANOVA) у дослідженнях харчових технологій

1. Мета лекції та результати навчання

Мета лекції — сформувати розуміння логіки однофакторного дисперсійного аналізу як методу порівняння трьох і більше технологічних режимів та навчити інтерпретувати ANOVA не лише як програмну таблицю, а як декомпозицію загальної мінливості експериментальних даних.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати, чому для порівняння трьох і більше середніх недоцільно виконувати серію окремих t-тестів;
- формулювати гіпотези однофакторного ANOVA;
- розрізняти міжгрупову, внутрішньогрупову та загальну мінливість;
- обчислювати SS_B, SS_W та SS_T;
- визначати ступені свободи і середні квадрати MS_B, MS_W;
- обчислювати F-статистику та приймати статистичне рішення;
- пояснювати припущення ANOVA;
- розуміти роль тестів Shapiro–Wilk і Levene;
- інтерпретувати post-hoc Tukey HSD;
- обчислювати η^2 як розмір ефекту;
- формулювати статистичний і технологічний висновки для порівняння декількох режимів.

2. План лекції

1. Проблема множинних порівнянь.
2. Ідея дисперсійного аналізу.
3. Фактор, рівень фактора та відгук.
4. Гіпотези однофакторного ANOVA.
5. Декомпозиція загальної мінливості.
6. Міжгрупова та внутрішньогрупова суми квадратів.
7. Ступені свободи та середні квадрати.
8. F-статистика.
9. Припущення ANOVA.
10. Post-hoc аналіз Tukey HSD.
11. Розмір ефекту η^2.
12. Короткий вступ до планування експерименту.

3. Чому не можна просто виконати багато t-тестів

Припустимо, досліджуються чотири температурні режими:

$$70°C, \quad 75°C, \quad 80°C, \quad 85°C.$$

Кількість усіх попарних порівнянь:

$$C_4^2 = \frac{4!}{2!\,2!} = 6.$$

Якщо кожний t-test виконується при $\alpha = 0.05$, ймовірність хоча б одного хибнопозитивного висновку в серії тестів стає більшою за 0.05.

3.1. Family-wise error rate

Якщо для спрощення вважати тести незалежними, ймовірність не зробити жодної помилки I роду в m тестах:

$$(1-\alpha)^m.$$

Тоді ймовірність хоча б одного хибнопозитивного результату:

$$\boxed{FWER = 1-(1-\alpha)^m}.$$

Для $\alpha = 0.05$ і $m = 6$:

$$FWER \approx 1 - 0.95^6 \approx 0.265.$$

Тобто приблизно 26.5% у цьому спрощеному розрахунку.

ANOVA вирішує іншу задачу: спочатку перевіряє одну глобальну гіпотезу про рівність усіх середніх, контролюючи помилку I роду на заданому рівні.

4. Що таке ANOVA

ANOVA — Analysis of Variance, тобто дисперсійний аналіз.

На перший погляд парадоксально: метод використовується для порівняння середніх, але називається аналізом дисперсії.

Причина: ANOVA порівнює два джерела мінливості:

- мінливість **між групами**;
- мінливість **всередині груп**.

якщо міжгрупова мінливість ≫ внутрішньогрупова, то середні, імовірно, відрізняються.

5. Фактор, рівень і відгук

Фактор — змінна, вплив якої досліджується.

Рівень фактора — конкретне значення або категорія фактора.

Відгук — кількісна результативна змінна.

Фактор — температура оброблення.

Рівні — 70, 75, 80, 85 °C.

Відгук — вихід готового продукту, %.

6. Структура однофакторного ANOVA

Нехай маємо k груп. У групі i є n_i спостережень:

$$x_{i1}, x_{i2}, \dots, x_{in_i}.$$

Загальний обсяг:

$$\boxed{N = \sum_{i=1}^{k} n_i}.$$

7. Гіпотези ANOVA

Нульова гіпотеза:

$$\boxed{H_0: \mu_1 = \mu_2 = \cdots = \mu_k}.$$

Альтернативна:

H_1:принаймні одне середнє відрізняється.

Значущий ANOVA не говорить, які саме групи відрізняються. Він лише відхиляє твердження «усі середні однакові».

8. Групові середні

Для кожної групи:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}.$$

9. Загальне середнє

Grand mean:

$$\boxed{\bar{x}_{..} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}}.$$

Якщо обсяги груп однакові, це також середнє групових середніх:

$$\bar{x}_{..} = \frac{1}{k} \sum_{i=1}^{k} \bar{x}_i.$$

лише за рівних n_i.

10. Загальна сума квадратів

Загальна мінливість відносно grand mean:

$$\boxed{SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{..}\right)^2}.$$

Це Total Sum of Squares.

11. Міжгрупова сума квадратів

Between-group sum of squares:

$$\boxed{SS_B = \sum_{i=1}^{k} n_i\,(\bar{x}_i - \bar{x}_{..})^2}.$$

Вона характеризує відмінності між груповими середніми.

12. Внутрішньогрупова сума квадратів

Within-group sum of squares:

$$\boxed{SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i}\right)^2}.$$

Це мінливість спостережень навколо власних групових середніх.

13. Декомпозиція варіації

Основна тотожність ANOVA:

$$\boxed{SS_T = SS_B + SS_W}.$$

Інтерпретація. Загальна мінливість даних складається з частини, пов'язаної з відмінностями між групами, і частини, що лишається всередині груп.

14. Ступені свободи

Для between:

$$\boxed{df_B = k - 1}.$$

Для within:

$$\boxed{df_W = N - k}.$$

Для total:

$$\boxed{df_T = N - 1}.$$

Перевірка:

$$df_T = df_B + df_W.$$

15. Середні квадрати

Міжгруповий:

$$\boxed{MS_B = \frac{SS_B}{df_B}}.$$

Внутрішньогруповий:

$$\boxed{MS_W = \frac{SS_W}{df_W}}.$$

16. F-статистика

$$\boxed{F = \frac{MS_B}{MS_W}}.$$

Якщо H_0 істинна, обидва середні квадрати оцінюють приблизно одну й ту саму генеральну дисперсію, тому:

$$F \approx 1.$$

Якщо групові середні істотно відрізняються:

$$MS_B \gg MS_W$$

і F стає великим.

17. F-розподіл в ANOVA

За H_0 статистика має F-розподіл із:

$$df_1 = k - 1, \qquad df_2 = N - k.$$

Статистичне рішення:

$$p < \alpha \Rightarrow H_0 \text{ відхиляємо.}$$

18. Демонстраційний приклад

Досліджується вплив температури на вихід продукту. По 6 спостережень у кожній групі.

70 °C	75 °C	80 °C	85 °C
91.8	93.0	94.2	92.7
92.2	93.4	94.5	93.0
91.5	92.8	94.1	92.9
92.0	93.1	94.4	93.1
91.7	93.3	94.3	92.8
92.1	92.9	94.6	93.0

19. Групові середні прикладу

$$\bar{x}_{70} = 91.8833,$$

$$\bar{x}_{75} = 93.0833,$$

$$\bar{x}_{80} = 94.3500,$$

$$\bar{x}_{85} = 92.9167.$$

20. Grand mean

$$\bar{x}_{\cdot\cdot} = 93.0583.$$

21. Суми квадратів прикладу

$$SS_B = 18.4183.$$

$$SS_W = 0.9000.$$

$$SS_T = 19.3183.$$

Перевірка:

$$18.4183 + 0.9000 = 19.3183.$$

22. Ступені свободи прикладу

Маємо:

$$k = 4, \qquad N = 24.$$

$$df_B = 4 - 1 = 3.$$

$$df_W = 24 - 4 = 20.$$

$$df_T = 24 - 1 = 23.$$

23. Mean Squares прикладу

$$MS_B = \frac{18.4183}{3} = 6.1394.$$

$$MS_W = \frac{0.9000}{20} = 0.0450.$$

24. F-статистика прикладу

$$F = \frac{6.1394}{0.0450} \approx 136.43.$$

При $df_1 = 3$, $df_2 = 20$:

$$F_{\text{crit}} \approx 3.098.$$

Оскільки:

$$136.43 \gg 3.098,$$

а:

$$p < 0.0001,$$

гіпотеза про рівність усіх середніх відхиляється.

25. ANOVA-таблиця

Джерело	SS	df	MS	F
Між групами	18.4183	3	6.1394	136.43
Всередині груп	0.9000	20	0.0450	—
Загальна	19.3183	23	—	—

26. Що саме довів ANOVA

ANOVA довів лише:

$$\boxed{\text{не всі } \mu_i \text{ однакові}}.$$

Він не визначив, які саме температури відрізняються.

27. Post-hoc аналіз

Після значущого глобального ANOVA виконують post-hoc порівняння, які контролюють множинність.

У курсі використовується Tukey HSD.

28. Tukey HSD

Tukey Honest Significant Difference порівнює всі пари середніх із контролем family-wise error rate.

Для рівних обсягів груп концептуально:

$$HSD = q_{\alpha;k,df_W}\sqrt{\frac{MS_W}{n}}.$$

У нашому курсі ручний розрахунок Tukey не є обов'язковим; студент інтерпретує програмну таблицю.

29. Результат Tukey для демонстраційного набору

Пара	Висновок
70–75	значуща різниця
70–80	значуща різниця
70–85	значуща різниця
75–80	значуща різниця
75–85	різниця незначуща
80–85	значуща різниця

30. Розмір ефекту η^2

Статистична значущість не показує, наскільки велика роль фактора.

Для однофакторного ANOVA:

$$\boxed{\eta^2 = \frac{SS_B}{SS_T}}.$$

Для прикладу:

$$\eta^2 = \frac{18.4183}{19.3183} \approx 0.953.$$

Отже, приблизно 95.3% загальної варіативності в цьому навчальному наборі пов'язано з відмінностями між температурними режимами.

31. Орієнтири для η^2

У різних дисциплінах орієнтири відрізняються. Часто умовно використовують:

η^2	Умовна оцінка
≈0.01	малий ефект
≈0.06	середній
≈0.14 і більше	великий

Але технологічний зміст важливіший за універсальні пороги.

32. Припущення однофакторного ANOVA

1. Незалежність спостережень.
2. Кількісна результативна змінна.
3. Приблизна нормальність у кожній групі або нормальність залишків.
4. Однорідність дисперсій.

33. Незалежність

Це насамперед властивість дизайну. Якщо один і той самий зразок вимірюється в кількох умовах, звичайний independent one-way ANOVA не є коректним.

34. Нормальність

У малих групах доцільно аналізувати:

- Q–Q plots;
- Shapiro–Wilk;
- викиди;
- симетрію.

35. Однорідність дисперсій

Перевіряється:

- графічно;
- Levene test;
- іноді Brown–Forsythe.

У ПР6 використовується Levene.

36. Що робити при порушенні однорідності

Можливі варіанти:

- Welch ANOVA;

- трансформація даних;
- робастні методи;
- непараметричний Kruskal–Wallis.

У базовій ПР6 ці альтернативи згадуються концептуально.

37. Kruskal–Wallis як непараметрична альтернатива

Для 3+ незалежних груп при порядкових або явно ненормальних даних може використовуватися Kruskal–Wallis test.

Він працює з рангами й є узагальненням логіки Mann–Whitney на кілька груп.

38. ANOVA і причинність

Якщо фактор був **контрольовано заданий і рандомізований** у добре спланованому експерименті, ANOVA може підтримувати причинну інтерпретацію.

Якщо ж аналізуються лише спостережувані групи без контролю змішувальних факторів, причинний висновок обмежений.

39. Вступ до планування експерименту

У технологічному дослідженні важливі поняття:

- фактор;
- рівень;
- відгук;
- повторність;
- рандомізація;
- блокування;
- взаємодія факторів.

40. Повторність

Повтори дозволяють оцінити внутрішньогрупову дисперсію.

Без повторів відокремити випадкову мінливість від ефекту фактора значно складніше.

41. Рандомізація

Порядок виконання дослідів бажано рандомізувати, щоб систематичні часові фактори не змішувалися з рівнями фактора.

42. Блокування

Якщо відомо важливе додаткове джерело варіації, наприклад партія сировини або день проведення експерименту, його можна врахувати через блоки.

43. Взаємодія факторів

Якщо одночасно вивчаються, наприклад, температура і вологість, ефект температури може залежати від рівня вологості.

Це називають interaction.

Однофакторний ANOVA взаємодій не аналізує; для цього потрібний двофакторний або багатофакторний дизайн.

44. Чому у ПР6 використовується однофакторний ANOVA

Методично важливо спочатку чітко засвоїти базову декомпозицію:

$$SS_T = SS_B + SS_W.$$

Лише після цього логічно переходити до більш складних моделей.

45. Excel для ANOVA

У Microsoft Excel:

- дані зручно розташувати по групах у стовпцях;
- Data → Data Analysis → ANOVA: Single Factor;
- вихід містить Summary та ANOVA table.

46. Що потрібно вміти читати у виході Excel

- Count;
- Sum;
- Average;
- Variance;
- SS;
- df;
- MS;
- F;
- P-value;
- F crit.

47. Чому ручний розрахунок потрібний попри Excel

Ручний розрахунок дозволяє побачити зміст ANOVA:

загальна мінливість → між групами + всередині груп.

Без цього ANOVA-table легко перетворюється на набір незрозумілих чисел.

48. Типові помилки

1. Порівнювати 4 групи шістьма t-тестами без корекції.
2. Інтерпретувати значущий ANOVA як доказ того, що всі пари відрізняються.

3. Не виконувати post-hoc аналіз.
4. Плутати SS_B і SS_W.
5. Неправильно обчислювати grand mean при нерівних n_i.
6. Не перевіряти $SS_T = SS_B + SS_W$.
7. Плутати SS і MS.
8. Звітувати лише p-value без η^2.
9. Ігнорувати Levene.
10. Вважати ANOVA автоматичним доказом причинності.

49. Зв'язок із практичною роботою №6

У ПР6 ANOVA є головною ручною частиною. Студент:

1. обчислює групові середні;
2. обчислює grand mean;
3. визначає SS_B, SS_W, SS_T;
4. перевіряє декомпозицію;
5. обчислює df;
6. визначає MS_B, MS_W;
7. обчислює F;
8. інтерпретує p-value;
9. обчислює η^2;
10. програмно аналізує Tukey HSD.

50. Підсумок лекції

3+ групи → $H_0: \mu_1 = \cdots = \mu_k \rightarrow SS_T = SS_B + SS_W \rightarrow MS_B, MS_W \rightarrow F \rightarrow p \rightarrow$ Tukey $\rightarrow \eta^2 \rightarrow$ технологічний в

ANOVA — це не «ще один тест середніх», а метод порівняння структури міжгрупової та внутрішньогрупової мінливості. Він дає глобальне рішення, яке після значущого результату доповнюється post-hoc аналізом і оцінкою величини ефекту.

51. Контрольні питання

1. Чому серія t-тестів для 4 груп збільшує ризик помилки I роду?
2. Що означає FWER?
3. Які гіпотези перевіряє однофакторний ANOVA?
4. Що таке фактор, рівень фактора та відгук?
5. Що характеризує SS_B?
6. Що характеризує SS_W?
7. Запишіть тотожність декомпозиції сум квадратів.
8. Як визначається df_B?
9. Як визначається df_W?

10. Що таке MS_B і MS_W?
11. Запишіть формулу F-статистики.
12. Чому при істинному H_0 очікуємо F близьке до 1?
13. Що означає дуже велике F?
14. Які припущення має ANOVA?
15. Для чого використовується Levene?
16. Що робити при неоднорідних дисперсіях?
17. Чому значущий ANOVA не показує, які пари відрізняються?
18. Для чого потрібен Tukey HSD?
19. Як обчислюється η^2?
20. Що означає велике η^2?
21. Що таке повторність у дизайні експерименту?
22. Навіщо потрібна рандомізація?
23. Що таке блокування?
24. Що називають interaction?
25. Чому ANOVA не завжди дозволяє робити причинний висновок?

Лекція 12

Кореляційний аналіз технологічних даних

1. Мета лекції та результати навчання

Мета лекції — сформувати розуміння кореляційного аналізу як інструменту кількісного оцінювання зв'язку між технологічними змінними та навчити критично інтерпретувати коефіцієнти кореляції разом із графічною структурою даних.

Після опрацювання лекції здобувач повинен уміти:

- пояснювати, що означає статистичний зв'язок між двома кількісними змінними;
- будувати та інтерпретувати scatterplot;
- розрізняти позитивний, негативний, слабкий, сильний та нелінійний зв'язок;
- обчислювати коефіцієнт кореляції Пірсона;
- пояснювати припущення для Pearson correlation;
- перевіряти статистичну значущість коефіцієнта r;
- розуміти роль довірчого інтервалу для кореляції;
- пояснювати вплив викидів і обмеження діапазону;
- розуміти логіку рангової кореляції Spearman;
- відрізняти лінійну кореляцію від причинного зв'язку;
- порівнювати висновки ANOVA та кореляційного аналізу для одного технологічного експерименту.

2. План лекції

1. Поняття статистичного зв'язку.
2. Scatterplot як перший етап кореляційного аналізу.
3. Коефіцієнт кореляції Пірсона.
4. Властивості r.
5. Перевірка значущості кореляції.
6. Довірчий інтервал для ρ.
7. Викиди та інші джерела викривлення.
8. Кореляція Spearman.
9. Кореляційна матриця — концептуально.
10. Кореляція та причинність.
11. Зв'язок кореляції з ANOVA і регресією.

3. Від групового порівняння до зв'язку між змінними

У ANOVA температура розглядалася як **категоріальний фактор** із чотирма рівнями:

$$70,\ 75,\ 80,\ 85°C.$$

Кореляційний аналіз ставить інше питання: якщо температуру розглядати як кількісну змінну X, чи існує систематична залежність між її значенням та виходом продукту Y?

Один і той самий набір даних може відповідати різним статистичним питанням. ANOVA порівнює групові середні; кореляція оцінює напрям і силу зв'язку між двома числовими змінними.

4. Парні спостереження

Кореляційний аналіз працює з парами:

$$(x_1, y_1),\ (x_2, y_2), \dots, (x_n, y_n).$$

Кожна пара повинна належати одному й тому самому об'єкту або досліду.

Наприклад:

Температура X, °C	Вихід Y, %
70	91.8
70	92.2
75	93.0
80	94.2
85	92.7

5. Scatterplot — обов'язковий перший крок

Діаграма розсіювання показує кожну пару (x_i, y_i) як точку на площині.

Вона дозволяє побачити:

- напрям зв'язку;
- приблизну форму;
- лінійність;
- кластери;
- викиди;
- неоднакове розсіювання;
- можливі нелінійні залежності.

Не слід починати аналіз із функції CORREL. Однакове або близьке значення r може відповідати дуже різним геометричним структурам даних.

6. Основні типи зв'язку

6.1. Позитивний

Зі збільшенням X значення Y мають тенденцію збільшуватися.

6.2. Негативний

Зі збільшенням X значення Y мають тенденцію зменшуватися.

6.3. Відсутність вираженого лінійного зв'язку

Зміни X не супроводжуються систематичним лінійним збільшенням або зменшенням Y.

6.4. Нелінійний зв'язок

Залежність може бути сильною, але не описуватися прямою. Наприклад, вихід продукту зростає до оптимальної температури, а потім зменшується.

$$\boxed{\text{сильний зв'язок} \neq \text{сильна лінійна кореляція}}.$$

7. Коефіцієнт кореляції Пірсона

Вибірковий коефіцієнт:

$$\boxed{r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}}.$$

Він є стандартизованою мірою спільної лінійної мінливості двох змінних.

8. Коваріація як основа кореляції

Вибіркова коваріація:

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y}).$$

Тоді:

$$\boxed{r = \frac{s_{xy}}{s_x s_y}}.$$

Стандартизація через s_x і s_y робить коефіцієнт безрозмірним.

9. Межі коефіцієнта

$$\boxed{-1 \leq r \leq 1}.$$

Значення	Зміст
$r = 1$ ідеальний позитивний	лінійний зв'язок
$r = -1$ ідеальний негативни	й лінійний зв'язок
$r = 0$ відсутність лінійної	кореляції

10. Властивості Pearson r

- безрозмірний;
- симетричний:

$$r_{XY} = r_{YX};$$

- не змінюється при переході від одиниць вимірювання до іншої лінійної шкали;
- чутливий до викидів;
- описує саме лінійну складову зв'язку.

11. Орієнтовна сила кореляції

Часто використовують умовні орієнтири:

$\lvert r\rvert$ Умовна х	арактеристика
0–0.3	слабка
0.3–0.5	помірна
0.5–0.7	помітна / середньо сильна
0.7–0.9	сильна
0.9–1.0	дуже сильна

Це лише орієнтири. У технологічному дослідженні значення $r = 0.4$ може бути важливим, а $r = 0.9$ — артефактом вузького або змішаного набору даних.

12. Коефіцієнт детермінації для простої лінійної залежності

У парній лінійній регресії:

$$\boxed{R^2 = r^2}.$$

Наприклад, якщо:

$$r = 0.80,$$

то:

$$R^2 = 0.64.$$

У межах простої лінійної моделі приблизно 64% варіації Y асоційовано з лінійною зміною X.

13. Припущення для Pearson correlation

Для коректної класичної інференції щодо ρ бажані:

- парні кількісні спостереження;
- незалежність пар;
- приблизна лінійність;
- відсутність грубих впливових викидів;
- для класичної перевірки — приблизна двовимірна нормальність або принаймні нормальність, достатня для моделі.

14. Популяційний і вибірковий коефіцієнти

Генеральний коефіцієнт:

$$\rho.$$

Вибіркова оцінка:

$$r.$$

Як і будь-яка вибіркова статистика, r змінюється від вибірки до вибірки.

15. Перевірка значущості Pearson r

Типова нульова гіпотеза:

$$\boxed{H_0: \rho = 0}.$$

Для двосторонньої альтернативи:

$$H_1: \rho \neq 0.$$

16. t-статистика для кореляції

$$\boxed{t = r\sqrt{\frac{n-2}{1-r^2}}}.$$

Ступені свободи:

$$\boxed{df = n-2}.$$

17. Приклад перевірки значущості

Для даних ПР6:

$$n = 24, \qquad r \approx 0.544.$$

Тоді:

$$t = 0.544\sqrt{\frac{22}{1-0.544^2}} \approx 3.05.$$

Для $df = 22$:

$$p \approx 0.006.$$

Отже, лінійна кореляція статистично відрізняється від нуля.

18. Чому значуща кореляція може бути лише помірною

Статистична значущість залежить не лише від r, а й від n.

При достатньому обсязі навіть помірний коефіцієнт може бути значущим.

$$\boxed{\text{значущість кореляції} \neq \text{сила кореляції}}.$$

19. Довірчий інтервал для кореляції

Розподіл r не є симетричним, особливо поблизу ± 1. Тому для інтервального оцінювання часто використовують Fisher z-transformation:

$$\boxed{z_F = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)}.$$

Стандартна похибка:

$$SE_z = \frac{1}{\sqrt{n-3}}.$$

Після побудови CI у z-шкалі виконують зворотне перетворення.

У базовій ПР6 ручний CI для кореляції не є обов'язковим, але студент повинен розуміти, що r також має вибіркову невизначеність.

20. Викиди і Pearson r

Одна крайня точка може різко:

- збільшити r;
- зменшити r;
- змінити знак кореляції.

Тому scatterplot є обов'язковим.

21. Впливова точка

Впливова точка — спостереження, вилучення якого суттєво змінює статистичний результат.

Вона не обов'язково є помилкою. Це може бути реальний режим процесу, що принципово відрізняється.

22. Обмеження діапазону

Якщо аналізувати лише вузьку область X, кореляція може стати слабшою.

Наприклад, у повному діапазоні 20–80 °C зв'язок може бути сильним, але в діапазоні 70–75 °C майже не проявлятися.

Коефіцієнт кореляції залежить від діапазону даних. Тому не можна порівнювати r між дослідженнями, не враховуючи область спостережень.

23. Змішування груп

Кореляція може виникнути через змішування підгруп.

Наприклад, дві виробничі лінії мають різні середні значення температури й виходу. Якщо об'єднати їх без урахування лінії, отримаємо кореляцію, яка частково відображає різницю між лініями, а не внутрішньолінійний зв'язок.

24. Simpson's paradox — концептуально

У крайніх випадках напрям зв'язку в об'єднаних даних може відрізнятися від напряму всередині кожної групи.

Це один із аргументів на користь обов'язкового врахування структури експерименту та групувальних факторів.

25. Нелінійний зв'язок і нульова кореляція

При U-подібній залежності Pearson r може бути близьким до нуля:

$$r \approx 0,$$

навіть якщо Y сильно залежить від X.

$$\boxed{r = 0 \neq \text{«зв'язку немає»}}.$$

Коректніше: немає вираженої **лінійної** кореляції.

26. Кореляція Spearman

Spearman rank correlation використовує ранги:

$$\boxed{\rho_s = Corr(R_X, R_Y)}.$$

Якщо ties відсутні, класична формула:

$$\rho_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

де d_i — різниця рангів.

За ties зручніше обчислювати Pearson correlation між ранговими змінними.

27. Що оцінює Spearman

Spearman оцінює силу **монотонного** зв'язку.

Монотонний означає, що при збільшенні X Y переважно рухається в одному напрямі, але не обов'язково по прямій.

28. Pearson vs Spearman

Ознака	Pearson	Spearman
Що оцінює	Лінійний зв'язок	Монотонний ранговий зв'язок
Дані	Кількісні	Кількісні або порядкові
Викиди	Чутливий	Зазвичай стійкіший
Нелінійна монотонність	Може недооцінювати	Може відображати краще
Розрахунок	За вихідними значеннями	За рангами

29. Приклад ПР6: Pearson і Spearman

Для 24 пар «температура — вихід» отримано:

$$r_P \approx 0.544, \qquad p \approx 0.006.$$

$$\rho_S \approx 0.501, \qquad p \approx 0.013.$$

Обидва коефіцієнти вказують на помірний позитивний загальний зв'язок.

30. Але графік показує більше

Середні виходи:

$$91.88 \to 93.08 \to 94.35 \to 92.92.$$

Тобто вихід зростає від 70 до 80 °C, а при 85 °C знижується.

Через цей максимум залежність не є простою монотонною прямою.

$$\boxed{\text{помірний } r > 0 \text{ приховує максимум близько } 80°C}.$$

31. Порівняння з ANOVA

Для того самого набору:

$$F \approx 136.43, \qquad p < 0.0001, \qquad \eta^2 \approx 0.953.$$

ANOVA показує надзвичайно сильний ефект температурного режиму.

Кореляція:

$$r \approx 0.544.$$

є лише помірною.

Це не суперечність. ANOVA відповідає: «чи різняться середні між температурними рівнями?». Pearson відповідає: «наскільки добре залежність описується загальним лінійним напрямом?».

32. Груповий фактор і кількісна змінна — різні моделі

Якщо температура використовується як фактор:

$$70,\ 75,\ 80,\ 85 \rightarrow 4 \text{ категорії}.$$

ANOVA не припускає, що ефект між 70 і 75 має ту саму форму, що між 80 і 85.

Кореляція використовує числовий порядок і відстані між температурними значеннями.

33. Кореляція не означає причинність

$$\boxed{\text{correlation} \neq \text{causation}}.$$

Кореляція може бути зумовлена:

- прямим причинним впливом;
- зворотним причинним впливом;
- третьою змінною;
- спільним часовим трендом;
- змішуванням груп;
- випадковою асоціацією.

34. Коли причинний висновок сильніший

Причинна інтерпретація обґрунтованіша, коли:

- фактор контрольовано змінюється дослідником;
- використана рандомізація;
- контролюються інші важливі фактори;
- є часовий порядок «причина → наслідок»;
- результат відтворюється.

У добре спланованому технологічному експерименті зв'язок між заданою температурою і відгуком має сильніше причинне обґрунтування, ніж у простому спостережному наборі.

35. Спільний тренд як хибна кореляція

Якщо дві змінні одночасно зростають у часі, між ними може виникнути висока кореляція навіть без прямого причинного зв'язку.

Тому для часових даних потрібно аналізувати часову структуру, а не лише r.

36. Кореляційна матриця

Якщо змінних багато, можна обчислити матрицю:

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}.$$

На діагоналі завжди:

$$r_{ii} = 1.$$

Матриця симетрична:

$$r_{ij} = r_{ji}.$$

37. Чому у ПР6 не використовується велика кореляційна матриця

Мета ПР6 — не перевантажувати студента десятками коефіцієнтів, а чітко побачити зв'язок між:

$$X = \text{температура}, \qquad Y = \text{вихід продукту}.$$

Кореляційна матриця буде корисною у складніших багатофакторних дослідженнях.

38. Часткова кореляція — концептуально

Часткова кореляція оцінює зв'язок X і Y після статистичного контролю третьої змінної Z.

Наприклад, зв'язок температури й виходу при контролі вологості сировини.

Це важлива концепція, але розрахунок не входить до базової ПР6.

39. Множинна кореляція — концептуально

У багатофакторній моделі результат може одночасно залежати від кількох змінних:

$$Y = f(X_1, X_2, \ldots, X_p).$$

Для такого аналізу потрібна множинна регресія, яка розглядатиметься в лекції 14.

40. Від кореляції до регресії

Кореляція є симетричною:

$$r_{XY} = r_{YX}.$$

Регресія — асиметрична: одна змінна визначається як predictor, інша — response.

$$X \to Y.$$

Кореляція відповідає на питання про силу асоціації, регресія — про форму залежності та прогноз.

41. Коли висока кореляція не гарантує добрий прогноз

Можливі причини:

- мала вибірка;
- вузький діапазон X;
- викиди;
- екстраполяція;
- нестабільність процесу;
- зміна умов у майбутньому.

42. Excel для Pearson correlation

Функції:

- `CORREL(array1,array2)`;
- `PEARSON(array1,array2)`.

Для p-value можна обчислити:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

і використати t-розподіл із $n-2$ df.

43. Scatterplot в Excel

Потрібно використовувати саме XY Scatter, а не звичайний Line chart, оскільки X є кількісною координатою.

Line chart може трактувати значення X як категорії з однаковими відстанями. Для кореляції та регресії потрібна XY-діаграма.

44. Spearman у програмному забезпеченні

У ПР6 Spearman доцільно обчислювати в Real Statistics або Jamovi, оскільки ручне ранжування вже опрацьовано у ПР5 і повторювати великий ранговий розрахунок методично недоцільно.

45. Що повинен містити висновок про кореляцію

- напрям;
- величину коефіцієнта;
- p-value;
- форму scatterplot;
- зауваження щодо викидів;
- предметну інтерпретацію;
- застереження щодо причинності, якщо дизайн не експериментальний.

46. Приклад статистичного висновку ПР6

Між температурою оброблення та виходом продукту в діапазоні 70–85 °C спостерігається статистично значуща помірна позитивна лінійна кореляція: $r = 0.544,\ p = 0.006$. Рангова кореляція Spearman також позитивна: $\rho_s = 0.501, p = 0.013$.

47. Приклад технологічного висновку

У середньому підвищення температури в досліджуваному діапазоні пов'язане зі збільшенням виходу продукту, але залежність не є просто лінійною: максимальний середній вихід спостерігається при 80 °C, після чого при 85 °C він знижується.

Тому для визначення оптимального режиму однієї лінійної кореляції недостатньо.

48. Типові помилки

1. Обчислювати Pearson r без scatterplot.
2. Трактувати $r = 0$ як повну відсутність будь-якого зв'язку.
3. Ігнорувати нелінійність.
4. Ігнорувати викиди.
5. Порівнювати коефіцієнти з різних діапазонів X без застережень.
6. Ототожнювати значущість із силою кореляції.
7. Ототожнювати кореляцію з причинністю.
8. Використовувати Pearson для порядкових даних без обґрунтування.
9. Ігнорувати Spearman при монотонній, але нелінійній залежності.
10. Вважати сильний ANOVA і помірний r суперечливими результатами.
11. Використовувати Line chart замість XY Scatter.
12. Робити прогноз за межами дослідженого діапазону лише на основі високого r.

49. Зв'язок із практичною роботою №6

У другій, коротшій частині ПР6 студент:

1. перетворює групову таблицю ANOVA на 24 пари (X, Y);
2. будує scatterplot;
3. обчислює Pearson r у Excel;
4. обчислює p-value;
5. обчислює Spearman у спеціалізованому ПЗ;
6. порівнює ANOVA та correlation;
7. виявляє, чи є залежність лінійною, монотонною або має максимум;
8. формулює технологічний висновок.

50. Підсумок лекції

парні дані → scatterplot → лінійність/викиди → r_P → p → ρ_S → порівняння моделей → інтерпретація без хи

Кореляційний аналіз починається з геометрії даних, а не з одного коефіцієнта. Pearson оцінює лінійну асоціацію, Spearman — монотонну рангову. Значущий коефіцієнт не доводить причинності, а слабка лінійна кореляція не виключає сильної нелінійної залежності.

51. Контрольні питання

1. Що означає статистичний зв'язок між двома змінними?
2. Чому scatterplot потрібно будувати до обчислення r?
3. Який зв'язок називають позитивним?
4. Який зв'язок називають негативним?
5. Запишіть формулу Pearson r.
6. Як пов'язані кореляція та коваріація?
7. У яких межах лежить r?
8. Що означає $r = 0$?
9. Чи означає $r = 0$ повну відсутність зв'язку?
10. Які припущення бажані для Pearson correlation?
11. Які гіпотези перевіряють при тестуванні r?
12. Запишіть t-статистику для перевірки $H_0: \rho = 0$.
13. Скільки ступенів свободи має цей тест?
14. Чому значущий r не обов'язково є сильним?
15. Для чого використовується Fisher z-transformation?
16. Як викид може вплинути на r?
17. Що означає range restriction?
18. Чому змішування груп може створювати хибну кореляцію?
19. Що таке кореляція Spearman?
20. Який тип зв'язку оцінює Spearman?
21. Чим Pearson відрізняється від Spearman?
22. Чому сильний ANOVA може поєднуватися з помірним Pearson r?
23. Що означає твердження «correlation does not imply causation»?
24. За яких умов причинна інтерпретація є сильнішою?
25. Що таке кореляційна матриця?
26. Що таке часткова кореляція?
27. У чому різниця між кореляцією та регресією?
28. Чому висока кореляція не гарантує надійної екстраполяції?

Лекція 13

Парна лінійна регресія: побудова, оцінювання та прогнозування

1. Мета лекції та результати навчання

Мета лекції — сформувати цілісне розуміння парної лінійної регресії як моделі кількісної залежності між фактором і результативною змінною та навчити використовувати її для пояснення, оцінювання й прогнозування технологічних показників.

Після опрацювання лекції здобувач повинен уміти:

- розрізняти кореляційний і регресійний аналіз;
- записувати модель простої лінійної регресії;
- пояснювати зміст коефіцієнтів β_0 і β_1;
- оцінювати коефіцієнти методом найменших квадратів;
- обчислювати S_{xx}, S_{xy}, b_0, b_1;
- розрізняти SST, SSR і SSE;
- інтерпретувати коефіцієнт детермінації R^2;
- перевіряти статистичну значущість коефіцієнта нахилу;
- аналізувати залишки;
- будувати точковий прогноз, довірчий інтервал для середньої відповіді та prediction interval;
- розрізняти інтерполяцію й екстраполяцію;
- оцінювати момент досягнення критичного технологічного рівня.

2. План лекції

1. Регресія і кореляція: принципова різниця.
2. Модель $Y = \beta_0 + \beta_1 X + \varepsilon$.
3. Метод найменших квадратів.
4. Оцінювання b_0 та b_1.
5. Залишки та декомпозиція варіації.
6. Коефіцієнт детермінації R^2.
7. Перевірка значущості регресії.
8. Аналіз залишків.
9. Точковий прогноз.
10. Confidence interval і prediction interval.
11. Інтерполяція та екстраполяція.
12. Прогноз критичного моменту.

3. Кореляція та регресія

Кореляція відповідає на питання: наскільки сильно дві змінні пов'язані?

Регресія відповідає на інше питання: як змінюється результативна змінна Y при зміні фактора X, і як за X прогнозувати Y?

$$\boxed{\text{кореляція — симетрична,} \qquad \text{регресія — напрямлена модель } X \to Y}.$$

4. Модель простої лінійної регресії

Теоретична модель:

$$\boxed{Y = \beta_0 + \beta_1 X + \varepsilon}.$$

де:

- Y — результативна змінна;
- X — предиктор або фактор;
- β_0 — intercept;
- β_1 — slope;
- ε — випадкова похибка.

5. Вибіркова регресійна пряма

За даними оцінюється:

$$\boxed{\hat{Y} = b_0 + b_1 X}.$$

b_0 і b_1 — вибіркові оцінки невідомих β_0 та β_1.

6. Інтерпретація коефіцієнта b_1

b_1 показує середню зміну прогнозованого Y при збільшенні X на одну одиницю.

Якщо:

$$\hat{Y} = 2.00 + 0.268t,$$

де t — час зберігання у днях, а Y — $\log_{10} CFU/g$, то збільшення часу на 1 добу асоціюється із середнім збільшенням показника приблизно на 0.268 log-одиниці.

7. Інтерпретація intercept

b_0 — прогнозоване значення Y при:

$$X = 0.$$

Предметний зміст intercept залежить від того, чи має $X = 0$ фізичний сенс і чи лежить він у дослідженому діапазоні.

8. Метод найменших квадратів

Для кожного спостереження:

$$e_i = y_i - \hat{y}_i.$$

Метод OLS обирає b_0, b_1 так, щоб мінімізувати:

$$\boxed{SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}.$$

Ідея OLS. Найкраща пряма — та, для якої сума квадратів вертикальних відхилень фактичних точок від прогнозованої прямої мінімальна.

9. Допоміжні суми

$$\boxed{S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

$$\boxed{S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})\,(y_i - \bar{y})}.$$

10. Оцінка slope

$$\boxed{b_1 = \frac{S_{xy}}{S_{xx}}}.$$

11. Оцінка intercept

$$\boxed{b_0 = \bar{y} - b_1 \bar{x}}.$$

12. Демонстраційний набір

Час зберігання, діб:

$$t = 0,2,4,6,8,10,12,14,16.$$

Мікробіологічний показник:

$$Y = 2.1,2.5,3.0,3.6,4.1,4.7,5.2,5.8,6.3.$$

13. Вибіркові середні

$$\bar{x} = 8.$$

$$\bar{y} = 4.14444.$$

14. Обчислення S_{xx} та S_{xy}

$$S_{xx} = 240.$$

$$S_{xy} = 64.4.$$

15. Коефіцієнти регресії

$$b_1 = \frac{64.4}{240} = 0.26833.$$

$$b_0 = 4.14444 - 0.26833 \cdot 8 = 1.99778.$$

Отже:

$$\boxed{\hat{Y} = 1.9978 + 0.26833t}.$$

16. Прогнозовані значення та залишки

Для кожної точки:

$$\hat{y}_i = b_0 + b_1 x_i,$$

$$e_i = y_i - \hat{y}_i.$$

Наприклад, при $t = 0$:

$$\hat{y} = 1.9978, \qquad e = 2.1 - 1.9978 = 0.1022.$$

17. Декомпозиція варіації

Загальна сума квадратів:

$$\boxed{SST = \sum (y_i - \bar{y})^2}.$$

Регресійна:

$$\boxed{SSR = \sum (\hat{y}_i - \bar{y})^2}.$$

Залишкова:

$$\boxed{SSE = \sum (y_i - \hat{y}_i)^2}.$$

Виконується:

$$\boxed{SST = SSR + SSE}.$$

18. Значення для демонстраційного прикладу

$$SST = 17.30222.$$

$$SSR = 17.28067.$$

$$SSE = 0.02156.$$

19. Коефіцієнт детермінації

$$\boxed{R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}}.$$

Для прикладу:

$$R^2 \approx 0.99875.$$

Тобто приблизно 99.875% варіативності Y у межах цієї простої лінійної моделі пояснюється зміною часу.

20. Зв'язок R^2 і r

Для простої лінійної регресії з intercept:

$$\boxed{R^2 = r^2}.$$

У прикладі:

$$r \approx 0.99938.$$

21. Залишкова стандартна похибка

Оцінка стандартного відхилення випадкової похибки:

$$\boxed{s_e = \sqrt{\frac{SSE}{n-2}}}.$$

Для прикладу:

$$s_e \approx 0.05549.$$

22. Стандартна похибка slope

$$\boxed{SE_{b_1} = \frac{s_e}{\sqrt{S_{xx}}}}.$$

Для прикладу:

$$SE_{b_1} \approx 0.003582.$$

23. Перевірка значущості slope

Гіпотези:

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 \neq 0.$$

Статистика:

$$\boxed{t = \frac{b_1}{SE_{b_1}}}.$$

Ступені свободи:

$$df = n - 2.$$

24. Значущість у прикладі

$$t \approx 74.91, \qquad df = 7.$$

$$p \approx 2 \cdot 10^{-11}.$$

Отже, коефіцієнт нахилу статистично значущо відрізняється від нуля.

25. Довірчий інтервал для slope

$$\boxed{b_1 \pm t_{1-\alpha/2;n-2} SE_{b_1}}.$$

Для прикладу:

$$95\%\, CI_{b_1} \approx [0.2599;\ 0.2768].$$

26. F-test загальної значущості

У простій регресії:

$$\boxed{F = t^2}$$

для тесту $H_0: \beta_1 = 0$.

У прикладі:

$$F \approx 5611.76.$$

27. Припущення лінійної регресії

- лінійність середньої залежності;
- незалежність залишків;
- приблизна сталість дисперсії залишків;
- для класичної інференції — приблизна нормальність залишків;
- відсутність надмірно впливових точок.

28. Чому перевіряють залишки, а не лише Y

Регресія моделює:

$$Y = \hat{Y} + \varepsilon.$$

Тому припущення стосуються передусім випадкової частини ε, яку в даних оцінюють залишками e_i.

29. Графік залишків

Будують:

$$e_i \text{ проти } \hat{y}_i$$

або:

$$e_i \text{ проти } x_i.$$

Для адекватної простої лінійної моделі бажано бачити випадкову хмару навколо нуля без систематичної структури.

30. Ознаки проблем на residual plot

- дуга — нелінійність;
- «воронка» — неоднакова дисперсія;
- довгі серії одного знака — можлива залежність;
- окремі дуже великі залишки — потенційні викиди.

31. Нормальність залишків

Перевіряється Q–Q plot та, за потреби, Shapiro–Wilk.

Нормальність особливо важлива для точних t-, F- та interval-оцінок у малих вибірках.

32. Гомоскедастичність

Класична модель припускає приблизно сталу дисперсію:

$$Var(\varepsilon \mid X = x) = \sigma^2.$$

Якщо розсіювання залишків зростає разом із X, маємо heteroskedasticity.

33. Точковий прогноз

Для заданого x_0:

$$\boxed{\hat{y}_0 = b_0 + b_1 x_0}.$$

Для $t_0 = 15$:

$$\hat{Y}(15) = 1.9978 + 0.26833 \cdot 15 \approx 6.0228.$$

34. Два різні інтервали прогнозування

34.1. Confidence interval для середньої відповіді

Оцінює середній Y для всіх об'єктів при $X = x_0$.

$$\boxed{\hat{y}_0 \pm t_{\text{crit}} s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}}.$$

34.2. Prediction interval для нового спостереження

Оцінює діапазон для однієї нової індивідуальної реалізації:

$$\boxed{\hat{y}_0 \pm t_{\text{crit}} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}}.$$

35. Чому prediction interval ширший

CI для середньої відповіді враховує невизначеність оцінки середнього.

PI додатково враховує випадкову варіативність окремого нового спостереження.

$$\boxed{PI \text{ ширший за } CI_{\text{mean}}}.$$

36. Приклад для $t_0 = 15$

$$\hat{Y} = 6.0228.$$

$$95\%\, CI_{\text{mean}} \approx [5.9491;\ 6.0965].$$

$$95\%\, PI \approx [5.8723;\ 6.1733].$$

37. Інтерполяція

Прогноз усередині дослідженого діапазону:

$$x_{\min} \le x_0 \le x_{\max}.$$

Зазвичай він надійніший, якщо модель адекватна.

38. Екстраполяція

Прогноз поза межами дослідженого діапазону:

$$x_0 < x_{\min} \quad \text{або} \quad x_0 > x_{\max}.$$

Екстраполяція ризикована. Висока R^2 усередині діапазону не гарантує, що той самий закон діятиме поза ним.

39. Прогноз моменту досягнення критичного рівня

Нехай задано критичне значення:

$$Y_{\text{crit}}.$$

Для лінійної моделі:

$$Y_{\text{crit}} = b_0 + b_1 t_{\text{crit}}.$$

Звідси:

$$\boxed{t_{\text{crit}} = \frac{Y_{\text{crit}} - b_0}{b_1}}.$$

40. Приклад критичного часу

Для:

$$Y_{\text{crit}} = 6.0,$$

$$b_0 = 1.9978, \qquad b_1 = 0.26833,$$

отримаємо:

$$t_{\text{crit}} \approx \frac{6.0 - 1.9978}{0.26833} \approx 14.92 \text{ доби.}$$

41. Чи є це «терміном придатності»?

Не автоматично.

Модельний момент перетину критичного рівня — це статистична оцінка за конкретним показником, а не юридично або нормативно встановлений термін придатності.

Для реального shelf-life потрібно враховувати:

- мікробіологічну безпеку;
- сенсорні показники;
- фізико-хімічні зміни;
- варіативність партій;
- умови зберігання;
- запас безпеки;
- нормативні вимоги.

42. Лінійність не повинна прийматися автоматично

Якщо scatterplot або residual plot демонструє кривину, можна розглядати:

- поліноміальну регресію;
- експоненційну модель;
- логістичну модель;
- трансформацію змінних.

43. Коротко про нелінійні моделі

Наприклад:

$$Y = a + bx + cx^2$$

може описувати оптимум або максимум.

Але у ПР7 основна мета — повністю засвоїти просту лінійну модель.

44. Excel для парної регресії

Основні функції:

- `SLOPE`;
- `INTERCEPT`;
- `RSQ`;
- `CORREL`;
- `STEYX`;
- `FORECAST.LINEAR`;

- `LINEST`.

45. Data Analysis → Regression

Вихід Excel містить:

- Multiple R;
- R Square;
- Adjusted R Square;
- Standard Error;
- ANOVA;
- Coefficients;
- Standard Error;
- t Stat;
- P-value;
- Lower/Upper 95%.

46. Чому потрібні residuals у Excel

У параметрах Regression потрібно вмикати:

- Residuals;
- Residual Plots;
- Line Fit Plots — за потреби.

Без residual diagnostics висока R^2 може створити хибне відчуття адекватності моделі.

47. Типові помилки

1. Будувати регресію без scatterplot.
2. Ототожнювати кореляцію і регресію.
3. Інтерпретувати intercept поза предметним змістом.
4. Оцінювати модель лише за R^2.
5. Не аналізувати залишки.
6. Плутати CI середньої відповіді та PI нового спостереження.
7. Використовувати PI там, де потрібен CI mean, і навпаки.
8. Беззастережно екстраполювати.
9. Вважати модельний t_{crit} автоматичним терміном придатності.
10. Округлювати b_0, b_1 надто рано.

48. Зв'язок із ПР7

У ПР7 студент виконує повний цикл:

1. будує scatterplot;
2. обчислює $\bar{x}, \bar{y}, S_{xx}, S_{xy}$;

3. визначає b_0, b_1;
4. обчислює прогнозовані значення та залишки;
5. визначає SST, SSR, SSE;
6. обчислює R^2;
7. перевіряє slope;
8. аналізує residuals;
9. виконує прогноз;
10. будує CI mean та PI;
11. оцінює критичний момент;
12. формулює технологічний висновок.

49. Підсумок лекції

scatterplot → $\hat{Y} = b_0 + b_1 X$ → OLS → $residuals$ → $SST = SSR + SSE$ → R^2 → t/F → CI/PI → прогноз → те

Лінійна регресія є не просто способом провести пряму через точки, а повною статистичною моделлю з оцінкою невизначеності, перевіркою припущень і обмеженнями застосування.

50. Контрольні питання

1. Чим регресія відрізняється від кореляції?
2. Запишіть модель простої лінійної регресії.
3. Що означає β_1?
4. Що означає intercept?
5. У чому полягає метод найменших квадратів?
6. Що таке залишок?
7. Як обчислюється S_{xx}?
8. Як обчислюється S_{xy}?
9. Запишіть формулу b_1.
10. Запишіть формулу b_0.
11. Що означають SST, SSR, SSE?
12. Яка тотожність виконується між ними?
13. Що означає R^2?
14. Як пов'язані R^2 і r у простій регресії?
15. Як перевіряється значущість slope?
16. Які припущення має лінійна регресія?
17. Чому потрібно аналізувати залишки?
18. Що означає «воронка» на residual plot?
19. Чим CI mean відрізняється від PI?
20. Чому PI ширший?

21. Що таке інтерполяція?
22. Що таке екстраполяція?
23. Чому екстраполяція ризикована?
24. Як визначається t_{crit}?
25. Чому модельний критичний час не слід автоматично називати терміном придатності?

Лекція 14

Множинна регресія, діагностика моделі та технологічне прогнозування

1. Мета лекції та результати навчання

Мета лекції — узагальнити регресійний підхід для ситуацій, коли результативний показник залежить від кількох технологічних факторів, та сформувати критерії оцінювання адекватності, стійкості й практичної придатності регресійної моделі.

Після опрацювання лекції здобувач повинен уміти:

- записувати модель множинної лінійної регресії;
- інтерпретувати часткові регресійні коефіцієнти;
- розуміти різницю між R^2 та adjusted R^2;
- пояснювати F-test загальної значущості моделі;
- пояснювати t-тести окремих коефіцієнтів;
- розуміти явище multicollinearity;
- інтерпретувати VIF;
- аналізувати залишки, leverage та influential observations;
- розрізняти explanatory та predictive моделі;
- оцінювати ризики overfitting;
- пояснювати межі інтерполяції та екстраполяції у багатофакторному просторі;
- формулювати коректний технологічний висновок за регресійною моделлю.

2. План лекції

1. Навіщо потрібна множинна регресія.
2. Загальна форма моделі.
3. Частковий зміст коефіцієнтів.
4. R^2 та adjusted R^2.
5. F-test загальної моделі.
6. t-тести окремих коефіцієнтів.
7. Multicollinearity та VIF.
8. Залишкова діагностика.
9. Leverage, outliers та influential observations.
10. Overfitting і складність моделі.
11. Explanatory vs predictive modeling.
12. Прогнозування та межі застосування.
13. Регресія для shelf-life та технологічної оптимізації.

3. Чому одного фактора часто недостатньо

У реальному харчовому виробництві результат рідко залежить лише від однієї змінної. Наприклад, термін зберігання може залежати одночасно від:

- температури;
- вологості;
- типу пакування;
- початкового мікробіологічного навантаження;
- pH;
- активності води;
- концентрації консервантів.

Множинна регресія дозволяє оцінити внесок кількох факторів у спільній моделі та відокремити частковий зв'язок одного фактора за фіксованих інших.

4. Модель множинної лінійної регресії

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon}.$$

Оцінена модель:

$$\boxed{\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p}.$$

5. Інтерпретація часткового коефіцієнта

Коефіцієнт b_j показує очікувану зміну Y при збільшенні X_j на одну одиницю **за фіксованих інших предикторів**.

Нехай:

$$\hat{Y} = 4.2 + 0.15T - 0.30pH.$$

Коефіцієнт 0.15 при температурі означає: при однаковому pH збільшення температури на 1 одиницю асоціюється зі збільшенням прогнозованого Y на 0.15.

6. «За інших рівних умов» — ключ до множинної регресії

На відміну від простої кореляції, множинна модель намагається оцінити **частковий ефект**.

$$\boxed{b_j\text{: } \Delta Y \text{ при зміні } X_j, \text{ коли інші } X \text{ фіксовані}}.$$

Саме тому знак або величина коефіцієнта можуть відрізнятися від простої парної кореляції.

7. Матричне представлення — концептуально

Множинну модель можна записати:

$$\boxed{\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}}.$$

OLS-оцінка:

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}.$$

У цьому курсі матричні обчислення вручну не виконуються. Формула потрібна для розуміння того, що всі коефіцієнти оцінюються одночасно.

8. R-squared у множинній моделі

$$\boxed{R^2 = 1 - \frac{SSE}{SST}}.$$

Воно показує частку варіативності Y, описану сукупністю предикторів.

9. Проблема звичайного R^2

При додаванні нового предиктора R^2 не зменшується, навіть якщо нова змінна майже не несе корисної інформації.

$$R^2_{\text{new}} \geq R^2_{\text{old}}.$$

Тому для порівняння моделей різної складності використовують adjusted R^2.

10. Adjusted R-squared

$$\boxed{R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n-1}{n-p-1}}.$$

де p — кількість предикторів без intercept.

Adjusted R^2 штрафує модель за зайві змінні.

11. Інтерпретація R^2 і adjusted R^2

Якщо додавання нового фактора значно підвищує R^2, але adjusted R^2 майже не змінюється або зменшується, користь нового предиктора сумнівна.

$$\boxed{\text{більше предикторів} \neq \text{автоматично краща модель}}.$$

12. Загальний F-test моделі

Нульова гіпотеза:

$$\boxed{H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0}.$$

Альтернативна:

$$H_1 : \text{принаймні один } \beta_j \neq 0.$$

13. F-статистика множинної регресії

$$\boxed{F = \frac{SSR/p}{SSE/(n-p-1)}}.$$

Ступені свободи:

$$df_1 = p, \qquad df_2 = n - p - 1.$$

Малий p-value F-test означає, що модель як ціле дає кращий опис Y, ніж модель лише з intercept.

14. Значущість окремого коефіцієнта

Для кожного:

$$H_0: \beta_j = 0.$$

Статистика:

$$\boxed{t_j = \frac{b_j}{SE(b_j)}}.$$

Важливо: цей тест оцінює внесок X_j **після врахування інших предикторів**.

15. Глобально значуща модель і незначущий окремий коефіцієнт

Можлива ситуація:

$$p_F < 0.001,$$

але один із:

$$p_j > 0.05.$$

Це не суперечність. Модель як сукупність предикторів значуща, але конкретний фактор після контролю інших може не мати незалежного внеску.

16. Multicollinearity

Multicollinearity виникає, коли предиктори сильно пов'язані між собою.

Наприклад:

$$Corr(X_1, X_2) \approx 0.95.$$

У такій ситуації модель важко визначає, який із двох факторів «відповідає» за спільний ефект.

17. Наслідки multicollinearity

- зростають стандартні похибки коефіцієнтів;
- окремі p-value стають великими;
- знаки коефіцієнтів можуть бути нестабільними;
- невеликі зміни вибірки сильно змінюють b_j;
- інтерпретація часткових ефектів ускладнюється.

Високе R^2 не виключає серйозної multicollinearity.

18. Variance Inflation Factor

Для предиктора X_j:

$$\boxed{VIF_j = \frac{1}{1 - R_j^2}}.$$

де R_j^2 — коефіцієнт детермінації при регресії X_j на всі інші предиктори.

19. Інтерпретація VIF

VIF	Орієнтовна інтерпретація
≈1	мультиколінеарність практично відсутня
2–5	помірна, потребує контексту
>5	можлива серйозна проблема
>10	часто вважається дуже сильною

Ці пороги не є абсолютними правилами.

20. Що робити при високому VIF

- перевірити, чи змінні не дублюють одна одну;
- видалити один із надмірно схожих предикторів, якщо це змістовно;
- об'єднати показники;
- збільшити обсяг даних;
- використати спеціальні методи regularization — у складніших задачах.

21. Чому не слід видаляти змінну лише через p>0.05

Фактор може бути важливим:

- як confounder;
- за предметною теорією;
- для стабільності інших коефіцієнтів;
- для прогнозної якості.

Модель будується не тільки автоматичним відбором «значущих колонок».

22. Припущення множинної лінійної регресії

- правильна функціональна форма;
- незалежність спостережень;
- відсутність сильної multicollinearity;
- приблизна гомоскедастичність залишків;
- приблизна нормальність залишків для класичної інференції;
- відсутність надмірно впливових спостережень.

23. Residual diagnostics

Необхідно аналізувати:

- residuals vs fitted;
- Q–Q plot;
- залишки у часовому порядку, якщо дані послідовні;

- studentized residuals;
- leverage;
- Cook's distance — у розширеному аналізі.

24. Стандартизовані та studentized residuals

Звичайні залишки мають різну теоретичну дисперсію. Тому для пошуку нетипових точок використовують стандартизовані варіанти.

Орієнтовно значення за межами:

$$|r_i| > 2$$

заслуговують на увагу, а:

$$|r_i| > 3$$

можуть бути сильним сигналом.

Але це не автоматичне правило вилучення.

25. Leverage

Leverage характеризує, наскільки незвичайним є спостереження у просторі предикторів X.

Точка з крайніми значеннями X може мати високе leverage навіть при невеликому залишку.

26. Outlier і high-leverage point — не одне й те саме

Тип точки	Особливість
Outlier in Y	великий вертикальний залишок
High leverage	нетипове положення у просторі X
Influential point	суттєво змінює оцінену модель

27. Cook's distance — концептуально

Cook's distance оцінює, наскільки зміняться fitted values при вилученні конкретного спостереження.

Високе значення — сигнал для детального аналізу, а не автоматична підстава видалити точку.

28. Influential observation

Впливова точка може бути:

- помилкою;
- реальним рідкісним режимом;
- важливим спостереженням на межі технологічного діапазону.

Потрібен sensitivity analysis — порівняння моделі з точкою і без неї.

29. Нелінійність у множинній моделі

Лінійна форма:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

не означає, що всі технологічні залежності реально прямолінійні.

Можна додавати:

$$X^2, \quad X^3, \quad \ln X, \quad X_1X_2.$$

30. Поліноміальний термін

Для оптимуму температури:

$$\boxed{Y = \beta_0 + \beta_1 T + \beta_2 T^2 + \varepsilon}.$$

Наявність T^2 дозволяє описати максимум або мінімум.

31. Interaction term

Якщо ефект температури залежить від вологості:

$$\boxed{Y = \beta_0 + \beta_1 T + \beta_2 W + \beta_3 TW + \varepsilon}.$$

Коефіцієнт β_3 описує interaction.

Взаємодія означає: ефект одного фактора змінюється залежно від рівня іншого.

32. Чому interaction важлива у харчових технологіях

Приклади:

- ефект температури залежить від pH;
- ефект пакування залежить від температури зберігання;
- ефект концентрації добавки залежить від часу оброблення.

33. Explanatory model

Мета explanatory modeling — зрозуміти структуру зв'язків і оцінити часткові ефекти.

Важливі:

- коефіцієнти;
- довірчі інтервали;
- предметна обґрунтованість;
- контроль confounding;
- стабільність інтерпретації.

34. Predictive model

Мета predictive modeling — якнайточніше прогнозувати нові значення.

Критично важливі:

- помилка на нових даних;
- валідація;
- стійкість;

- відсутність overfitting.

добре пояснювати ≠ автоматично добре прогнозувати.

35. Overfitting

Overfitting виникає, коли модель надто добре підлаштовується під випадковий шум конкретної вибірки.

Ознаки:

- дуже багато параметрів відносно n;
- майже ідеальний fit на навчальних даних;
- нестабільні коефіцієнти;
- поганий прогноз на нових даних.

36. Чому $R^2 = 0.99$ не гарантує якісну модель

Високе R^2 може співіснувати з:

- нелінійною структурою залишків;
- multicollinearity;
- впливовими точками;
- overfitting;
- неправильною екстраполяцією.

37. Валідація моделі

Найкраще оцінювати прогнозну якість на даних, які не використовувались для оцінювання коефіцієнтів.

Базові підходи:

- training/test split;
- cross-validation;
- зовнішня незалежна вибірка.

У базовій ПР7 ці процедури лише вводяться концептуально.

38. Метрики прогнозної помилки

Типові:

$$MAE = \frac{1}{n}\sum|y_i - \hat{y}_i|.$$

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}.$$

RMSE сильніше карає великі помилки.

39. Модель shelf-life з кількома факторами

Концептуальний приклад:

$$\hat{Y} = b_0 + b_1 t + b_2 T + b_3 pH + b_4 (t \cdot T).$$

де:

- t — час;
- T — температура;
- pH — кислотність;
- tT — interaction.

Така модель дозволяє врахувати, що швидкість зміни показника у часі залежить від температури.

40. Критичний час у множинній моделі

Якщо модель лінійна за часом:

$$\hat{Y} = A + Bt,$$

де A і B залежать від інших фіксованих факторів, тоді:

$$t_{\text{crit}} = \frac{Y_{\text{crit}} - A}{B}.$$

Отже, критичний час може відрізнятися для різних температур, pH або пакування.

41. Необхідність prediction uncertainty

Для реального shelf-life недостатньо одного:

$$\hat{t}_{\text{crit}}.$$

Потрібно враховувати:

- невизначеність коефіцієнтів;
- варіативність між партіями;
- варіативність окремих продуктів;
- похибку вимірювання;
- зміну зовнішніх умов.

42. Інтерполяція у багатовимірному просторі

Навіть якщо кожний окремий predictor лежить у знайомому діапазоні, їх комбінація може бути новою.

Наприклад, модель могла бачити:

$$(T = 5°C,\ pH = 4.0)$$

та:

$$(T = 15°C,\ pH = 6.0),$$

але не бачити комбінацію:

$$(T = 15°C,\ pH = 4.0).$$

Тому поняття екстраполяції у множинній регресії складніше, ніж просте виходження одного X за min/max.

43. Технологічна правдоподібність моделі

Коефіцієнти повинні оцінюватися не лише статистично, а й предметно.

Якщо модель прогнозує фізично неможливі значення, це сигнал:

- неправильної форми моделі;
- екстраполяції;
- помилок у даних;
- відсутнього важливого фактора.

44. Модель і причинний висновок

Множинна регресія може статистично «контролювати» змінні, але сама по собі не перетворює спостережне дослідження на експеримент.

regression adjustment ≠ автоматичний доказ причинності.

Причинність залежить від дизайну, рандомізації, часової послідовності та контролю confounders.

45. Вибір предикторів

Доцільно враховувати:

- предметну теорію;
- дизайн експерименту;
- якість вимірювання;
- обсяг вибірки;
- multicollinearity;
- ціль моделі.

Автоматичний stepwise selection не повинен бути єдиною стратегією.

46. Принцип ієрархії

Якщо в модель включено interaction:

$$X_1X_2,$$

зазвичай залишають і main effects:

$$X_1, \qquad X_2.$$

Аналогічно, якщо є X^2, зазвичай залишають X.

47. Масштабування предикторів

Коефіцієнти в різних одиницях не можна прямо порівнювати за абсолютною величиною.

Наприклад:

- температура — °C;
- pH — одиниці pH;
- концентрація — mg/L.

Для порівняння впливів інколи використовують стандартизовані коефіцієнти.

48. Стандартизовані коефіцієнти — концептуально

Якщо X та Y стандартизувати:

$$Z = \frac{X - \bar{x}}{s_x},$$

коефіцієнти стають безрозмірними й можуть бути зручні для порівняння відносної сили факторів.

Але для технологічної інтерпретації часто корисніші незастандартизовані коефіцієнти у фізичних одиницях.

49. Excel для множинної регресії

Data → Data Analysis → Regression.

У поле Input Y Range задається одна результативна змінна, а Input X Range може містити кілька сусідніх стовпців предикторів.

Потрібно аналізувати:

- R Square;
- Adjusted R Square;
- Standard Error;
- ANOVA F і Significance F;
- Coefficients;
- Standard Error;
- t Stat;
- P-value;
- 95% CI;
- Residuals.

50. Що Excel не дає автоматично достатньо зручно

Для повної діагностики часто зручніші Jamovi, JASP або R, особливо для:

- VIF;
- standardized residuals;

- Cook's distance;
- interaction plots;
- розширених diagnostic plots.

51. Приклад структури статистичного висновку

Множинна лінійна модель була статистично значущою за загальним F-test. Adjusted R^2 показав, що модель пояснює суттєву частину варіативності результативного показника з урахуванням числа предикторів. Значущі коефіцієнти інтерпретовано як часткові ефекти за фіксованих інших факторів. VIF не вказував на критичну multicollinearity, а аналіз залишків не виявив грубого порушення лінійності чи гомоскедастичності.

52. Приклад технологічного висновку

Модель дозволяє кількісно оцінювати зміну показника якості залежно від сукупності технологічних умов і може використовуватися для інтерполяційного прогнозу в межах дослідженого факторного простору. Для використання як виробничої прогнозної моделі потрібна перевірка на незалежних даних.

53. Повний алгоритм оцінювання регресійної моделі

дизайн → scatterplots → модель → R^2/R^2_{adj} → F → t_j → VIF → residual diagnostics → influential points → va

54. Типові помилки

1. Додавати багато предикторів лише для збільшення R^2.
2. Ігнорувати adjusted R^2.
3. Інтерпретувати коефіцієнт без фрази «за фіксованих інших факторів».
4. Ігнорувати multicollinearity.
5. Видаляти змінні лише за p-value.
6. Не аналізувати residuals.
7. Вважати outlier, high leverage та influential point одним поняттям.
8. Видаляти впливову точку без предметної перевірки.
9. Ігнорувати interaction.
10. Порівнювати абсолютні коефіцієнти в різних одиницях як міру важливості.
11. Використовувати training fit як доказ прогнозної якості.
12. Екстраполювати у невивчену комбінацію факторів.
13. Ототожнювати регресійну асоціацію з причинністю.
14. Називати прогнозований критичний час нормативним shelf-life без валідації.

55. Зв'язок із практичною роботою №7

ПР7 зосереджується на **простій лінійній регресії**, щоб студент повністю відпрацював ручний алгоритм. Лекція 14 показує, як ця логіка розширюється:

$$\hat{Y} = b_0 + b_1 X \quad \rightarrow \quad \hat{Y} = b_0 + b_1 X_1 + \cdots + b_p X_p.$$

У практичній роботі студент повинен особливо добре засвоїти універсальні для обох моделей принципи:

- не оцінювати модель лише за R^2;
- аналізувати residuals;
- перевіряти статистичну значущість;
- розрізняти інтерполяцію і екстраполяцію;
- подавати uncertainty разом із прогнозом;
- формулювати предметно обмежений технологічний висновок.

56. Підсумок лекції

$Y = f(X_1, \dots, X_p) \rightarrow R^2_{adj} \rightarrow F \rightarrow t_j \rightarrow VIF \rightarrow$ residuals → influence → validation → prediction → технологічне

Множинна регресія є природним продовженням простої лінійної моделі. Її цінність полягає не в максимізації R^2, а в побудові статистично обґрунтованої, предметно правдоподібної та перевіреної моделі, здатної підтримувати технологічні рішення.

57. Контрольні питання

1. Коли потрібна множинна регресія?
2. Запишіть загальну модель множинної лінійної регресії.
3. Як інтерпретується b_j?
4. Що означає «за фіксованих інших предикторів»?
5. Чому парна кореляція і частковий регресійний коефіцієнт можуть мати різні значення?
6. Чому R^2 не зменшується при додаванні предиктора?
7. Для чого потрібний adjusted R^2?
8. Які гіпотези перевіряє загальний F-test?
9. Що перевіряє t-test окремого коефіцієнта?
10. Чи може модель бути глобально значущою при незначущому окремому коефіцієнті?
11. Що таке multicollinearity?
12. Які наслідки високої multicollinearity?
13. Як обчислюється VIF?
14. Що означає VIF близький до 1?
15. Чому не слід механічно видаляти предиктор лише через високий p-value?
16. Які припущення має множинна регресія?
17. Що таке leverage?
18. Чим outlier відрізняється від high-leverage point?
19. Що називають influential observation?
20. Для чого використовується Cook's distance?
21. Що таке interaction term?
22. Наведіть технологічний приклад interaction.

23. Чим explanatory model відрізняється від predictive model?
24. Що таке overfitting?
25. Чому високий R^2 не гарантує добрий прогноз?
26. Для чого потрібна validation?
27. Що таке MAE та RMSE?
28. Чому екстраполяція в множинній регресії складніша, ніж у простій?
29. Чому regression adjustment не є автоматичним доказом причинності?
30. Які принципи ПР7 залишаються актуальними для множинної регресії?

Список літератури

1. Шелестовський Б. Г., Габрусєв Г. В., Габрусєва І. Ю. Вища математика: теорія ймовірностей та математична статистика : навч. посіб. Тернопіль : СМП «Тайп», 2023. 142 с. URL: http://elartu.tntu.edu.ua/handle/lib/41009.
2. Плічко А. М., Акбаш К. С., Луньова М. В. Математична статистика : навчальний посібник. Кропивницький : Видавець Лисенко С. В., 2024. 220 с.
3. Білущак Г. І. Аналітичні та чисельні методи дослідження. Статистичні методи аналізу даних : навчальний посібник для здобувачів вищої освіти. Львів : Растр-7, 2024. 315 с.
4. Дударєв І. М., Кузьмін О. В. Практикум з методології наукових досліджень : навчальний посібник. Одеса : Олді+, 2023. 278 с.
5. Михайлицька О. Р., Сливка Н. Б., Наговська В. О., Білик О. Я. Методологія наукових досліджень : навчальний посібник для здобувачів вищої освіти спеціальності 181 «Харчові технології». Львів : Простір-М, 2024. 258 с.
6. Горбачук В. М., Кушлик-Дивульська О. І. Теорія ймовірностей та математична статистика : підручник для здобувачів ступеня бакалавра за технічними та економічними спеціальностями. Київ : КПІ ім. Ігоря Сікорського, 2023. 351 с.
7. Бідюк П. І., Данилов В. Я., Жиров О. Л. Прикладна статистика : навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2023. 186 с.
8. Костюк В. О., Мількін І. В., Славута О. І. Статистика : підручник. Харків : ХНУМГ ім. О. М. Бекетова, 2023. 204 с.
9. Григорків В. С., Вінничук О. Ю., Григорків М. В., Маханець Л. Л. Статистика: основи теорії та практикум : навчальний посібник. Чернівці : Чернівецький нац. ун-т ім. Юрія Федьковича : Рута, 2022. 304 с.
10. Герасименко С. С., Голубова Г. В., Потапова М. Ю., Червона С. П. Статистика : навчальний посібник / за ред. О. Г. Осауленка. Київ : НАСОА, 2022. 265 с.
11. Загальна теорія статистики : підручник / І. А. Дмитрієв, О. А. Дмитрієва, О. М. Гіржева, А. В. Непран, Н. О. Бірченко, А. А. Воронкова, Н. В. Чуйко ; за ред. А. В. Непрана, І. А. Дмитрієва. Харків : ПП Іванченка, 2022. 720 с.