ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД УКООПСПІЛКИ
«ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ» (ПУЕТ)
Кафедра комп'ютерних наук та інформаційних технологій

О. О. Ємець, Т. О. Парфьонова

ТЕОРІЯ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА
(Частина 2)

НАВЧАЛЬНО-МЕТОДИЧНИЙ ПОСІБНИК
для самостійного вивчення навчальної
дисципліни студентами денної форми навчання
спеціальності 122 Комп'ютерні науки освітня програма
«Комп'ютерні науки» ступеня бакалавра

Полтава
ПУЕТ
2023

Автори: ***О. О. Ємець,*** д. ф.-м. н., професор;
Т. О. Парфьонова, к. ф.-м. н., доцент, доцент кафедри комп'ютерних наук та інформаційних технологій Вищого навчального закладу Укоопспілки «Полтавський університет економіки і торгівлі».

Рецензенти: ***Т. В. Чілікіна,*** к. ф.-м. н., доцент, доцент кафедри комп'ютерних наук та інформаційних технологій Вищого навчального закладу Укоопспілки «Полтавський університет економіки і торгівлі»;
О. П. Кошова, к. п. н., доцент, доцент кафедри комп'ютерних наук та інформаційних технологій Вищого навчального закладу Укоопспілки «Полтавський університет економіки і торгівлі».

Рекомендовано до видання, розміщення в електронній бібліотеці та використання в освітньому процесі на засіданні кафедри комп'ютерних наук та інформаційних технологій ПУЕТ
15 вересня 2022 р., *протокол №* ***2***

Ємець О. О.
Теорія ймовірностей і математична статистика (Частина 2) : навчально-методичний посібник для самостійного вивчення навчальної дисципліни студентами денної форми навчання спеціальності 122 Комп'ютерні науки освітня програма «Комп'ютерні науки» ступеня бакалавра / О. О. Ємець, Т. О. Парфьонова. – Полтава : ПУЕТ, 2023. – 54 с. – 1 електрон. опт. диск (CVD-ROM).

Відповідальні за зміст навчально-методичного видання автори, рецензенти та завідувач кафедри комп'ютерних наук та інформаційних технологій ***О. В. Ольховська***

ВСТУП

Дисципліна *«Теорія ймовірностей і математична статистика» вивчається студентами очної форми навчання спеціальності 122 Комп'ютерні науки, вона є складовою частиною галузі знань 12 «Інформаційні технології» і розроблена на основі «Освітньо-професійної програми», робочого навчального плану бакалавра з комп'ютерних наук.*

Основною метою вивчення дисципліни «Теорія ймовірностей і математична статистика» є формування знань, умінь і практичних навичок застосування основних методів теорії ймовірностей і математичної статистики, які необхідні для аналізування і прогнозування законів, що описують економічні і соціальні явища та процеси.

Головним завданням дисципліни є здобуття фундаментальних теоретичних знань сучасної теорії ймовірностей і основних методів обробки статистичних даних.

Предметом вивчення дисципліни «Теорія ймовірностей і математична статистика» є випадкові події, випадкові величини та їх функції, розподіли випадкових величин та їх статистичні закономірності. Після вивчення дисципліни студент повинен:

знати:

- *основні завдання математичної статистики;*
- *основні методи оцінювання невідомих параметрів;*
- *як проводити точкове й інтервальне оцінювання невідомих параметрів;*
- *як перевіряються статистичні гіпотези, основні критерії перевірки статистичних гіпотез;*
- *критерії згоди: Стьюдента, Пірсона, Фішера-Снедекора;*
- *критерії Барлетта, Кочрена;*
- *класифікацію регресій;*
- *метод найменших квадратів, необхідні умови його застосування для прогнозування;*

уміти:

- *будувати варіаційні ряди;*
- *будувати емпіричну функцію розподілу, полігон частот, гістограму;*
- *знаходити вибіркове середнє, дисперсію, коефіцієнт кореляції;*
- *знаходити точкові й інтервальні оцінки невідомих параметрів, перевіряти їх незсуненість та ефективність;*
- *використовувати критерії Пірсона, Стюдента, Фішера-Снедекора, Кочрена задля перевірки статистичних гіпотез;*
- *будувати лінійну регресію.*

Вивчення дисципліни базується на елементах знань з дисциплін «Дискретна математика», «Алгебра та геометрія», «Математичний аналіз».

Дисципліна є основою для дисциплін «Архітектура обчислювальних систем», «Елементи комбінаторної оптимізації», «Курсовий проект із фаху», «Методи оптимізації і дослідження операцій», «Обчислювальні методи», «Системний аналіз і теорія прийняття рішень», «Теорія програмування», «Аналіз даних і прикладні пакети статистичної обробки інформації», «Дипломне проектування Атестація», «Теорія інформації і кодування».

НАВЧАЛЬНА ПРОГРАМА ДИСЦИПЛІНИ

Модуль 3. Основи математичної статистики, оцінювання параметрів генеральної сукупності

Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики

Первинна обробка вибірок. Інтервальний варіаційний ряд. Вибіркові характеристики статистичних рядів.

Тема 4. Оцінки невідомих параметрів генеральних сукупностей

Оцінки невідомих параметрів генеральних сукупностей. Інтервали надійності математичного сподівання, дисперсії, середньоквадратичного відхилення. Мінімальний об’єм вибірки.

Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналізи

Тема 5. Перевірка статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей

Основні поняття про статистичну перевірку гіпотез. Перевірка гіпотез про рівність параметрів нормальних генеральних сукупностей із гіпотетичним значенням Перевірка гіпотез про рівність невідомих параметрів двох нормальних генеральних сукупностей. Критерій згоди Пірсона.

Тема 6. Елементи факторного аналізу

Факторний аналіз. Однакова кількість спостережень на всіх рівнях фактору.

Тема 7. Кореляційний аналіз

Регресійний аналіз: парні регресії. Метод найменших квадратів для множинної лінійної регресії. Класифікація регресій. Аналіз лінійних і нелінійних регресій. Кореляційний аналіз.

ТЕМАТИЧНИЙ ПЛАН ДИСЦИПЛІНИ

Назва теми	**Кількість годин за видами занять**					
	разом	**аудиторні заняття**				**самостійна робота**
		лекції	**семінарські**	**практичні**	**лабораторні**	
Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності						
Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики	25	4	–	6	–	15
Тема 4. Оцінки невідомих параметрів генеральних сукупностей	28	2	–	6	–	20
Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналіз						
Тема 5. Перевірка статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей	22	4	–	4	–	14
Тема 6. Елементи факторного аналізу	22	4	–	4	–	14
Тема 7. Кореляційний аналіз	23	4	–	4	–	15
Усього	**120**	**18**	**–**	**24**	**–**	**78**

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ЩОДО ВИВЧЕННЯ ДИСЦИПЛІНИ

Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності

Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики

У процесі вивчення теми слід ознайомитись з означеннями, способами побудови та способами обчислення таких понять: вибірка, варіаційні ряди (інтервальний і дискретний), полігон частот і полігон відносних частот, гістограма, емпірична функція розподілу, вибіркове середньоарифметичне, вибіркова дисперсія, вибіркове середньоквадратичне, вибіркові моді та медіана.

Термінологічний словник

Завдання математичної статистики. Основні поняття

Завдання математичної статистики полягають в тому, щоб на основі знання деяких властивостей підмножиних елементів, що взяті з деякої множини, зробити який-небудь висновок щодо властивостей цієї множини. *Перше завдання:* указати способи збору і групування статистичних даних. *Друге завдання:* розробити методи аналізування статистичних даних.

Означення 1. *Вибіркова сукупність, або просто* ***вибірка*** *– це сукупність об'єктів, які випадково відібрані з усієї сукупності об'єктів, яка називається* ***генеральною сукупністю****.*

Обсяг сукупності – це кількість об'єктів цієї сукупності.

Повторною називають вибірку, за якої відібраний об'єкт (перед вибором наступного) повертається в генеральну сукупність. ***Безповторною*** називають вибірку, за якої відібраний об'єкт у генеральну сукупність не повертається.

Вибірка повинна правильно показувати пропорції генеральної сукупності, тобто повинна бути ***репрезентативною*** (***представницькою***).

Статистичний розподіл вибірки

Емпірична функція розподілу. Полігон і гістограма

Нехай із генеральної сукупності зроблено вибірку, причому x_1 спостерігалося n_1 разів, x_2 спостерігалося n_2 разів, …, x_k спостерігалося n_k разів і $\sum n_i = n$ – обсяг вибірки.

Значення x_i називають ***варіантами,*** а числа n_i – ***частотами,*** $n_i/n = w_i$ – ***відносними частотами.***

Статистичним розподілом вибірки називають перелік варіант і відповідних частот (табл. 1), або відносних частот (табл. 2).

Таблиця 1 – Статистичний розподіл частот

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

$$\sum n_i = n$$

Таблиця 2 – Статистичний розподіл відносних частот

x_i	x_1	x_2	…	x_k
w_i	w_1	w_2	…	w_k

$$\sum w_i = 1$$

Нехай відомо статистичний розподіл частот кількісної ознаки *X*, n_x – кількість спостережень, за яких спостерігалися значення ознаки менші за *x*.

Відносна частота події $X < x$ дорівнює n_x/n. Якщо *x* змінюється, то змінюється і відносна частота, тобто n_x/n є функцією від *x*.

Емпіричною функцією розподілу (функцією розподілу вибірки) називають функцію $F^*(x)$, яка визначає для кожного значення *x* відносну частоту події $X < x$, отже, $F^*(x) = n_x/n$.

Полігоном частот називають ламану, відрізки якої з’єднують точки $(x_1, n_1), \ldots, (x_k, n_k)$.

Полігоном відносних частот називають ламану, відрізки якої з’єднують точки $(x_i, w_i), i = 1, 2, \ldots, k$.

У випадку, коли ознака є неперервною, будують *гістограму*.

Гістограмою частот називають ступінчату фігуру, що складається з прямокутників, основами яких є часткові інтервали довжиною *h*, а висоти рівні відношенню n_i/h (n_i/h – ***щільність частоти***). Площа гістограми дорівнює сумі всіх частот, тобто обсягу вибірки.

Для ***гістограми відносних частот*** висоти прямокутників дорівнюють w_i/h (w_i/h – ***щільність відносної частоти***). Площа гістограми відносних частот дорівнює одиниці.

Генеральна і вибіркові середні. Генеральна і вибіркова дисперсії

Означення 2. ***Генеральною середньою*** $\overline{X_{\Gamma}}$ *називають середнє арифметичне значення генеральної сукупності.*

Якщо значення $X_1, X_2, ..., X_k$ мають відповідно частоти $N_1, N_2, ..., N_k$, то $\overline{X_{\Gamma}} = \dfrac{X_1 N_1 + ... + X_k N_k}{N}$ або $\overline{X_{\Gamma}} = \dfrac{\sum X_i N_i}{N}$, де $N = \sum N_i$. Нехай із генеральної сукупності зроблено вибірку обсягу n, тоді ***вибірковою середньою*** $\overline{X_B}$ називають середнє арифметичне значень ознаки вибіркової сукупності $\overline{X_B} = \dfrac{\sum x_i n_i}{n}$.

Означення 3. ***Генеральною дисперсією*** D_{Γ} *називають середнє арифметичне квадратів відхилень значень ознаки генеральної сукупності від їх середнього значення* $\overline{X_{\Gamma}}$ $D_{\Gamma} = \dfrac{\sum \left(x_i - \overline{X_{\Gamma}}\right)^2 n_i}{n}$ або $D_{\Gamma} = \overline{X}^2 - \left(\overline{X_{\Gamma}}\right)^2$, де $\overline{X^2}$ – середнє квадратів по вибірці або генеральній сукупності відповідно.

Означення 4. ***Вибірковою дисперсією*** D_B *називають середнє арифметичне квадратів відхилень значень ознаки вибіркової сукупності від їх середнього значення* $\overline{X_B}$

$$D_B = \frac{\sum \left(x_i - \overline{X_B}\right)^2 n_i}{n} \text{ або } D_B = \overline{X}^2 - \left(\overline{X_B}\right)^2.$$

Вибіркове середнє квадратичне відхилення $\sigma_B = \sqrt{D_B}$.

Генеральне середнє квадратичне відхилення: $\sigma_{\Gamma} = \sqrt{D_{\Gamma}}$.

Модою вибірки називають варіанту з максимальною частотою (мода може бути не одна). Позначення M_0.

Приклад 1. У таблиці 3 мода – це 6 ($n_i = 7$ – максимальна частота), $M_0 = 6.$

Таблиця 3 – ДВР до прикладу 1

x_i	1	5	6	8	9
n_i	2	4	7	3	1

Медіана вибірки – серединний з урахуванням частот елемент вибірки, яка упорядкована за неспаданням варіант. Позначення M_e.

$$M_e = \begin{cases} x_{\frac{(n+1)}{2}}, \text{ коли } n = 2k - 1 \left(\text{непарна кількість варіант}\right), \\ \dfrac{x_{\frac{n}{2}+1} + x_{\frac{n}{2}}}{2}, \text{ коли } n = 2k. \end{cases}$$

У попередньому прикладі (табл. 3) $M_e = 6$ ($n = \sum n_i = 17$ $n + 1 = 18,$ $\dfrac{n+1}{2} = 9,$ $x_9 = 6$).

Приклад 2. Нехай маємо вибірку

$$\{x_i\} = \{1,1,5,5,5,5,5,5,6,6,8,8,8,9,10,10,10,10,10,10\}.$$

Таблиця 4 – ДВР до прикладу 2

x_i	1	5	6	8	9	10
n_i	2	6	2	3	1	6

У таблиці 4 $n = 20,$ $n/2 = 10,$ $(n/2) + 1 = 11,$ $x_{10} = 6,$ $x_{11} = 8,$ $\dfrac{x_{10} + x_{11}}{2} = \dfrac{6+8}{2} = 7.$ $M_e = 7.$

Тема 4. Оцінки невідомих параметрів генеральних сукупностей

У процесі вивчення теми слід ознайомитись з означеннями таких понять як точкові та інтервальні оцінки, способами побудови надійних інтервалів і способами обчислення мінімальних обсягів вибірки

Точковою називають оцінку, що визначається одним числом. Якщо вибірка малого обсягу, то точкова оцінка може *дуже відрізнятися* від параметра, що оцінюється, тобто призводить до грубих помилок. По цій причині за невеликого обсягу вибірки слід користуватися ***інтервальними*** оцінками, тобто оцінками, що визначаються двома числами – кінцями інтервалу.

Нехай є вибірка, по якій оцінюється невідомий параметр Θ. Позначимо через Θ^* його оцінку. Якщо $\left|\Theta-\Theta^*\right|<\delta,\ \delta>0,$ то δ *характеризує точність оцінки*.

Надійністю (надійною ймовірністю) оцінки Θ за Θ^* називають ймовірність $\gamma,$ з якою виконується нерівність $\left|\Theta-\Theta^*\right|<\delta,$ тобто $P\left(\left|\Theta-\Theta^*\right|<\delta\right)=\gamma.$ Зазвичай γ задають близьким до одиниці (0,95; 0,99; 0,999).

Інтервал $\left(\Theta^*-\delta;\ \Theta^*+\delta\right)$ називають ***надійним.*** Це інтервал, який із заданою ймовірністю γ покриває невідомий параметр Θ. Величина $\alpha=1-\gamma$ називається ***рівнем значущості***.

Надійний інтервал для оцінки математичного сподівання нормального розподілу при відомому σ

Нехай кількісна ознака X генеральної сукупності розподілена нормально, причому ***середнє квадратичне відхилення σ цього розподілу відомо***. Потрібно оцінити невідоме математичне сподівання a за вибірковою середньою $\overline{x}.$ Слід знайти надійні інтервали, що покривають параметр a з надійністю $\gamma.$

Надійний інтервал $\left(\overline{x}-t\frac{\sigma}{\sqrt{n}};\ \overline{x}+t\frac{\sigma}{\sqrt{n}}\right)$ покриває невідомий параметр a; точність оцінки $\delta=t\frac{\sigma}{\sqrt{n}}.$ t шукаємо із $2\Phi(t)=\gamma.$

Зауваження. Якщо потрібно оцінити математичне сподівання з наперед заданою точністю δ і надійністю $\gamma,$ то мінімальний обсяг вибірки, який забезпечує цю точність, знаходять за формулою (σ відоме): $n=\left[\frac{t^2\sigma^2}{\delta^2}\right],$ тут $[a]$ – найближче ціле, не менше за a.

Зсуненою оцінкою називають точкову оцінку, математичне сподівання якої не дорівнює параметру, що оцінюється.

Незсуненою оцінкою називають точкову оцінку, математичне сподівання якої дорівнює параметру, що оцінюється, при будь-якому обсязі вибірки.

Можна довести, що незсуненою оцінкою генеральної середньої є ***вибіркова середня***.

Зсуненою оцінкою генеральної дисперсії служить *вибіркова дисперсія*. Можна показати, що $M\left(D_B\right)=\frac{n-1}{n}D_Г$.

Незсуненою оцінкою генеральної дисперсії є *виправлена вибіркова дисперсія*, яку часто позначають s^2 і обчислюють так:

$$s^2=\frac{n}{n-1}D_B=\frac{\sum n_i\left(X_i-\overline{X}_B\right)^2}{n-1}.$$

Надійний інтервал для оцінки математичного сподівання нормального розподілу при невідомому σ

По даним вибірки можна побудувати випадкову величину (її можливі значення будемо позначати через t) $T=\frac{\overline{X}-a}{S/\sqrt{n}}$, яка має розподіл Ст'юдента з $k=n-1$ ступенями вільності, тут $\overline{X}$ – вибіркова середня, S – «виправлене» середнє квадратичне відхилення.

Використовуючи розподіл Ст'юдента, знайдемо надійний інтервал $\left(\bar{x}-t_\gamma\frac{S}{\sqrt{n}};\bar{x}+t_\gamma\frac{S}{\sqrt{n}}\right)$, що покриває невідоме a з надійністю γ. Тут величини $\overline{X}$ і S замінено невипадковими величинами $\bar{x}$ та s, що знайдені по вибірці, t_γ знаходять з таблиці критичних точок критерію Ст'юдента для двобічної області з $n-1$ ступенями вільності.

Пошук мінімального обсягу вибірки з нормальної сукупності для оцінки математичного сподівання (при невідомій дисперсії генеральної сукупності)

Формула має вигляд: $n_{min}=\left[\frac{t_\gamma s^2}{\delta^2}\right]$, де s^2 – виправлена вибіркова дисперсія, а t_γ шукається за таблицями розподілу Ст'юдента з $n-1$ ступенями вільності (для двосторонньої критичної області).

Надійний інтервал для дисперсії нормальної сукупності

Можна довести справедливість таких формул:

$$\frac{n-1}{u_2}s^2 \le D \le \frac{n-1}{u_1}s^2; \quad \sqrt{\frac{n-1}{u_2}}s \le \sigma \le \sqrt{\frac{n-1}{u_1}}s,$$

де s^2 – виправлена вибіркова дисперсія, u_1, u_2 визначаються за таблицями критичних точок розподілу Пірсона (χ^2):

$$u_1 = \chi^2_{\text{кр}}\left(1-\frac{\alpha}{2}; n-1\right); \quad u_2 = \chi^2_{\text{кр}}\left(1-\frac{\alpha}{2}; n-1\right).$$

План практичних занять

Зміст практичного заняття	Обсяг годин
Практичне заняття 13–14. Первинна обробка вибірок: варіаційні ряди	4
Практичне заняття 15. Полігон відносних частот, гістограма. Емпірична функція розподілу	2
Практичне заняття 16. Основні вибіркові характеристики вибірок. Оцінки невідомих параметрів генеральних сукупностей. Точкові й інтервальні оцінки	2
Практичне заняття 17. Інтервали надійності математичного сподівання, дисперсії, середньоквадратичного відхилення. Мінімальний об’єм вибірки	2
Практичне заняття 18. Підсумкове заняття на тему «Основи математичної статистики, оцінки параметрів генеральної сукупності». Модульна контрольна робота 3	2

Практичне заняття 13–14. Первинна обробка вибірок: варіаційні ряди

Завдання 1. Фірма має три торгівельні точки, вибіркові значення прибутків по яким представлені у вибірках A_1, A_2, A_3 Знайти та порівняти середні, максимальні прибутки та стабільність роботи по кожному магазину та фірми загалом, провести однофакторний дисперсійний аналіз, виявивши, наскільки суттєво впливає на результати фактор, покладений в основу групування (розбиття по магазинах), зробивши аналіз результатів, побудувавши

1) за вибірками A_1, A_2, A_3 дискретні варіаційні ряди;

2) за вибіркою *A*, яка є об'єднанням вибірок A_1, A_2, A_3 і яка є вибіркою із генеральної сукупності *X* – прибутку по фірмі загалом, – інтервальний варіаційний ряд ($b_i - a_i = k$ – задано).

Рекомендації. Вибірки до завдання 1 студент отримує на занятті від викладача.

Завдання 2. За вибіркою A_1 (табл. 5) побудувати дискретний варіаційний ряд.

Таблиця 5 – Вибірка до завдання 2

A_1	151	152	151	153	152	154	155	152	156	158

Рекомендації. ***Дискретний варіаційний ряд*** – це упорядкована послідовність пар «варіанта – частота», розташованих у порядку зростання варіант, де n_i – частота (кількість повторень) варіанти x_i. Спочатку знаходимо мінімальний та максимальний елемент вибірки: $x_{min} = 151$, $x_{max} = 158$. Далі упорядковуємо по зростанню варіанти і записуємо відповідні їм частоти (табл. 6). Для знаходження частот засобами Excel використовуйте вбудовану функцію «ЧАСТОТА» (або «FREQUENCY»), де перший масив – це вибірка, а другий – масив послідовних значень, записаних у порядку зростання від мінімального до максимального елементів вибірки. У результаті цього можуть бути отримані нульові частоти, які слід приховати за допомогою фільтра. Для роботи з масивами використовуйте комбінацію клавіш <Ctrl+Shift+Enter>.

Таблиця 6 – Дискретний варіаційний ряд до завдання 2

i	1	2	3	4	5	6	7	
x_i	151	152	153	154	155	156	158	$\sum$
n_i	2	3	1	1	1	1	1	10

Завдання 3. За вибіркою *A*, яка є об'єднанням заданих вибірок A_1, A_2, A_3 (табл. 7) побудувати інтервальний варіаційний ряд (довжина інтервалу $h = 2$).

Таблиця 7 – Вибірка до завдання 3

A_1	151	152	151	153	152	154	155	152	156	158
A_2	155	156	155	157	152	153	154	155	157	158
A_3	152	154	156	158	160	154	151	154	155	154

Рекомендації. ***Інтервальний варіаційний ряд*** – це упорядкована послідовність пар «інтервал – частота» $(a_i, b_i]$, розташованих у порядку зростання меж інтервалів, де a_i та b_i – відповідно ліва (верхня) та права (нижня) межі *i*-го інтервалу.

x_i – середина інтервалу. Нижня межа першого інтервалу знаходять за формулою $a_1 = x_{min} - h/2$, верхню межу першого інтервалу – $b_1 = a_1 + h$. Нижню межу другого інтервалу збігається із верхньою межею першого інтервалу, тобто $a_2 = b_1$. Далі аналогічно знаходять інші інтервали. Знайдемо інтервали (табл. 8). $a_1 = 151 - 2/2 = 150$, $b_1 = 150 + 2 = 152$, $a_2 = b_1 = 152$, $b_2 = 152 + 2 = 154$ і т. д. Упорядкуємо варіанти вибірки *A*:

151	151	151	152	152	152	152	152	153	153
154	154	154	154	154	154	155	155	155	155
155	156	156	156	157	157	158	158	158	160

До першого інтервалу належать елементи 151 та 152, їх частоти відповідно рівні 3 та 5, тому $n_1 = 8$. До другого належать 153 та 154 (частоти відповідно – 2 та 6), тому $n_2 = 2 + 6 = 8$, і т. д.

Для знаходження частот засобами Excel також можна використовувати вбудовану функцію «ЧАСТОТА» (або «FREQUENCY»). Але тут перший масив – це вибірка, а другий – масив значень b_i. Якщо будуть отримані інтервали з нульовими частотами, їх слід приховати за допомогою фільтра.

Таблиця 8 – Інтервальний варіаційний ряд до завдання 2

i	1	2	3	4	5	
a_i	150	152	154	156	158	
b_i	152	154	156	158	160	
x_i	151	153	155	157	159	$\sum$
n_i	8	8	8	5	1	30

Практичне заняття 15. Полігон відносних частот, гістограма. Емпірична функція розподілу

Завдання 1. За умовою завдання 1 із практичного заняття 13–14:

− обчислити відносні та накопичені частоти для вибірок A, A_1;

− скласти емпіричні функції розподілу для A, A_1 та побудувати їх графіки;

– побудувати полігон частот для A_1 та гістограму для A.

Завдання 2. За умовою завдання 2 із практичного заняття 13–14:

– обчислити відносні та накопичені частоти для вибірок A, A_1;

– скласти емпіричні функції розподілу для A, A_1 та побудувати їх графіки;

– побудувати полігон частот для A_1 та гістограму для A.

Рекомендації. Доповнимо необхідними значеннями таблицю із завдання 2 практичного заняття 13–14 (табл. 10). Тут n_i/n – відносні частоти, n_x – накопичені частоти, n_x/n – відносні накопичені частоти, n – обсяг вибірки, $n=10$.

Таблиця 10 – Обчислення частот для вибірки A_1 до завдання 2

i	x_i	n_i	n_i/n	n_x	n_x/n
1	151	2	0,2	2	0,2
2	152	3	0,3	5	0,5
3	153	1	0,1	6	0,6
4	154	1	0,1	7	0,7
5	155	1	0,1	8	0,8
6	156	1	0,1	9	0,9
7	158	1	0,1	10	1
$\sum$		**10**	**1**		

Емпірична функція по вибірці A_1 має вигляд:

$$F(x)=\begin{cases} 0; & x\le 151, \\ 0{,}2; & 151<x\le 152, \\ 0{,}5; & 152<x\le 153 \\ 0{,}6; & 153<x\le 154 \\ 0{,}7; & 154<x\le 155 \\ 0{,}8; & 155<x\le 156 \\ 0{,}9; & 156<x\le 158 \\ 1; & x>158 \end{cases}$$

Для її побудови ми використовували стовпець n_x/n із таблиці 10. Для побудови емпіричної функції по вибірці A аналогічно A_1 доповнюємо таблицю інтервального ряду значеннями частот n_x/n.

Графіки емпіричних функцій для вибірок A, A_1, полігон і гістограма представлені на рис. 1–4.

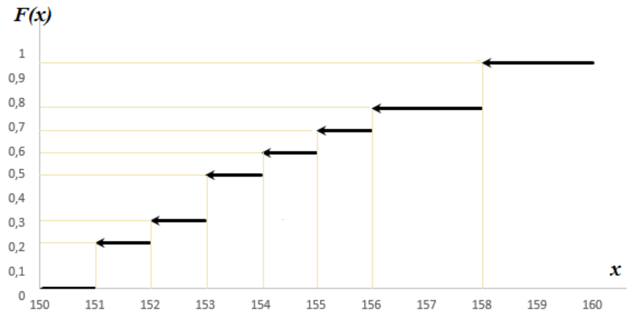

Рисунок 1 – Графік емпіричної функції по вибірці A_1

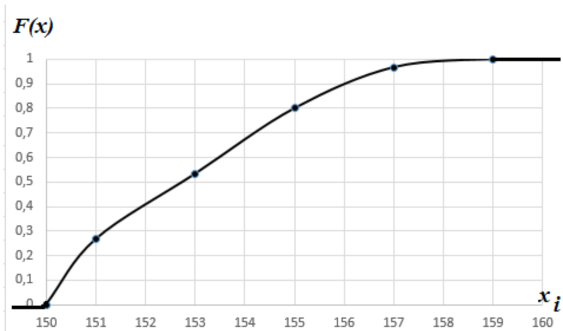

Рисунок 2 – Графік емпіричної функції по вибірці A

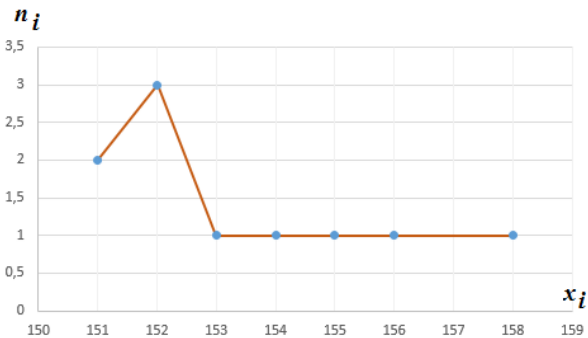

Рисунок 3 – Полігон частот по вибірці A_1

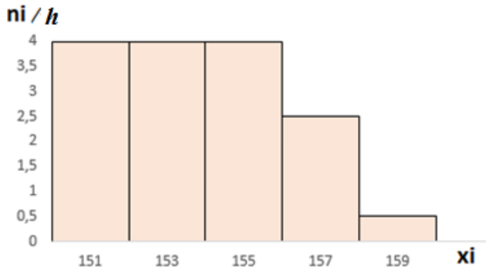

Рисунок 4 – Гістограма по вибірці *A*

Практичне заняття 16. Основні вибіркові характеристики вибірок. Оцінки невідомих параметрів генеральних сукупностей. Точкові і інтервальні оцінки

Завдання 1. За умовою завдання 1 із практичного заняття 13–14:

– обчислити вибіркові характеристики варіаційних рядів (для A, A_1, A_2, A_3):

– вибіркове середнє арифметичне $\overline{x_B}$;

– моду M_o;

– медіану M_e;

– вибіркову дисперсію D_B;

– вибіркове середньоквадратичне σ_B;

– обчислити незсунені оцінки математичного сподівання, дисперсії і середньоквадратичного генеральних сукупностей для A, A_1, A_2, A_3.

Рекомендації. Обчислення виконуємо за формулами згідно з лекцією.

Практичне заняття 17. Інтервали надійності математичного сподівання, дисперсії, середньоквадратичного відхилення. Мінімальний об'єм вибірки

Завдання 1. Випадкова величина X має нормальний розподіл з відомим $\sigma = 4$. Знайти надійний інтервал для оцінки невідомого математичного сподівання a по вибірковому середньому $\overline{x} = 15{,}6$, якщо $n = 49$, а $\gamma = 0{,}99$.

Завдання 2. Кількісна ознака X генеральної сукупності розподілена нормально. По вибірці обсягу $n=16$ знайдені вибіркова середня $\overline{x}=22{,}4$ і «виправлене» середнє квадратичне відхилення $s=2$. Оцінити невідоме математичне сподівання за допомогою інтервалу з надійністю $0{,}99$.

Завдання 3. Побудувати надійний інтервал для оцінки з надійністю γ невідомого математичного сподівання a нормально розподіленої генеральної сукупності X, якщо відомі середнє квадратичне відхилення σ, вибіркове середнє $\overline{x}$ і об'єм вибірки n.

№	σ	$\overline{x}$	n	γ	№	σ	$\overline{x}$	n	γ
1	3	10,2	36	0,95	6	3,5	18,2	36	0,95
2	4	11,4	64	0,99	7	3	12,4	64	0,99
3	4,5	15,6	100	0,99	8	4,5	11,6	81	0,999
4	5	13,2	64	0,95	9	6	19,4	100	0,95
5	5,5	11	144	0,999	10	5	18,6	81	0,95

Завдання 4. Побудувати надійний інтервал для оцінки з надійністю γ невідомого середнього квадратичного відхилення σ нормально розподіленої генеральної сукупності X, якщо відомі вибіркове середнє квадратичне відхилення s і об'єм вибірки n.

Завдання 5. Побудувати надійний інтервал для оцінки з надійністю невідомого математичного сподівання а нормально розподіленої генеральної сукупності X, якщо відомі вибіркове середнє х, вибіркове середнє квадратичне відхилення s і об'єм вибірки n.

№	s	$\overline{x}$	n	γ	№	s	$\overline{x}$	n	γ
1	4	10,2	9	0,95	6	4,5	18,2	25	0,95
2	4	11,4	16	0,99	7	2	12,4	9	0,99
3	3	15,6	25	0,99	8	3,5	11,6	16	0,999
4	5	13,2	9	0,95	9	5	19,4	25	0,95
5	2,5	11	16	0,999	10	4	18,6	9	0,95

Завдання 6. Випадкова величина X має нормальний розподіл. Знайти мінімальний об'єм вибірки для оцінки математичного сподівання з відомим $\sigma=2$, якщо надійність $\gamma=0{,}993$, а $n=49$.

Завдання 7. Випадкова величина X має нормальний розподіл. Знайти мінімальний об'єм вибірки для оцінки математичного сподівання (при невідомій дисперсії генеральної сукупності), якщо надійність $\gamma=0{,}995$, а $n=81$ та $s=0{,}85$.

Завдання 8. Знайти надійний інтервал для дисперсії нормальної сукупності, якщо відомо: $\gamma = 0{,}99,\ n = 31,\ s = 0{,}87.$

Завдання 9. Випадкова величина X має нормальний розподіл з відомим $\sigma = 2.$ Знайти надійний інтервал для оцінки невідомого математичного сподівання a по вибірковому середньому $\bar{x} = 4{,}6,$ якщо $n = 49,$ а $\gamma = 0{,}993.$

Завдання 10. Кількісна ознака X генеральної сукупності розподілена нормально. По вибірці обсягу $n = 81$ знайдені вибіркова середня $\bar{x} = 23{,}6$ і «виправлене» середнє квадратичне відхилення $s = 0{,}85.$ Оцінити невідоме математичне сподівання за допомогою інтервалу з надійністю.

Питання та інформаційні джерела для самостійного вивчення тем модуля

Назва теми, з якої виносяться питання на самостійне опрацювання	Перелік питань, що вивчаються студентом самостійно	Інформаційне джерело (порядковий номер за переліком)
Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності		
Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики	Дискретний варіаційний ряд	[1–7]
	Полігон частот та полігон відносних частот	[1–7]
	Інтервальний варіаційний ряд	[1–7]
	Емпірична функція розподілу	[1–7]
	Вибіркове середньо арифметичне	[1–8]
	Вибіркова дисперсія	[1–8]
	Вибіркове середньо квадратичне	[1–8]
	Вибіркові мода та медіана	[1–8]
Тема 4. Оцінки невідомих параметрів генеральних сукупностей	Точкові й інтервальні оцінки невідомих параметрів	[1–8]
	Незсунені, обґрунтовані, ефективні оцінки невідомих параметрів	[1–7]
	Методи знаходження точкових оцінок	[1–7]
	Інтервали надійності математичного сподівання, дисперсії, середньоквадратичного відхилення. Мінімальний об’єм вибірки	[1–7]

Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналіз

Тема 5. Перевірка статистичних гіпотез щодо невідомих параметрів та закони розподілу генеральних сукупностей

У процесі вивчення теми слід ознайомитись з означеннями статистичних гіпотез, основними відомостями щодо перевірки статистичних гіпотез, рівень значущості, критичну область, область прийняття гіпотези, розглянути способи перевірки статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей

Термінологічний словник

Статистичною називають гіпотезу про вигляд невідомого розподілу, або про параметри відомих розподілів. Поряд із гіпотезою, що висунуто, розглядають і гіпотезу, що їй суперечить. Якщо висунена гіпотеза буде відкинута, то має місце гіпотеза, що їй суперечить. Тому ці гіпотези потрібно розрізняти.

Нульовою (основною) називають гіпотезу, що висунута. Позначають її H_0. ***Конкуруючою (альтернативною)*** називають гіпотезу H_1, яка суперечить основній.

Простою називають гіпотезу, що містить тільки одне припущення. ***Складною*** називають гіпотезу, яка складається із скінченої або нескінченної сукупності простих гіпотез.

Помилка першого роду полягає в тому, що буде відхилена правильна гіпотеза. *Помилка другого роду* полягає в тому, що буде прийнята неправильна гіпотеза.

Зауваження 1. Ймовірність зробити помилку першого роду прийнято позначати α, її називають ***рівнем значущості.*** Найчастіше рівень значущості приймають рівним 0,05 або 0,01. Рівень значущості рівний 0,05 означає, що в п'яти випадках зі ста ми ризикуємо допустити помилку першого роду (відхилити справедливу гіпотезу).

Задля перевірки нульової гіпотези використовують спеціально підібрану випадкову величину, точний або наближений розподіл якої відомий. Цю величину позначають: *U* або *Z,* якщо вона розподілена нормально; *F* або υ^2 – по закону Фішера-Снедекора; *T* – по закону

Ст'юдента; χ^2 – по закону «хі квадрат» (Пірсона) тощо. Коли вигляд розподілу значення немає, будемо позначати цю величину *K*.

Статистичним критерієм (або просто критерієм) називають випадкову величину *K*, яку використовують для перевірки нульової гіпотези.

Для перевірки гіпотези по даним вибірок обчислюють значення величин, що входять у критерій, і таким чином одержують числове значення критерію (значення критерію, яке спостерігається). *Значенням критерію, яке спостерігається* $K_{\text{спост}}$ (або емпіричним) називають значення критерію, яке обчислено за вибірками.

Після вибору певного критерію множину всіх його можливих значень розбивають на дві підмножини, які не перерізаються: одна містить значення критерію, при яких нульова гіпотеза відхиляється, а друга – при яких вона приймається.

Критичною областю називають сукупність значень критерію, при яких нульову гіпотезу відхиляють.

Областю прийняття гіпотези (областю допустимих значень) називають сукупність значень критерію, при яких гіпотезу приймають.

Основний принцип перевірки статистичних гіпотез: якщо значення критерію, яке спостерігається, належить критичній області – гіпотезу відхиляють, якщо $K_{\text{спост}}$ належить області прийняття гіпотези – гіпотезу приймають.

Оскільки критерій *K* – одновимірна випадкова величина, всі її можливі значення належать деякому інтервалу. Тому критична область і область прийняття гіпотези теж є інтервалами і, отже, існують точки, що їх відділяють.

Критичними точками (границями, межами) $K_{\text{кр}}$ називають точки, що відділяють критичну область від області прийняття гіпотези.

Розрізняють однобічну (правобічну або лівобічну) критичні області і двобічну критичну область.

Правобічною називають критичну область, що визначається нерівністю $K > K_{\text{кр}}$, $K_{\text{кр}} > 0$.

Лівобічною називають критичну область, що визначається нерівністю $K < K_{\text{кр}}$, $K_{kp} < 0$.

Однобічною називають правобічну або лівобічну області.

Двобічною називають критичну область, що визначається нерівністю $K < K_1$; $K > K_2$; $K_2 > K_1$. У частковому випадку, якщо $K_1 = -K_{\text{кр}}$, $K_2 = K_{\text{кр}}$, то $|K| > K_{\text{кр}}$.

Порівняння двох дисперсій нормальних генеральних сукупностей

Правило 1. Для того, щоб при заданому рівні значущості перевірити нульову гіпотезу $H_0 : D(X) = D(Y)$ про рівність генеральних дисперсій нормальних сукупностей при конкуруючій гіпотезі $H_1 : D(X) = D(Y)$, потрібно обчислити відношення більшої виправленої дисперсії до меншої, тобто

$$F_{\text{спост}} = S_{\text{б}}^2 / S_{\text{м}}^2$$

і за таблицею критичних точок розподілу Фішера-Снедекора за заданим рівнем значущості α і кількостями ступенів вільності k_1, k_2 (k_1 – кількість ступенів вільності більшої виправленої дисперсії) знайти критичну точку $F_{\text{кр}} = F_{\text{кр}}(\alpha, k_1, k_2)$. Якщо $F_{\text{спост}} < F_{\text{кр}}$, то немає підґрунтя відхилити нульову гіпотезу. Якщо $F_{\text{спост}} > F_{\text{кр}}$ – нульову гіпотезу відхиляють.

Правило 2. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу про рівності генеральних дисперсій нормально розподілених сукупностей при конкуруючій гіпотезі $H_1 : D(X) \neq D(Y)$, потрібно обчислити відношення більшої виправленої дисперсії до меншої, тобто $F_{\text{спост}} = S_{\text{б}}^2 / S_{\text{м}}^2$ і за таблицею критичних точок розподілу Фішера-Снедекора за рівнем значущості $\alpha/2$ (вдвічі меншим заданого) і кількостями ступенів вільності k_1, k_2 (k_1 – кількість ступенів вільності більшої дисперсії) знайти критичну точку $F_{\text{кр}}(\alpha/2, k_1, k_2) = F_{\text{кр}}$.

Якщо $F_{\text{спост}} < F_{\text{кр}}$, то немає підґрунтя відхилити нульову гіпотезу. Якщо $F_{\text{спост}} > F_{\text{кр}}$ – нульову гіпотезу відхиляють.

Перевірка гіпотези про нормальний закон розподілу генеральної сукупності. Критерій згоди Пірсона

Якщо закон розподілу генеральної сукупності не відомий, але є підстава вважати, що він має певний вигляд (назвемо його A), то перевіряють гіпотезу: генеральна сукупність розподілена за законом A. Це робиться за допомогою спеціально підібраної випадкової величини – так званого критерію згоди.

Критерієм згоди називають критерій перевірки гіпотези про передбачуваний закон невідомого розподілу.

Критерій згоди Пірсона про перевірку гіпотези про нормальний розподіл генеральної сукупності. Будемо *порівнювати емпіричні* (ті, що спостерігаються) і *теоретичні* (обчислені, вважаючи розподіл нормальним) частоти.

Нехай при рівні значущості α потрібно перевірити нульову гіпотезу: генеральна сукупність розподілена нормально.

Правило 3. Для того, щоб при заданому рівні значущості перевірити нульову гіпотезу H_0: «генеральна сукупність розподілена нормально» – потрібно спочатку обчислити теоретичні частоти, а потім – значення критерію, що спостерігається $\chi^2_{\text{спост}} = \sum_{i=1}^{s} \frac{\left(n_i - m_i\right)^2}{m_i}$ і за таблицею критичних точок розподілу χ^2, за заданим рівнем значущості α і кількістю ступенів вільності $k = s - 3$ знайти критичну точку $\chi^2_{\text{кр}}\left(\alpha; k\right)$.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{кр}}$ – немає підґрунтя відхилити нульову гіпотезу. Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{кр}}$ – нульову гіпотезу відхиляють.

Зауваження. Обсяг вибірки повинен бути достатньо великим, у всякому випадку не менше 50. Кожна група повинна мати не менше 5–8 варіант; малочисельні групи об’єднуються в одну, підсумовуючи частоти (при цьому $k = m - 3$, де m – кількість груп, що залишилися).

Правило обчислення теоретичних частот при нормальному розподілі: $a = \overline{x}, \sigma = S$, $m_i = n \cdot p_i$, $p_i = \Phi\left(\frac{b_i - a}{\sigma}\right) - \Phi\left(\frac{a_i - a}{\sigma}\right)$, де Φ – функція Лапласа, $a_1 = -\infty, b_s = +\infty$.

Порівняння математичних сподівань нормальних генеральних сукупностей з однаковими дисперсіями

Нехай параметри нормальних сукупностей $N\left(M\left(X\right)\right);\ N\left(M\left(Y\right)\right);$ $\alpha -$ рівень значущості та $H_0 : M\left(X\right) = M\left(Y\right).$

Нехай знайдено $\overline{x}, \overline{y}, S_x^2, S_y^2$ – виправлені вибіркові й середні сукупностей.

Якщо гіпотеза H_0 справедлива, то випадкова величина

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{S_x^2\left(n_x - 1\right) + S_y^2\left(n_y - 1\right)}}\sqrt{\frac{n_x n_y\left(n_x + n_y - 2\right)}{n_x + n_y}} \tag{1}$$

має розподіл Ст'юдента з $k = n_x + n_y - 2$ ступенями вільності, де n_x, n_y – обсяги вибірки з першої і другої сукупностей відповідно. За формулою (1) обчислюємо $T_{\text{спост}}$.

Критичні точки шукаємо залежно від альтернативної гіпотези H_1 (табл. 1).

Таблиця 1 – Умови прийняття H_0 (порівняння математичних сподівань нормальних генеральних сукупностей з однаковими дисперсіями)

	Вид критичної області	**Критичні точки**	**Умови прийняття H_0**
$H_1 : M\left(X\right) > M\left(Y\right)$	правобічна	$T_{\text{кр}} = T_{\text{кр}}^{\text{одн}}\left(\alpha; n_x + n_y - 2\right) =$ $= T_{\text{кр}}^{\text{дв}}\left(2\alpha; n_x + n_y - 2\right)$	$T_{\text{спост}} \le T_{\text{кр}}$
$H_1 : M\left(X\right) < M\left(Y\right)$	лівобічна	$T_{\text{кр}} = -T_{\text{кр}}^{\text{одн}}\left(\alpha; n_x + n_y - 2\right) =$ $= -T_{\text{кр}}^{\text{дв}}\left(2\alpha; n_x + n_y - 2\right)$	$T_{\text{спост}} \ge T_{\text{кр}}$
$H_1 : M\left(X\right) \ne M\left(Y\right)$	двобічна	$T_{\text{кр}} = T_{\text{кр}}^{\text{одн}}\left(\frac{\alpha}{2}; n_x + n_y - 2\right) =$ $= T_{\text{кр}}^{\text{дв}}\left(\alpha; n_x + n_y - 2\right)$	$\left\|T_{\text{спост}}\right\| \le T_{\text{кр}}$

Порівняння дисперсії нормальної сукупності з гіпотетичним значенням

Позначимо σ_0^2 – гіпотетичне значення дисперсії нормальної сукупності, α – рівень значущості.

Необхідно перевірити гіпотезу $H_0 : D(X) = \sigma_0^2$.

Знайдемо за вибіркою обсягу n величину S^2 – виправлену вибіркову дисперсію, яка є незсуненою оцінкою $D(X)$, тобто $M(S^2) = D(X)$, тоді $H_0 : M(S^2) = \sigma_0^2$.

Зазвичай, $S^2 \neq \sigma_0^2$, чи значуща ця нерівність. Якщо H_0 правильна, то випадкова величина

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{2}$$

має розподіл Пірсона з $k = n-1$ ступенями вільності, її вибирають за критерій перевірки гіпотези H_0, обчислюючи за (2) $\chi^2_{\text{спост}}$. Критичні області шукають залежно від альтернативної гіпотези (табл. 2).

Таблиця 2 – Умови прийняття H_0 (порівняння дисперсії нормальної сукупності з гіпотетичним значенням)

	Вид критичної області	**Критичні точки**	**Умови прийняття H_0**
$H_1 : D(x) > \sigma_0^2$	правобічна	$\chi^2_{\text{кр}} = \chi^2_{\text{кр}}(\alpha; n-1)$	$\chi^2_{\text{спост}} \leq \chi^2_{\text{кр}}$
$H_1 : D(x) < \sigma_0^2$	лівобічна	$\chi^2_{\text{кр}} = \chi^2_{\text{кр}}(1-\alpha; n-1)$	$\chi^2_{\text{спост}} \geq \chi^2_{\text{кр}}$
$H_1 : D(x) \neq \sigma_0^2$	двобічна	$\chi^2_{\text{кр}_1} = \chi^2_{\text{кр}}\left(1-\frac{\alpha}{2}; n-1\right)$; $\chi^2_{\text{кр}_2} = \chi^2_{\text{кр}}\left(\frac{\alpha}{2}; n-1\right)$	$\chi^2_{\text{кр}_1} \leq \chi^2_{\text{спост}} \leq \chi^2_{\text{кр}_2}$

Порівняння математичного сподівання нормальної генеральної сукупності з гіпотетичним математичним сподіванням

I. Нехай є нормальна сукупність $N(M(X), \sigma)$, ***σ відоме,*** є вибірка обсягу n, рівень значущості α.

Перевіримо гіпотезу $H_0 : M(X) = m_0$, де m_0 – гіпотетичне значення математичного сподівання. Нехай $\bar{X}$ – вибіркове середнє. Якщо H_0 справедлива, то випадкова величина

$$z = \frac{\bar{X} - m_0}{\sigma / \sqrt{n}} \qquad (3)$$

має нормальний розподіл $N(0;1)$. Її вибирають за критерій, обчислюють за (3) $z_{\text{спост}}$. Критичні точки шукають за функцією Лапласа $\Phi(x)$. (див. табл. 3).

Тут $\Phi_0(x) = \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \int\limits_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \int\limits_{0}^{x} e^{-\frac{t^2}{2}} dt = 0{,}5 + \Phi(x)$.

II. Умови ті ж, що й у випадку I, але *σ* ***невідоме.*** Знайдемо ще виправлене *S*. Якщо H_0 правильна, то випадкова величина

$$T = \frac{\bar{X} - m_0}{S / \sqrt{n}} \qquad (4)$$

має розподіл Ст'юдента з $k = n - 1$ ступенями вільності. Її вибирають за критерій. За (4) обчислюють $T_{\text{спост}}$.

Критичні точки шукаємо залежно від альтернативної гіпотези (табл. 4).

Таблиця 3 – Умови прийняття H_0 (порівняння математичного сподівання нормальної генеральної сукупності з гіпотетичним математичним сподіванням, *σ* відоме)

H_1	Вид критичної області	Критичні точки	Умови прийняття H_0
$H_1 : M(X) > m_0$	правобічна	$z_{\text{кр}} : \Phi_0(z_{\text{кр}}) = 1 - \alpha;$ $\Phi(z_{\text{кр}}) = 0{,}5 - \alpha$	$z_{\text{спост}} \le z_{\text{кр}}$

$\boldsymbol{H_1}$	**Вид критичної області**	**Критичні точки**	**Умови прийняття $\boldsymbol{H_0}$**
$H_1 : M(X) < m_0$	лівобічна	$z_{\text{кр}} : \Phi_0(-z_{\text{кр}}) = 1 - \alpha;$ $\Phi(z_{\text{кр}}) = 0{,}5 - \alpha$	$z_{\text{спост}} \geq z_{\text{кр}}$
$H_1 : M(X) \neq m_0$	двобічна	$z_{\text{кр}} : \Phi_0(z_{\text{кр}}) = 1 - \alpha/2;$ $\Phi_0(z_{\text{кр}}) = 0{,}5 - \alpha/2$	$\lvert z_{\text{спост}} \rvert \leq z_{\text{кр}}$

Таблиця 4 – Умови прийняття $\boldsymbol{H_0}$ (порівняння математичного сподівання нормальної генеральної сукупності з гіпотетичним математичним сподіванням, *σ* невідоме)

$\boldsymbol{H_1}$	**Вид критичної області**	**Критичні точки**	**Умови прийняття $\boldsymbol{H_0}$**
$H_1 : M(X) > m_0$	правобічна	$T_{\text{кр}} = T_{\text{кр}}^{\text{одн}}(\alpha; n-1) =$ $= T_{\text{кр}}^{\text{одн}}(2\alpha; n-1)$	$T_{\text{спост}} \leq T_{\text{кр}}$
$H_1 : M(X) < m_0$	лівобічна	$T_{\text{кр}} = -T_{\text{кр}}^{\text{одн}}(\alpha; n-1) =$ $= -T_{\text{кр}}^{\text{одн}}(2\alpha; n-1)$	$T_{\text{спост}} \geq T_{\text{кр}}$
$H_1 : M(X) \neq m_0$	двобічна	$T_{\text{кр}} = T_{\text{кр}}^{\text{одн}}\left(\frac{\alpha}{2}; n-1\right) =$ $= T_{\text{кр}}^{\text{дв}}(\alpha; n-1)$	$\lvert T_{\text{спост}} \rvert \leq T_{\text{кр}}$

Тема 6. Елементи факторного дисперсійного аналізу

У процесі вивчення теми слід ознайомитись з означеннями факторної та залишкової дисперсії, із основними гіпотезами, які перевіряються у факторному аналізі, розглянути перевірку гіпотези про рівність дисперсій на всіх рівнях фактору (критерій Кочрена) та гіпотези про рівність математичних сподівань на всіх рівнях фактору (критерій Фішера).

Термінологічний словник

Для того, щоб застосувати критерій Фішера-Снедекора для перевірки нульової гіпотези $H_0 : M\left(X\right)=M\left(Y\right)=M\left(Z\right)$, слід перевірити (і прийняти) гіпотезу про рівність дисперсій $H_0 : D\left(X\right)=D\left(Y\right)=D\left(Z\right)$ із заданим рівнем значущості *α*, альтернативна гіпотеза H_1: *існують дві вибірки серед* A_1, A_2, A_3, *дисперсії яких не рівні*.

Відомо, що при правильній гіпотезі H_0 випадкова величина, за якою перевіряють цю гіпотезу

$$G=\frac{S_{max}^2}{S_1^2+S_2^2+S_3^2+...+S_l^2}, \quad (5)$$

має розподіл Кочрена з $k=n-1$ ***ступенями вільності*** та з l_***вибірками,*** де S_{max}^2 – максимальна з l виправлених вибіркових дисперсій $S_1^2,...,S_l^2$ для кожної з l вибірок. За (5) знаходять спостережене значення критерію.

При вибраній альтернативній гіпотезі критична область правобічна; критичну точку шукаємо за таблицею розподілу Кочрена $G_{\text{кр}}=G_{\text{кр}}\left(\alpha;k;l\right)$. Якщо H_0 відхиляється, то виключаємо з розгляду фактор, у якого дисперсія суттєво відрізняється від двох інших і перевіряємо цю ж гіпотезу для двох факторів, що залишились. Якщо $G_{\text{спост}}<G_{\text{кр}}$, гіпотеза про рівність дисперсій підтверджується. Тоді переходимо до перевірки гіпотези ***про рівність математичних сподівань на всіх рівнях фактору*** $H_0 : M\left(X\right)=M\left(Y\right)=M\left(Z\right)$. Користуємося критерієм Фішера-Снедекора. Якщо H_0 справедлива, то випадкова величина, що є критерієм,

$$F=\frac{S_{\text{факторна}}^2}{S_{\text{залишкова}}^2}$$

має розподіл ***Фішера-Снедекора*** з $k_1=l-1;\ k_2=n\cdot l-l$ ступенями вільності, де l – ***кількість рівнів фактору***. У формулі $S_{\text{факторна}}^2$ факторна дисперсія, далі $S_{\text{факт}}^2$, $S_{\text{залишкова}}^2$ – залишкова дисперсія, далі $S_{\text{зал}}^2$.

У факторному аналізі знайдена за вибіркою A виправлена дисперсія S^2 називається ***загальною дисперсією***, далі $S^2_{\text{заг}}$ ($S^2_{\text{заг}} = S^2$ для вибірки A). Залишкова дисперсія обчислюється за формулою: $S^2_{\text{зал}} = \frac{1}{l}\left(S_1^2 + \ldots + S_l^2\right)$.

Факторна дисперсія дорівнює: $S^2_{\text{факт}} = \frac{Q_{\text{факт}}}{l-1}$.

$$Q_{\text{факт}} = Q_{\text{заг}} - Q_{\text{зал}};\; Q_{\text{заг}} = (n-1)S^2_{\text{заг}};\; Q_{\text{зал}} = (n-l)S^2_{\text{зал}}.$$

При альтернативній гіпотезі H_1: «існують вибірки, для яких математичні сподівання різні», критична область правобічна. Знайдемо $F_{\text{кр}} = F_{\text{кр}}(\alpha; l-1; n\cdot l - l)$. Якщо $F_{\text{спост}} > F_{\text{кр}}$, гіпотеза H_0 про рівність математичних сподівань $\left(M(X), M(Y), M(Z)\right)$ відхиляється. Після цього виділяється рівень фактору, на якому $\overline{X_B}$ відрізняється від інших. Висуваємо дві гіпотези.

I. $H_0: M(X) = M(Y)$ при альтернативній гіпотезі $H_1: M(X) \neq M(Y)$. Якщо $l = 3$, то H_0 перевіряють за T-критерієм, якщо $l > 3$, то – за допомогою факторного аналізу.

II. Друга гіпотеза – для перевірки того, чи той рівень фактору ми виключили (це суттєво при $l > 3$, при $l = 3$ цього можна не перевіряти). Якщо її перевірити, матимемо додаткову інформацію. $H_0: M(X) = M(Z)$; $H_1: M(X) < M(Z)$.

Однофакторний дисперсійний аналіз

Нехай генеральні сукупності $X_1, X_2, \ldots, X_p$ розподілені нормально та мають *однакову*, хоч і невідому, *дисперсію*, математичне сподівання також невідоме, але вони можуть бути різні.

Потрібно при заданому рівні значущості по вибірковим середнім перевірити нульову гіпотезу про рівність всіх математичних сподівань $H_0: M(X_1) = M(X_2) = \ldots = M(X_p), p > 2$.

Для порівняння декількох середніх користуються дисперсійним аналізом (ДА), що ґрунтується на порівнянні дисперсій. На практиці дисперсійний аналіз застосовують, щоб встановити, чи має суттєвий вплив деякий *якісний* фактор *F,* який має *р* рівнів $F_1, F_2, ..., F_p$ на величину *X,* що вивчається.

Основна ідея дисперсійного аналізу полягає в порівнянні *«факторної дисперсії»,* що породжується впливом фактору і *«залишкової дисперсії»,* обумовленої випадковими причинами.

Якщо відмінність між цими дисперсіями значуща, то фактор суттєво впливає на *X.* У цьому випадку середні величини значень, що спостерігаються на кожному рівні (групові середні) будуть відрізнятися також значуще.

Якщо вже встановлено, що фактор суттєво впливає на *X,* а потрібно з'ясувати, який з рівнів найбільше впливає, то додатково роблять попарне порівняння середніх.

Інколи дисперсійний аналіз застосовують, щоб встановити *однорідність* декількох сукупностей (дисперсії однакові за припущенням; якщо ДА покаже, що і математичні сподівання однакові, то в цьому розумінні сукупності однорідні). Однорідні сукупності можна об'єднувати в одну і тим самим одержувати про неї більш повну інформацію, а отже, більш надійні висновки.

Отже, сутність дисперсійного аналізу: перевірка нульової гіпотези про рівність групових середніх нормальних сукупностей з однаковими дисперсіями зводиться до перевірки за *F*-критерієм гіпотези про рівність факторної і залишкової дисперсій.

Тема 7. Кореляційний аналіз

У процесі вивчення теми слід ознайомитись з означеннями функціонального, стохастичного та кореляційного зв'язку, кореляційного поля і регресії, коефіцієнта кореляції та його властивості, розглянути задачі кореляції, методи побудови регресій: метод «натягнутої нитки», метод сум, метод найменших квадратів.

Термінологічний словник

Функціональна залежність – залежність, у якій кожному можливому значенню аргумента *X* поставлено у відповідність одне строго визначене значення *Y*.

Статистичною називають залежність, при якій зміна однієї з величин викликає зміну розподілу ймовірності іншої.

Статистична залежність проявляється в тому, що при зміні однієї з величин змінюється середнє значення іншої, у цьому випадку статистична залежність називається ***кореляційною.***

Нехай вивчається зв'язок між випадковою величиною Y і випадковою величиною X.

Умовним середнім $\overline{y_x}$ називають середнє арифметичне значень Y, відповідних значенню $X = x$.

Очевидно, що умовна середня є функцією ***x***, у цьому випадку кажуть, що величина Y залежить від X ***кореляційно***.

Кореляційною залежністю Y від X називають функціональну залежність умовної середньої $\overline{y_x}$ від x:

$$\overline{y_x} = f\left(x\right). \tag{1}$$

Рівняння (1) називають ***рівнянням регресії*** Y на X, функцію $f\left(x\right)$ називають ***регресією*** Y на X, а її графік – ***лінією регресії*** Y на X.

$$\overline{x_y} = \varphi\left(y\right). \tag{2}$$

Рівняння (2) – це ***рівняння регресії*** X на Y, $\varphi\left(y\right)$ ***регресія*** X на Y, графік – ***лінія регресії*** X на Y.

Перша задача теорії кореляції – знайти форму кореляційного зв'язку, тобто вид функції регресії (лінійна, квадратична, показникова тощо).

Часто регресія є лінійною. Якщо $f\left(x\right)$, $\varphi\left(y\right)$ – лінійні, то кореляцію називають *лінійною*, в іншому разі – нелінійною.

Друга задача теорії кореляції – оцінити силу кореляційного зв'язку. Сила зв'язку оцінюється за величиною розсіювання значень Y навколо умовного середнього $\overline{y_x}$. Велике розсіювання – слабка залежність Y від X, або відсутність залежності. Велика залежність (а, може, й функціональна) – мале розсіювання.

За ***методом «натягнутої нитки»*** регресію визначають так. Будують кореляційне поле. Через точки кореляційного поля проводять пряму так, щоб в обох півплощинах знаходилася приблизно однакова кількість точок. На цій прямій обирають дві точки $A\left(x_A, y_A\right)$ та

$B(x_B, y_B)$ і розв'язують систему $\begin{cases} y_A = a \cdot x_A + b \\ y_B = a \cdot x_B + b \end{cases}$, з якої знаходять параметри a і b регресії $y = ax + b$ [8].

За ***методом сум*** регресію визначають так. Поділяють вибірку на дві наближено рівні за кількістю елементів частини. Кількість k пар факторів у першій частині визначають як $k = \left[(n+1)/2 \right]$. Відповідно, друга вибірка буде містити $n - k$ пар значень. [8] Параметри регресії a і b знаходять із системи $\begin{cases} \sum\limits_{i=1}^{k} y_i = a \sum\limits_{i=1}^{k} x_i + kb \\ \sum\limits_{i=k+1}^{n} y_i = a \sum\limits_{i=k+1}^{n} x_i + (n-k)b \end{cases}$.

Згідно з ***методом найменших квадратів*** параметри регресії a і b визначають так, щоб сума квадратів відхилень між теоретичними та спостережуваними значеннями факторів була мінімальною. [8] Параметри регресії a і b за цим методом знаходять за формулами:

$$a = \frac{n \cdot \sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \cdot \sum\limits_{i=1}^{n} y_i}{n \cdot \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2}, \quad b = \frac{\sum\limits_{i=n}^{n} x_i^2 \cdot \sum\limits_{i=n}^{n} y_i - \sum\limits_{i=n}^{n} x_i \cdot \sum\limits_{i=n}^{n} x_i y_i}{n \cdot \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2}.$$

Вибірковий коефіцієнт кореляції

$$r_B = \rho_{yx} \frac{\sigma_x}{\sigma_y} = \frac{\sum n_{xy} xy - n\overline{x}\overline{y}}{n \sigma_x \sigma_y},$$

де ρ_{yx} – вибірковий коефіцієнт регресії Y на X ($Y = \rho_{yx} x + b$).

Властивості r_B ***:*** 1) $|r_B| \le 1.$ 2) Якщо $r_B = 0$ і вибіркові лінії регресії – прямі, то X та Y не пов'язані кореляційною залежністю. 3) Якщо $r_B = \pm 1,$ то значення X та Y пов'язані *лінійною функціональною залежністю*. 4) Зі зростанням абсолютної величини r_B кореляційна залежність стає більш тісною та при $|r_B| = 1$ переходить у функціональну залежність.

План практичних занять

Зміст практичного заняття	Обсяг годин
Практичне заняття 19. Перевірка статистичних гіпотез про невідомі параметри генеральних сукупностей	2
Практичне заняття 20. Перевірка статистичних гіпотез про закони розподілу генеральних сукупностей	2
Практичне заняття 21–22. Факторний аналіз. Однакова кількість спостережень на всіх рівнях фактора. Неоднакова кількість спостережень на всіх рівнях фактора	4
Практичне заняття 23. Регресійний аналіз: парні регресії. Наближені методи побудови регресій	2
Практичне заняття 24. Підсумкове заняття на тему «Перевірка статгіпотез, дисперсійний і кореляційний аналізи». Модульна контрольна робота 4	2

Практичне заняття 19. Перевірка статистичних гіпотез про невідомі параметри генеральних сукупностей

Порівняння двох дисперсій нормальних генеральних сукупностей

Завдання 1. За двома незалежними вибірками обсягу $n_1 = 12$, $n_2 = 15$, вибраним із генеральних сукупностей X та Y, знайдено виправлені вибіркові дисперсії $S_X^2 = 11{,}41$; $S_Y^2 = 6{,}52$. При рівні значущості 0,05 перевірити нульову гіпотезу $H_0 : D(X) = D(Y)$ про рівність генеральних дисперсій, якщо конкуруюча гіпотеза $H_1 : D(X) > D(Y)$.

Рекомендації. Знайдемо відношення більшої виправленої дисперсії до меншої $F_{\text{спост}} = \dfrac{11{,}41}{6{,}52} = 1{,}75$. За таблицею згідно з рівнем значущості $\alpha = 0{,}05$ і кількостями ступенів вільності $k_1 = 12 - 1 = 11$; $k_2 = 15 - 1 = 14$, знаходимо критичну точку $F_{\text{кр}}(0{,}05; 11; 14) = 2{,}57$.

Оскільки $F_{\text{спост}} < F_{\text{кр}}$, немає підґрунтя відхилити нульову гіпотезу про рівність генеральних дисперсій.

Завдання 2. За двома незалежними вибірками обсягів $n_1 = 10$, $n_2 = 18$, які вибрані з нормальних генеральних сукупностей X та Y, знайдено виправлені вибіркові дисперсії $S_X^2 = 1{,}23;\ S_Y^2 = 0{,}41.$ При рівні значущості $\alpha = 0{,}1$ перевірити нульову гіпотезу про рівність генеральних дисперсій, якщо конкуруюча гіпотеза $H_1 : D(X) \neq D(Y).$

Рекомендації. Знайдемо відношення більшої виправленої дисперсії до меншої $F_{\text{спост}} = 1{,}23/0{,}41 = 3.$ Знаходимо $\alpha/2 = 0{,}05,$ $k_1 = 10 - 1 = 9;$ $k_2 = 18 - 1 = 17$ та критичну точку $F_{\text{кр}}(0{,}05;9;17) = 2{,}50.$ Оскільки $F_{\text{спост}} = 3 > F_{\text{кр}} = 2{,}5,$ нульову гіпотезу про рівність генеральних дисперсій відхиляємо.

Порівняння математичних сподівань нормальних генеральних сукупностей з однаковими дисперсіями

Завдання 3. За двома незалежними вибірками обсягу $n_1 = 10,$ $n_2 = 8,$ вибраним із нормальних генеральних сукупностей X та Y, знайдено вибіркові середні $\overline{x} = 142{,}3$ та $\overline{y} = 145{,}3$ та виправлені вибіркові дисперсії $S_X^2 = 2{,}7;\ S_Y^2 = 3{,}2.$

При рівні значущості 0,01 перевірити нульову гіпотезу $H_0 : M(X) = M(Y),$ якщо конкуруюча гіпотеза а) $H_1 : M(X) \neq M(Y),$ б) $H_1 : M(X) > M(Y),$ в) $H_1 : M(X) < M(Y).$

Рекомендації. Попередньо слід перевірити гіпотезу про рівність дисперсій $H_0 : D(X) = D(Y),$ нехай $H_1 : D(X) > D(Y).$

Знайдемо відношення більшої виправленої дисперсії до меншої $F_{\text{спост}} = 3{,}2/2{,}7 = 1{,}19.$ За таблицею згідно з рівнем значущості $a = 0{,}01$ і кількостями ступенів вільності $k_1 = 8 - 1 = 7,$ $k_2 = 10 - 1 = 9,$ знаходимо критичну точку $F_{\text{кр}}(0{,}01;7;9) = 5{,}62.$ Оскільки $F_{\text{спост}} < F_{\text{кр}},$ немає підґрунтя відхилити нульову гіпотезу про рівність генеральних дисперсій.

Далі перевіримо $H_0 : M(X) = M(Y),$ $H_1 : M(X) \neq M(Y).$

Знайдемо спостережуване значення за формулою

$$T_{\text{спост}} = \frac{\overline{x} - \overline{y}}{\sqrt{S_x^2(n_x - 1) + S_y^2(n_y - 1)}} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}.$$

$$T_{\text{спост}} = \frac{142{,}3-145{,}3}{\sqrt{2{,}7(10-1)+3{,}2(8-1)}}\sqrt{\frac{10\cdot 8\cdot(10+8-2)}{10+8}} \approx -3{,}7.$$

$|T_{\text{спост}}| = 3{,}7.$

За таблицею розподілу Ст'юдента згідно з рівнем значущості $a = 0{,}01$ і зі ступенями вільності $k = n_x + n_y - 2 = 10+8-2 = 16$ знаходимо критичну точку

$$T_{\text{кр}}^{\text{дв}}\left(\alpha, n_x + n_y - 2\right) = T_{\text{кр}}^{\text{дв}}\left(0{,}01;16\right) = 2{,}92.$$

Оскільки $|T_{\text{спост}}| = 3{,}7 > T_{\text{кр}} = 2{,}92,$ нульову гіпотезу відхиляють.

Порівняння виправленої вибіркової дисперсії із гіпотетичним значенням генеральної дисперсії нормальної сукупності

Завдання 4. Із нормальної генеральної сукупності отримано вибірку об'ємом $n = 31$:

Варіанти x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
Частоти n_i	1	3	7	10	6	3	1

Необхідно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу $H_0 : \sigma^2 = \sigma_0^2 = 0{,}18,$ прийнявши у якості конкуруючої гіпотези $H_1 : \sigma^2 > 0{,}18.$

Рекомендації. Обчислимо вибіркову середню $\overline{X} = \left(\sum_i x_i \cdot n_i\right)\Big/ n =$ $= 11{,}08$ та дисперсію

$$D(X) = \left(\left(\sum_i x_i^2 \cdot n_i\right)\Big/ n\right) - \left(\left(\sum_i x_i \cdot n_i\right)\Big/ n\right)^2 = 123{,}11 - 122{,}85 = 0{,}26.$$

Виправлена дисперсія $S^2 = \left(n/(n-1)\right)\cdot D(X) = 0{,}26\cdot 1{,}03 = 0{,}27.$

Знайдемо спостережуване значення $\chi^2_{\text{спост}} = (n-1)\cdot s^2 / \sigma_0^2 =$ $= \left((31-1)\cdot 0{,}27\right)/0{,}18 = 45.$

Критичне значення

$$\chi^2_{\text{кр}} = \chi^2_{\text{кр}}(\alpha; n-1) = \chi^2_{\text{кр}}(0{,}05; 30) = 43{,}773. \ \ \chi^2_{\text{спост}} = 45 > \chi^2_{\text{кр}} = 43{,}773.$$

Отже, $H_0 : \sigma^2 = \sigma_0^2 = 0{,}18$ – відхиляється.

Порівняння математичного сподівання нормальної генеральної сукупності із гіпотетичним математичним сподіванням

Завдання 5. Із нормальної генеральної сукупності отримано вибірку об'ємом $n = 31$ (див. завдання 4).

Необхідно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу $H_0 : M(X) = m_0 = 10$, прийнявши у якості конкуруючої гіпотези $H_0 : M(X) > m_0$. σ – відоме, $\sigma = 32$.

Рекомендації. Використаємо обчислену в завданні 4 $\overline{X} = 11{,}08$. Спостережуване значення

$$z_{\text{спост}} = \left(\overline{X} - m_0\right) \Big/ \left(\sigma / \sqrt{n}\right) = (11{,}08 - 10) \Big/ \left(32 / \sqrt{31}\right) = 1{,}08 / 5{,}75 = 0{,}19.$$

Критичне значення шукаємо за функцією Лапласа

$$\Phi\left(z_{\text{кр}}\right) = 0{,}5 - \alpha = 0{,}5 - 0{,}05 = 0{,}45. \ \ z_{\text{кр}} = 1{,}65. \ \ z_{\text{спост}} = 0{,}19 < z_{\text{кр}} = 1{,}65.$$

Отже, $H_0 : M(X) = m_0 = 10$ приймається.

Завдання 6. Розв'язати задачу 5 за умови, що σ – невідоме.

Рекомендації. $S^2 = 0{,}27 \to S = 0{,}52.$

$$T_{\text{спост}} = \left(\overline{X} - m_0\right) \Big/ \left(S / \sqrt{n}\right) = (11{,}08 - 10) \Big/ \left(0{,}52 / \sqrt{31}\right) = 1{,}08 / 0{,}093 = 11{,}61.$$

Критичне значення шукаємо за розподілом Ст'юдента із $k = n - 1 = 30$ ступенями вільності

$$T_{\text{кр}} = T_{\text{кр}}(\alpha; n-1) = T_{\text{кр}}(0{,}05; 30) = 1{,}7. \ \ T_{\text{спост}} = 11{,}61 > T_{\text{кр}} = 1{,}7.$$

Отже, $H_0 : M(X) = m_0 = 10$ не приймається.

Практичне заняття 20. Перевірка статистичних гіпотез про закони розподілу генеральних сукупностей

Завдання 1. Використовуючи критерій Пірсона, при рівні значущості α перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності X зі статистичними даними, які подано у вигляді інтервального варіаційного ряду (в першому та другому рядках вказано кінці часткових інтервалів $a_i - b_i$, в третьому – відповідні їм частоти n_i). [1].

Рекомендації. Розглянемо розв'язання на прикладі Var_0.

Var_0 $\alpha = 0{,}1$

a_i	134	139	143	148	153	160	163
b_i	139	143	148	153	160	163	167
n_i	5	6	6	6	7	6	6

Перевіримо нульову гіпотезу H_0: «генеральна сукупність розподілена за нормальним законом». Спершу обчислимо: $X_1 = \left(a_i - \overline{X}\right)/S$, $X_2 = \left(b_i - \overline{X}\right)/S$, $p_i = \Phi\left(X_2\right) - \Phi\left(X_1\right)$, $m_i = n \cdot p_i$, де $n = 42$, $\overline{X} = 152{,}74$; $S = 11{,}1$. $\Phi\left(X_1\right)$, $\Phi\left(X_2\right)$ знайдемо за таблицею значень інтегральної функції Лапласа. Занесемо отримані значення в таблицю 5. Обчислимо останні три стовпці таблиці 5.

Знайдемо кількість ступенів вільності: $k = m - 3 = 7 - 3 = 4$, де $m = 7$ – кількість інтервалів; за таблицею критичних точок розподілу χ^2 з рівнем значущості $\alpha = 0{,}1$ і кількістю ступенів вільності $k = 4$ знаходимо $\chi^2_{\text{кр}}\left(0{,}1;4\right) = 7{,}779$. Оскільки $\chi^2_{\text{спост}} = 5{,}487 < \chi^2_{\text{кр}} = 7{,}779$ – немає підґрунтя відхилити нульову гіпотезу. Отже, розбіжність емпіричних і теоретичних частот незначуща. Тому дані спостережень узгоджуються з гіпотезою про нормальний розподіл генеральної сукупності.

Таблиця 5 – Обчислення до завдання 1

i	***a_i***	***b_i***	***X*₁**	***X*₂**	**$\Phi\left(X_1\right)$**	**$\Phi\left(X_2\right)$**	***p_i***
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	139	–13,76	–1,24	–0,5	–0,3925	0,108
2	139	143	–1,24	–0,88	–0,3925	–0,3106	0,082
3	143	148	–0,88	–0,43	–0,3106	–0,1664	0,144
4	148	153	–0,43	0,02	–0,1664	0,0080	0,174
5	153	160	0,02	0,65	0,0080	0,2422	0,234
6	160	163	0,65	0,92	0,2422	0,3212	0,079
7	163	1000	0,92	76,37	0,3212	0,5	0,179
Σ							

i	…	n_i	m_i	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
9	**10**	**11**	**12**	**13**	**14**	**15**
1	...	5	4,52	0,485	0,235	0,052
2	...	6	3,44	2,560	6,555	1,906
3	...	6	6,06	–0,056	0,003	0,001
4	...	6	7,32	–1,325	1,755	0,240
5	...	7	9,84	–2,836	8,045	0,818
6	...	6	3,32	2,682	7,193	2,168
7	...	6	7,51	–1,510	2,279	0,303
Σ	...	42	42			$\chi^2_{\text{спост}} = 5{,}487$

Практичне заняття 21–22. Факторний аналіз. Однакова кількість спостережень на всіх рівнях фактора. Неоднакова кількість спостережень на всіх рівнях фактора

Завдання 1. Нехай знайдено вибіркові характеристики сукупностей A, A_1, A_2, A_3, які представлено в таблиці 6.

За вибіркою A з генеральної сукупності перевірити гіпотезу про рівність математичних сподівань на трьох рівнях фактору (фактор – розподіл по вибірках, наприклад, прибутку A_1, A_2, A_3 трьох магазинів фірми, що має їх) при рівні значущості α. Для застосування F-критерію Фішера спочатку перевірити гіпотезу про рівність дисперсій. Якщо вона не приймається, виокремити рівень фактору з відмінною від інших дисперсією, для всіх інших рівнів фактору окремо перевіряти гіпотези про рівність математичних сподівань.

Рекомендації. Приклад розв’язування наведено в лекції.

Таблиця 6 – Приклад варіанта 1 до завдання 1 (Var_1_α = 0,01)

Сукупність	n	$\overline{X}$	S^2	S
A_1	42	150,86	123,905	11,131
A_2	42	148,43	135,976	11,661
A_3	42	158,26	152,262	12,339
A	126	152,52	150,793	12,280

Завдання 2. Проведено по чотири випробування на кожному із трьох рівнів фактору F. Методом дисперсійного аналізу при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу про рівність групових середніх. Уважаємо, що вибірки вибрані із нормальних сукупностей із однаковими дисперсіями. Результати спостережень наведено в табл. 7.

Рекомендації. Спростимо обчислення, зменшивши всі спостереження на загальну середню $c = 29$. Розрахунки наведено в табл. 8.

Враховуючи, що $p = 3$; $q = 4$ обчислюємо:

$$S_{\text{заг}} = \sum_{j=1}^{p} S_j - \left[\sum_{j=1}^{p} T_j\right]^2 \Big/ pq = 312 - 0 = 312;$$

$$S_{\text{факт}} = \sum_{j=1}^{p} T_j^2 \Big/ q - \left[\sum_{j=1}^{p} T_j\right]^2 \Big/ pq = 32/4 - 0 = 8;$$

$$S_{\text{залиш}} = S_{\text{заг}} - S_{\text{факт}} = 312 - 8 = 304.$$

Таблиця 7 – Результати спостереження до завдання 2

Номер випробування i	Рівень фактору F_j		
	F_1	F_2	F_3
1	34	31	35
2	23	26	21
3	31	33	37
4	28	22	27
$\overline{x}_{гр.j}$	29	28	30

Отже, дисперсії такі: $\sigma^2_{\text{факт}} = \dfrac{S_{\text{факт}}}{p-1} = \dfrac{8}{3-1} = 4;$

$$\sigma^2_{\text{залиш}} = \frac{S_{\text{залиш}}}{p(q-1)} = \frac{304}{3(4-1)} = \frac{304}{9} = 33{,}8.$$

Обчислимо: $F_{\text{спост}} = \dfrac{\sigma^2_{\text{факт}}}{\sigma^2_{\text{зал}}} = \dfrac{4}{33{,}8} = 0{,}12.$

Таблиця 8 – Розрахунки зі спрощеннями

Номер випробу-вання i	Рівень фактору F_j						Суми
	F_j		F_j		F_3		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	
1	5	25	2	4	6	36	
2	–6	36	–3	9	–8	64	
3	2	4	4	16	8	64	
4	–1	1	–7	49	–2	4	
$S_j = \sum_{i=1}^{4} y_{ij}^2$		66		78		168	$\sum S_j = 312$
T_j	0		–4		4		$\sum T_j = 0$
T_j^2	0		16		16		$\sum T_j^2 = 32$

Кількість ступенів вільності чисельника і знаменника відповідно 2 і 9, отже, за таблицею критичних точок розподілу Фішера маємо $F_{\text{кр}}\left(0{,}05;2;9\right) = 4{,}26 > 0{,}12 = F_{\text{спост}}$. Тому нульова гіпотеза про рівність загалом (одночасно) групових середніх приймається.

Практичне заняття 23. Регресійний аналіз: парні регресії. Наближені методи побудови регресій

Завдання 1. Знайти вибіркове рівняння прямої регресії $y = \rho x + b$ за даними п'яти спостережень x_i, y_i над величинами X та Y. Побудувати рисунок, на якому вказати експериментальні дані та побудувати пряму регресії. Знайти вибірковий коефіцієнт кореляції.

(x_i, y_i)	(1; 4,9)	(2; 5,9)	(3; 4,4)	(4; 3,4)	(5; 2,9)

Завдання 2. Зроблено випадкову вибірку з генеральної сукупності працівників галузі промисловості N. Досліджувалися: X – вік; Y_1 – відстань від дому до роботи; Y_2 – стаж роботи за спеціальністю; Y_3 – річна зарплата ($\times 100$ грн) (див табл. 9).

1. Побудувати діаграми розсіяння і лінії регресії для пар змінних:
- «річна зарплата – вік» ($Y_3 - X$);
- «річна зарплата – відстань від дому» ($Y_3 - Y_1$);
- «річна зарплата – стаж роботи» ($Y_3 - Y_2$).

2. Знайти вибірковий коефіцієнт кореляції для кожної пари.

3. Дати інтерпретацію побудованих ліній регресії і відповідних коефіцієнтів кореляції.

Таблиця 9 – Спостережувані дані для завдання 2

X	**31**	**56**	**51**	**33**	**29**	**22**	**29**	**36**	**32**	**21**
Y_1	24	20	9	28	15	29	17	25	26	10
Y_2	7	20	21	10	5	1	4	11	2	2
Y_3	19	42	31	20	29	22	28	17	30	11

Завдання 3. Бюро економічного аналізу фірми «Світоч» оцінює ефективність витрат на рекламу. Для такої оцінки вони мають реальні спостереження обсягу продажу продукції в п’яти географічних зонах y_i та витрат на рекламу x_i.

i	1	2	3	4	5
y_i	25	30	35	45	65
x_i	5	6	9	12	18

Визначити залежність між (x_i, y_i) а) методом «натягнутої нитки»; б) методом сум; в) методом найменших квадратів.

Завдання 4. Досліджувалась залежність пружності Y сталевих болтів від вмісту в них нікелю X. Результати досліджень наведені в таблиці:

$X=x_i$, %	2,20	2,35	2,42	2,58	2,65	2,69	2,74	2,88	2,91
$Y=y_i$, %	35,4	35,0	35,8	36,2	36,7	36,9	37,3	37,8	38,2

Чи існує кореляційний зв’язок між величинами Y та X? Записати рівняння регресії Y на X і побудувати його графік.

Питання та інформаційні джерела для самостійного вивчення тем модуля

Назва теми, з якої виносяться питання на самостійне опрацювання	**Перелік питань, що вивчаються студентом самостійно**	**Інформаційне джерело (порядковий номер за переліком)**
Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналізи		
Тема 5. Перевірка статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей	Основні відомості про перевірку статистичних гіпотез. Рівень значущості	[1–8]
	Критичні точки. Лівостороння, правостороння і двостороння критичні області та їх пошук	[1–7]

Продовж. питань та інформаційні джерела для самостійного вивчення тем модуля

Назва теми, з якої виносяться питання на самостійне опрацювання	**Перелік питань, що вивчаються студентом самостійно**	**Інформаційне джерело (порядковий номер за переліком)**
Тема 5. Перевірка статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей	Перевірка гіпотез про рівність параметрів нормальних генеральних сукупностей гіпотетичним значенням	[1–7]
	Перевірка гіпотез про рівність невідомих параметрів двох нормальних генеральних сукупностей	[1–7]
	Перевірка гіпотези про нормальний розподіл. Критерій згоди Пірсона	[1–7]
Тема 6. Елементи факторного аналізу	Основні гіпотези, які перевіряються у факторному аналізі	[3, 6–7]
	Факторна та залишкова дисперсії	[3, 6–7]
	Критерій Кочрена.	[3, 6–7]
	Перевірка гіпотези про рівність дисперсії на всіх рівнях фактору	[3, 6–7]
	Критерій Бартлетта	[3, 6–7]
	Факторна і залишкова дисперсії при неоднаковій кількості спостережень на різних рівнях фактору	[3, 6–7]
Тема 7. Кореляційний аналіз	Функціональний, стохастичний та кореляційний зв’язки	[3, 6–7]
	Кореляційні поля і емпіричні регресії. Фактори і показники	[3, 6–7]
	Класифікація регресій. Методи переходу від нелінійних до лінійних регресій	[3, 6–7]
	Квазілінійні регресії, заміни, що приводять їх до лінійних регресій	[3, 6–7]
	Метод «натягнутої нитки»	[3, 6–7]
	Метод сум	[3, 6–7]
	Основні задачі кореляційного аналізу	[3, 6–7]

ІНДИВІДУАЛЬНІ ЗАВДАННЯ (НАВЧАЛЬНО-ДОСЛІДНІ ПРОЕКТИ) ДЛЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ

Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності

Індивідуальним завданням за третім модулем є домашні завдання до практичних занять, які видає студентам викладач.

Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналізи

Індивідуальним завданням за четвертим модулем є домашні завдання до практичних занять, які видає студентам викладач.

ПОРЯДОК І КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАНЬ СТУДЕНТІВ

Поточне оцінювання знань студентів

Метою та завданням поточного контролю є визначення рівня засвоєння матеріалу студентами для корекції їх навчальної роботи в разі потреби, а також накопичення балів рейтингу студента з дисципліни.

Засоби поточного контролю вивчення дисципліни:

– опитування на заняттях;
– перевірка підготовки до практичних занять;
– перевірка виконання домашнього завдання;
– перевірка виконання модульної контрольної роботи;
– розв’язування практичних завдань біля дошки;
– опитування у процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять.

Об’єктами поточного контролю є: відвідування занять, відповіді на заняттях, виконання домашніх завдань, виконання модульних контрольних робіт, додаткові види робіт, які вказано в таблицях нарахування балів (дод. А, Б).

Питання для підготовки до поточного модульного контролю

Назва теми	Питання
Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності	
Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики	Дискретний варіаційний ряд
	Інтервальний варіаційний ряд
	Полігон частот, гістограма
	Емпірична функція розподілу, її графік

Продовж. питань для підготовки до поточного модульного контролю

Назва теми	Питання
Тема 3. Первинна обробка вибірок: варіаційні ряди, вибіркові характеристики	Вибіркові характеристики статистичних рядів: вибіркова середня, дисперсія, середньоквадратичне відхилення
	Вибіркові характеристики статистичних рядів: мода, медіана
Тема 4. Оцінки невідомих параметрів генеральних сукупностей	Точкові й інтервальні оцінки невідомих параметрів
	Оцінки невідомих параметрів генеральних сукупностей
	Інтервали надійності математичного сподівання, дисперсії, середньоквадратичного відхилення. Мінімальний об’єм вибірки
Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналізи	
Тема 5. Перевірка статистичних гіпотез про невідомі параметри та закони розподілу генеральних сукупностей	Основні поняття про статистичну перевірку гіпотез
	Критичні точки. Лівостороння, правостороння і двостороння критичні області та їх пошук
	Перевірка гіпотез про рівність параметрів нормальних генеральних сукупностей із гіпотетичним значенням
	Перевірка гіпотез про рівність невідомих параметрів двох нормальних генеральних сукупностей
	Перевірка гіпотези про нормальний розподіл. Критерій згоди Пірсона
Тема 6. Елементи факторного аналізу	Основні гіпотези, які перевіряються у факторному аналізі
	Факторна та залишкова дисперсії. Однакова кількість спостережень на всіх рівнях фактору
	Критерій Кочрена
	Перевірка гіпотези про рівність дисперсії на всіх рівнях фактора
	Факторна та залишкова дисперсії. Неоднакова кількість спостережень на різних рівнях фактору
Тема 7. Кореляційний аналіз	Функціональний, стохастичний і кореляційний зв’язки
	Кореляційні поля й емпіричні регресії. Фактори і показники
	Метод «натягнутої нитки», метод сум, метод найменших квадратів
	Основні завдання кореляційного аналізу. Коефіцієнт кореляції, його властивості

Зразок модульної контрольної роботи

Модульна контрольна робота 3

1. Задано статистичний розподіл вибірки

x_i	2	5	7	8
n_i	1	3	2	4

Знайти вибіркове середнє, вибіркову дисперсію і середнє квадратичне відхилення.

2. Побудувати емпіричну функцію розподілу вибірки із завдання 1 та побудувати її графік.

3. Випадкова величина X має нормальний розподіл з відомим $\sigma = 2.$ Знайти надійний інтервал для оцінки невідомого математичного сподівання a по вибірковому середньому $\overline{x} = 4{,}6;$ якщо $n = 49,$ а $\gamma = 0{,}993.$

4. За двома незалежними вибірками обсягу $n_X = 9,$ $n_Y = 15,$ вибраним із генеральних сукупностей X та Y, знайдено виправлені вибіркові дисперсії $S_X^2 = 9{,}6;$ $S_Y^2 = 5{,}7.$ При рівні значущості $\alpha = 0{,}01$ перевірити нульову гіпотезу $H_0 : D(X) = D(Y)$ про рівність генеральних дисперсій, якщо конкуруюча гіпотеза $H_1 : D(X) > D(Y)$.

5. Сформулювати поняття моди та медіани дискретного варіаційного ряду.

Модульна контрольна робота 4

1. Нехай є три вибірки з нормальних сукупностей з однаковими дисперсіями. Результати спостережень наведено в таблиці. При рівні значущості $\alpha = 0{,}01$ перевірити нульову гіпотезу про рівність групових середніх, користуючись дисперсійним аналізом (спростити обчислення,зменшивши всі спостереження на $c = 25$).

	Рівні фактору		
Спостереження	F_1	F_2	F_3
1	23	25	26
2	24	22	29
3	25	24	26
4	28	21	31

2. За даними таблиці спостережень

x_i	1	2	1	3	4
y_i	6	9	7	8	7

скласти рівняння регресій Y на X та Y на X. Побудувати кореляційні поля і лінії регресій.

Критерії оцінювання знань та система нарахування балів

Виконання модульних контрольних робіт оцінюється у відсотках до максимальної суми за них. Систему нарахування балів представлено в таблицях додатків А, Б.

Підсумкове оцінювання знань студентів

Перелік питань для підготовки до іспиту (4 семестр)

1. Дискретний варіаційний ряд. Полігон частот.
2. Інтервальний варіаційний ряд. Гістограма.
3. Вибіркове середньо арифметичне.
4. Вибіркова дисперсія, вибіркове середньо квадратичне відхилення.
5. Мода та медіана.
6. Завдання математичної статистики. Поняття вибірки та генеральної сукупності. Повторна, безповторна вибірки.
7. Емпірична функція розподілу, її властивості, графік.
8. Полігон і гістограма відносних частот.
9. Оцінки невідомих параметрів. Точкові й інтервальні оцінки невідомих параметрів.
10. Надійні інтервали невідомих параметрів та їх імовірнісний зміст. Рівень значущості.
11. Незсунені оцінки математичного сподівання, дисперсії, середньоквадратичного.
12. Надійний інтервал для оцінки математичного сподівання нормального розподілу при відомому σ.
13. Надійний інтервал для оцінки математичного сподівання нормального розподілу при невідомому σ.
14. Надійний інтервал для дисперсії нормальної сукупності.
15. Мінімальний обсяг вибірки для пошуку надійного інтервалу математичного сподівання заданої довжини (при відомій і невідомій дисперсії генеральної сукупності).

16. Розподіл χ^2.
17. Розподіл студента або *t*-розподіл.
18. Розподіл *F* (Фішера-Снедекора).
19. Статистичні гіпотези. Основна й альтернативна гіпотези. Проста і складна гіпотеза.
20. Помилки першого і другого роду. Статистичний критерій перевірки нульової гіпотези. Значення критерію, яке спостерігається.
21. Критична область. Область прийняття гіпотез. Критичні точки.
22. Відшукання правобічної критичної області.
23. Відшукання лівобічної і двобічної критичних областей. Потужність критерію.
24. Порівняння двох дисперсій нормальної генеральної сукупностей. Випадок $H_0 : D(X) = D(Y);\ \ H_1 : D(X) > D(Y)$.
25. Порівняння двох дисперсій нормальної генеральної сукупностей. Випадок $H_0 : D(X) = D(Y);\ \ H_1 : D(X) \neq D(Y)$.
26. Порівняння дисперсії нормальної сукупності з гіпотетичним значенням.
27. Порівняння математичного сподівання генеральної сукупності з гіпотетичним математичним сподіванням при відомій і невідомій дисперсії.
28. Порівняння математичного сподівання нормальних генеральних сукупностей з однаковими дисперсіями.
29. Факторний аналіз. Основні гіпотези, які перевіряються у факторному аналізі.
30. Критерій Кочрена та Барлетта.
31. Кореляційний аналіз. Функціональний статистичний і кореляційний зв'язки. Умовна середня.
32. Два основних завдання теорії кореляції.
33. Метод найменших квадратів для парної регресії.
34. Кореляційний аналіз. Основні статистичні гіпотези кореляційного аналізу.
35. Кореляційні таблиці. Інтервальні та дискретні кореляційні таблиці. Емпіричні регресії.
36. Матрична форма методу найменших квадратів (МНК).
37. Загальна мультиколінеарність. Метод Фаррара-Глобера встановлення мультиколінеарності.

38. Матриці вибіркових виправлених коефіцієнтів коваріацій і кореляцій.
39. Матриця частинних коефіцієнтів коваріацій і кореляцій.
40. Перевірка гіпотези про значущість кореляції між парами факторів регресії.

Зразок екзаменаційного білету

1. Дискретний варіаційний ряд. Полігон частот.
2. Емпірична функція розподілу, її властивості, графік.
3. По вибірках *A, B* при рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про рівність дисперсій при альтернативній гіпотезі $H_1 : \sigma_1 > \sigma_2$:

A	6	3	6	6	5	9	8	5	5	3
B	3	7	1	9	2	4	4	2	2	8

4. По вибірці з генеральної сукупності побудувати гістограму, склавши інтервальний варіаційний ряд з інтервалом $h = 2$:

X	3	10	1	3	6	9	3	6	9	2	3	4

Загальна підсумкова оцінка з дисципліни

Критерії, параметри та шкала оцінювання знань студентів та відповідність оцінок в різних шкалах наведено в додатках А, Б.

Загальна підсумкова оцінка з дисципліни за семестр – це сума балів поточного контролю і балів, отриманих на екзамені.

Рейтингом студента за дисципліною є кількість балів, отриманих студентом у результаті вивчення дисципліни.

СПИСОК РЕКОМЕНДОВАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ

1. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – 5-те вид. – Київ : Центр учбової л-ри, 2010. – 424 с.
2. Грищенко В. О. Теорія ймовірностей і математична статистика: навч. посіб. / В. О. Грищенко. – Київ : Київ. торг.-екон. ун-т, 2002. – 164с.
3. Жлуктенко В. І. Теорія ймовірностей і математична статистика: навч.-метод. посіб. : у 2 ч. – Ч. II. Математична статистика / В. І. Жлуктенко, С. І. Наконечний, С. С. Савіна. – Київ : КНЕУ, 2001. – 336 с.
4. Кармелюк Г. І. Теорія ймовірностей та математична статистика. Посібник з розв’язування задач : навч. посіб. / Г. І. Кармелюк. – Київ : Центр учбової л-ри, 2007. – 576 с.
5. Кушлик-Дивульська О. І. Теорія ймовірностей та математична статистика : навч. посіб. / Н. В. Поліщук, Б. П. Орел, П. І. Штабалюк. – Київ : НТУУ «КПІ», 2014. – 212 с.
6. Медведєв М. Т. Теорія ймовірностей та математична статистика : підручник / М. Т. Медведєв, І. О. Пащенко. – Київ : Ліра-К, 2020. – 536 с.
7. Огірко О. І. Теорія ймовірностей та математична статистика : навч. посіб. / О. І. Огірко, Н. В. Галайко. – Львів : ЛьвДУВС, 2017. – 292 с.
8. Роскладка О. В. Теорія ймовірностей і математична статистика : навч. посіб. / О. В. Роскладка. – Полтава : ПУСКУ, 2007.

ДОДАТКИ

Додаток А

Таблиця відповідності результатів контролю знань за різними шкалами та критерії оцінювання з дисципліни

Оцінка за шкалою ЄКТС*	**Оцінка за баловою шкалою, що використовується в ПУЕТ**	**Оцінка за 4-баловою шкалою**
F	1–34 балів	2 – Незадовільно з обов’язковим повторним вивченням дисципліни
FX	35–59 балів	2 – Незадовільно з можливим повторним складанням іспиту
D E	60–65 балів 66–70 балів	3 – Задовільно
C B	71–78 балів 79–85 балів	4 – Добре
A	86–100 балів	5 – Відмінно

** ЄКТС – Європейська кредитна трансферно-накопичувальна система.*

Додаток Б

Система нарахування балів за видами навчальної роботи з дисципліни «Теорія ймовірностей і математична статистика»

Вид навчальної роботи	**Максимальна кількість балів**
1. Аудиторна (лекції і практичні заняття) Відвідування занять (за дистанційного навчання – тестування) (20 балів)	20
Модуль 3. Основи математичної статистики, оцінки параметрів генеральної сукупності Правильна відповідь при опитуванні (2 бали за відповідь, 3 відповіді за 3 модуль) 6 балів. 2. Модульний контроль. МКР (підсумкове тестування) (15 балів)	21
Модуль 4. Перевірка статгіпотез, дисперсійний та кореляційний аналіз Правильна відповідь при опитуванні (2 бали за відповідь, 2 відповіді за 1 модуль) 4 бали. 2. Модульний контроль. МКР (підсумкове тестування) (15 балів)	19
Поточне оцінювання	60
Екзамен	40
Разом	**100**

Дисципліна вважається зарахованою, коли студент набрав 60 і більше балів.

ЗМІСТ

Навчально-методичне видання

ЄМЕЦЬ Олег Олексійович
ПАРФЬОНОВА Тетяна Олександрівна

ТЕОРІЯ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА (Частина 2)

НАВЧАЛЬНО-МЕТОДИЧНИЙ ПОСІБНИК

Редагування *Л. М. Діденко*
Комп'ютерне верстання *О. С. Корніліч*

Формат 60x84/16. Ум. друк. арк. 3,1.
Зам. № 267/2048.

Видавець і виготовлювач
Вищий навчальний заклад Укоопспілки
«Полтавський університет економіки і торгівлі»,
к. 115, вул. Коваля, 3, м. Полтава, 36014; ☎(0532) 50-24-81

Свідоцтво про внесення до Державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції ДК № 3827 від 08.07.2010 р.