

Міністерство освіти і науки України
Національний університет водного господарства та
природокористування
Навчально-науковий інститут автоматики, кібернетики і
обчислювальної техніки
Кафедра вищої математики

04-02-54М

МЕТОДИЧНІ ВКАЗІВКИ

до виконання практичних робіт з навчальної дисципліни
«Теорія ймовірностей та математична статистика»
для здобувачів вищої освіти першого (бакалаврського)
рівня за освітньо-професійною програмою «Цифрові
технології дистанційної освіти» спеціальності
015.39 «Професійна освіта (Цифрові технології)»
денної та заочної форм навчання

Рекомендовано
науково-методичною
радою з якості ННІ АКOT
Протокол № 4
від 28 лютого 2022 р.

Рівне – 2022

Методичні вказівки до виконання практичних робіт з навчальної дисципліни «Теорія ймовірностей і математична статистика» для здобувачів вищої освіти першого (бакалаврського) рівня за освітньо-професійною програмою «Цифрові технології дистанційної освіти» спеціальності 015.39 «Професійна освіта (Цифрові технології)» денної та заочної форм навчання [Електронне видання] / Кушнір О. О., Кушнір В. П. – Рівне : НУВГП, 2022. – 46 с.

Укладачі:

Кушнір О. О., к.ф.-м.н., доцент кафедри вищої математики;
Кушнір В. П., к.ф.-м.н., доцент кафедри вищої математики.

Відповідальний за випуск: Тадеєв П. О., д.пед.н., проф.,
завідувач кафедри вищої математики.

Керівник групи забезпечення
спеціальності
«Професійна освіта
(Цифрові технології)»

Рощенко А.М.,
к.пед.н., доцент.

© Кушнір О. О.,
Кушнір В. П., 2022
© НУВГП, 2022

Зміст

Вступ.....	3
§ 1. Розподіли випадкових величин	4
§ 2. Системи випадкових величин.....	10
§ 3. Обробка статистичної інформації.....	14
§ 4. Надійні інтервали	17
§ 5. Параметричні статистичні критерії	22
§ 6. Перевірка статистичних гіпотез про дискретний розподіл та про відсутність зв'язку.....	28
§ 7. Перевірка гіпотези про неперервний розподіл за критеріями χ^2 -квадрат та Колмогорова	33
§ 8. Кореляційний аналіз.....	38
§ 9. Дисперсійний аналіз.....	43
Література.....	46

Вступ

Для розв'язування задач математичної статистики розроблені статистичні програми. Для початкового ознайомлення з предметом доцільно користуватися вільною програмою з відкритим кодом LibreOffice Calc.

З метою надання студентам методичної допомоги з дисципліни "Теорія ймовірностей і математична статистика" до кожної лабораторної роботи подано короткі теоретичні відомості, які стосуються переважно особливостям вбудованих функцій програми LibreOffice Calc, та наведені алгоритми виконання типових завдань з методичних вказівок 04-02-47 з репозиторію НУВГП. Нумерація завдань збережена без змін.

Програму LibreOffice Calc можна завантажити за посиланням

<https://www.libreoffice.org/download/download/?type=win->

[x86_64&lang=uk](#). Із двох запропонованих версій варто вибрати старішу, бо вона стабільніша.

§ 1. Розподіли випадкових величин

Мета роботи: Навчитися використовувати електронні таблиці для знаходження ймовірностей набування випадковими величинами значень із заданого числового проміжку, познайомитися з поняттям квантилів.

Вступна інформація

У LibreOffice Calc можна використовувати наступні функції, які стосуються нормального розподілу:

1. PNI — щільність стандартного нормального розподілу

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

2. GAUSS — функція Лапласа

$$\Phi_0(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx = \int_0^t \varphi(x) dx.$$

3. NORM.S.DIST — функція стандартного нормального

розподілу $\Phi(t) = \frac{1}{2} + \Phi_0(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx = \int_{-\infty}^t \varphi(x) dx,$

якщо другий аргумент 1 та його щільність

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ якщо другий аргумент } 0.$$

4. NORM.S.INV — квантиль u_Y стандартного нормального розподілу ($\Phi(u_Y) = Y$).
5. STANDARDIZE — z-стандартизація $z = \frac{t - \mu}{\sigma}$.
6. NORM.DIST=NORMDIST — функція нормального розподілу $F(t) = \Phi\left(\frac{t - \mu}{\sigma}\right)$, якщо четвертий аргумент 1 та його щільність $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$, якщо четвертий аргумент 0. (Ця функція є композицією NORM.S.DIST з STANDARDIZE у випадку 1).
7. NORM.INV=NORMINV — квантиль x_Y нормального розподілу ($F(x_Y) = Y$).
8. ERF.PRECISE — функція помилок
- $$\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-x^2} dx = 2\Phi_0(t\sqrt{2}).$$
- У свою чергу функція Лапласа $\Phi_0(t) = \frac{1}{2} \operatorname{erf}\left(\frac{t}{\sqrt{2}}\right)$.
9. ERFC.PRECISE — доповнювальна функція помилок
- $$\operatorname{erfc}(t) = 1 - \operatorname{erf}(t).$$

Для показникового розподілу використовуємо математичну функцію EXP, а для розподілу Пуассона — POISSON=POISSON.DIST. Ці функції при значенні 0 останнього параметра видають закон розподілу, а при значенні 1 —

правобічну границю функції розподілу. Якщо ж 1-й аргумент цієї функції 0, то вона видає EXP(-x). Оберненою до функції EXP є математична функція LN.

Виконання роботи № 1

Задача № 83. $\mu=72; \sigma=7$.

$$a) = \text{NORMDIST}(80; 72; 7; 1) \approx 0,8735;$$

$$б) = \text{NORMDIST}(60; 72; 7; 1) \approx 0,0432;$$

$$в) = 1 - \text{NORMDIST}(90; 72; 7; 1) \approx 0,0051;$$

$$г) = \text{NORMDIST}(75; 72; 7; 1) - \text{NORMDIST}(65; 72; 7; 1) \approx 0,5072;$$

$$= \text{NORMINV}(0,95; 72; 7) \approx 83,51.$$

Щоб знайти відповідь на останнє питання задачі, знаходимо дисперсію маси однієї людини $DX = \sigma^2 = 49 \text{ кг}^2$.

Тоді, за властивістю дисперсії суми незалежних випадкових величин, дисперсія маси 4-х людей

$$\begin{aligned} DY &= D(X_1 + X_2 + X_3 + X_4) = D(X_1) + D(X_2) + D(X_3) + D(X_4) = \\ &= 49 * 4 = 196 \text{ кг}^2, \end{aligned}$$

а середнє квадратичне відхилення $\sigma(Y) = \sqrt{DY} = 14 \text{ кг}$.

$$\begin{aligned} MY &= M(X_1 + X_2 + X_3 + X_4) = M(X_1) + M(X_2) + M(X_3) + M(X_4) = \\ &= 72 * 4 = 288 \text{ кг}. \end{aligned}$$

Далі знаходимо $= \text{NORMDIST}(320; 288; 14; 1) \approx 0,9889$.

Задача № 84. $\sigma=15, n=16$.

Тоді $DX = \sigma^2 = 225$. За властивостями дисперсії,

$$D\bar{x} = D\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2}(D(X_1) + D(X_2) + \dots + D(X_n)) = \\ = \frac{DX}{n} = \frac{225}{16} \Rightarrow \sigma(\bar{x}) = \sqrt{D(\bar{x})} = \frac{15}{4} = \frac{\sigma(X)}{\sqrt{n}} = 3,75.$$

Далі знаходимо

$$a) = 2 * \text{GAUSS}(3 / 3,75) \approx 0,5763;$$

$$б) = 1 - 2 * \text{GAUSS}(10 / 3,75) \approx 0,00766.$$

Для відповіді на останнє запитання задачі спочатку потрібно розв'язати рівняння $2\Phi_0(x) = 0,95$, що рівносильне $2(\Phi(x) - 0,5) = 0,95$, звідки $\Phi(x) = 0,975$, тому $x = U_{0,975}$. Знаходимо його, набравши $=\text{NORM.S.INV}(0,975) \approx 1,96$.

$$\text{Тоді } \frac{5}{\sigma} \sqrt{n} \approx 1,96, \text{ звідки знаходимо } n \approx (3 * 1,96)^2 \approx 34,6.$$

Для гарантії заокруглювати потрібно до більшого. Отже, $n=35$.

Задача № 85. $\lambda=2/\text{год}$.

а) Очевидно, що відповідь не залежить від того, хто із працівників раніше відремонтує свій прилад та візьметься за ремонт 3-го. Але, оскільки показниковий розподіл — “нестаріючий”, то він має однакові шанси відремонтувати 3-й прилад раніше чи пізніше

від того, коли його напарник відремонтує свій 1-й прилад. Тому відповідь 0,5.

б) Оскільки ймовірність добутку незалежних подій дорівнює добуткові їх імовірностей, то потрібно піднести до квадрату ймовірність того, що один працівник відремонтує свій прилад за годину: $p = (1 - e^{-\lambda t})^2$. Тоді набираємо $p = (1 - \text{EXP}(-2 \cdot 1))^2 \approx 0,7476$.

в) Переводимо хвилини в години: 20 хв = 1/3 год, 40 хв = 2/3 год. Далі знаходимо $p = \text{EXP}(-2 \cdot 1/3) - \text{EXP}(-2 \cdot 2/3) \approx 0,2498$.

Задача № 86. $t=10$ с.

а) Позначимо x ймовірність перемоги 2-го гравця. Ця подія може статися лише в тому випадку, коли у 1-го гравця не випаде 6. Ймовірність цього 5/6. Якщо ж після цього у 2-го гравця випаде 6 (ймовірність цього 1/6), то він уже виграє. Але він може виграти і тоді, коли в нього 6 не випаде. Для цього необхідно, щоб у 3-го гравця не випала 6 (ймовірність цього 5/6). Після цього знову підкидатиме кубик 1-й гравець — усе починається спочатку. За цих умов ймовірність перемоги 2-го гравця x . Таким чином,

$$x = \frac{5}{6} \left(\frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot x \right).$$

З цього рівняння знаходимо $x = 30 / 91$.

б) Щоб гра тривала довше 2-х хвилин, тобто 120 секунд, необхідно, щоб було більше від $120/10=12$ підкидань кубика. Отже, принаймні 13 разів має не випадати 6. Ймовірність цієї події $p = (5/6)^{13} \approx 0,09346$.

в) Математичне сподівання геометричного розподілу $\frac{q}{p^2}$. Такою буде середня кількість промахів — невипадання 6. Завершується гра випаданням 6. Тоді математичне сподівання кількості підкидань $\frac{q}{p^2} + 1$. Кожне підкидання триває $t=10$ с.

Отже, середня тривалість гри $\left(\frac{q}{p^2} + 1\right)t = 310\text{с} = 5 \text{ хв. } 10 \text{ с.}$

Задача № 87. $\lambda=3/\text{хв.}$ Тоді $a=3*2=6$.

а) =POISSON.DIST(4;6;0) $\approx 0,1339$;

б) =POISSON.DIST(3,5;6;1) $\approx 0,1512$;

в) = 1 - 0,1512 = 0,8488.

Середній час між замовленнями $T=1/\lambda=3$ хв. Цей розподіл — показниковий. Його дисперсія $\frac{1}{\lambda^2} = 9\text{хв}^2$, а середнє квадратичне відхилення $1/\lambda=3$ хв.

Задача № 88. $p=0,99$.

Ймовірність того, що буде менше однієї родзинки, тобто 0 родзинок $\text{EXP}(-a) = 1-p=0,01$. Тоді $a = -\text{LN}(0,01) \approx 4,605$. Середнє значення розподілу Пуассона дорівнює його параметрові a . Отже, в булочках у середньому має бути по 4,605 родзинок.

§ 2. Системи випадкових величин

Мета роботи: Навчитися знаходити числові характеристики двовимірних випадкових величин, познайомитися з двовимірним нормальним розподілом, знаходити рівняння лінійної регресії.

Вступна інформація. Електронні таблиці зручні для дослідження двовимірних дискретних розподілів, оскільки дозволяють автоматично виконувати прості операції множення рядків та знаходження сум їх елементів. Для виконання обчислень на початку комірки набираємо “=”.

Виконання роботи № 2

Задача № 89.

1. Переносимо таблицю в LibreOffice Calc. Додаємо всі ймовірності в рядках і стовпчиках (не включаючи в суму значення X та Y). Від 1 віднімаємо загальну суму. Отримаємо число p . Замінюємо в таблиці літеру p на отримане число.

$Y \setminus X$	0	1	2	
2	0,3	0,1	0	0,4
3	0	0,2	0,1	0,3
5	0	0,1	0,2	0,3
	0,3	0,4	0,3	1

2. Перший та останній рядки таблиці задають закон розподілу випадкової величини X , а 1-й та останній стовпчики — закон розподілу Y (без вільних полів). Перепишемо їх у окремі таблиці.

X	0	1	2
-----	---	---	---

P	0,3	0,4	0,3
Y	2	3	5
P	0,4	0,3	0,3

- Для обчислення математичних сподівань множимо числа на їх імовірності та знаходимо суму добутків.
- У коваріаційну матрицю записуються дисперсії та коваріація. Для знаходження DX віднімаємо від кожного значення випадкової величини X її математичне сподівання, підносимо до квадрату, множимо на відповідну ймовірність та додаємо. Так само знаходимо і $DY=3,2$.

X	0	1	2	Σ	
P	0,3	0,4	0,3	1	
$x_i p_i$	0	0,4	0,6	1	= MX
$X-MX$	-1	0	1		
$(X-MX)^2$	1	0	1		
$(x_i-MX)^2 p_i$	0,3	0	0,3	0,6	= DX

- Для знаходження коваріації усі ймовірності двовимірного закону розподілу p_{ij} множимо на (x_i-MX) та на (y_j-MY) та додаємо усі ці добутки.

$Y-MY/X-MX$	-1	0	1	Σ	
-1,2	0,36	0	0	0,36	
-0,2	0	0	-0,02	-0,02	
1,8	0	0	0,36	0,36	
Σ	0,36	0	0,34	0,7	= $cov(X,Y)$

6. Записуємо кореляційну матрицю $K(X, Y) = \begin{pmatrix} 0,6 & 0,7 \\ 0,7 & 1,56 \end{pmatrix}$.

7. Для знаходження коефіцієнта кореляції спочатку добуваємо корені з обох дисперсій, отримуючи таким чином середні квадратичні відхилення. Потім ділимо коваріацію на їхній добуток. Отримуємо $\rho(X, Y) \approx 0,7235$.

8. В рівнянні лінійної регресії Y на X коефіцієнт a біля $(X - MX)$ обчислюємо, поділивши коваріацію на DX . Записуємо рівняння лінійної регресії $Y_X = MY + a(X - MX)$.

$$Y_X \approx 3,2 + 1,1667(X - 1).$$

9. Аналогічно знаходимо $X_Y = MX + \frac{\text{cov}(Y, X)}{DY}(Y - MY)$.

$$X_Y \approx 1 + 0,4487(Y - 3,2).$$

10. Знаходимо $D(Y - X) = DY - 2\text{cov}(Y, X) + DX = 0,76$.

11. Знаходимо

$$D(3X + 2Y) = 3^2 DX + 2 * 3 * 2 * \text{cov}(Y, X) + 2^2 DY = 20,04.$$

12. Випишемо частину двовимірного закону закону, що відповідає $X=1$.

Y	2	3	5	Σ
P	0,1	0,2	0,1	0,4

13. Знаходимо умовні ймовірності, поділивши безумовні на їхню суму.

Y	2	3	5	Σ
P	0,25	0,5	0,25	1

14. Тепер знаходимо $M(Y/X=1)$, помноживши значення Y на умовні ймовірності та додавши.

Y	2	3	5	Σ	
P	0,1	0,2	0,1	0,4	
$P(Y/X=1)$	0,25	0,5	0,25	1	
	0,5	1,5	1,25	3,25	$= M(Y/X=1)$

15. Аналогічно знаходимо умовне математичне сподівання Y при інших значеннях X .

16. Записуємо таблицю закону розподілу умовного математичного сподівання Y відносно X , взявши ймовірності відповідних значень X .

$M(Y/X)$	2	3,25	4,333
P	0,3	0,4	0,3

17. За цією таблицею обчислюємо дисперсію умовного математичного сподівання та його середнє квадратичне відхилення.

$M(Y/X)$	2	3,25	4,333	
P	0,3	0,4	0,3	
$z_i p_i$	0,6	1,3	1,3	3,2 $=MZ$
$Z-MZ$	-1,2	0,05	1,133	
$(Z-MZ)^2$	1,44	0,0025	1,2844	
$(Z-MZ)^2$	1,44	0,0025	1,2844	
	0,432	0,001	0,3853	0,8183 $=DM(Y/X)$

18. Поділивши останнє на середнє квадратичне відхилення Y , отримаємо кореляційне відношення $\eta(Y/X) \approx 0,7243$. Воно виявилось не набагато більшим від коефіцієнта кореляції

$\rho(X, Y) \approx 0,7235$. Отже, регресія Y на X близька до лінійної функції.

19. Оскільки кореляційне відношення лежить в межах від 0,7 до 0,9, то кореляційний зв'язок між випадковими величинами сильний. Врахувавши, що коефіцієнт кореляції додатний, стверджуємо, що кореляційний зв'язок прямий.

Задача № 90.

Оскільки випадкові величини нормально розподілені, то обидві регресії — лінійні функції, обидва кореляційні відношення дорівнюють модулю коефіцієнта кореляції. Останній знаходимо за формулою “косинуса кута між векторами”:

$$\rho = \frac{5 * 4 + 3 * (-1)}{\sqrt{5^2 + 3^2} \cdot \sqrt{4^2 + (-1)^2}} = \frac{17}{\sqrt{34 * 17}} = \frac{\sqrt{2}}{2} \approx 0,7071.$$

Отже, кореляційний зв'язок прямий, сильний.

Рівняння регресії $U_V = -7 + (V - 2)$; $U_V = V - 9$;

$$V_U = 2 + 0,5 (U + 7); \quad V_U = 0,5 U + 5,5.$$

§ 3. Обробка статистичної інформації

Мета роботи: Навчитися обробляти статистичну інформацію, знаходити емпіричну функцію розподілу, точкові та інтервальні оцінки параметрів.

Вступна інформація

Вибіркове середнє можна знайти, використавши значок Σ

— Середнє значення. Крім того можна використати статистичні функції AVERAGE, AVERAGEA.

Вибіркову дисперсію S^2 можна знайти за допомогою кожної з наступних статистичних функцій: VARP=VARPA= VAR.P,

а незсунену оцінку дисперсії $S_e^2 = \frac{n}{n-1} S^2$ — VAR=VARA= VAR.S.

Вибіркове середнє квадратичне відхилення S знаходимо, добуваємо квадратний корінь (функція SQRT) з S^2 або безпосередньо за допомогою статистичних функцій STDEVP=STDEVPA=STDEV.P.

S_e знаходимо, добуваючи квадратний корінь із S_e^2 або використовуючи статистичні функції STDEV=STDEVA=STDEV.S.

Виконання роботи № 3

Задача № 91.

1. Вносимо вибірку в електронну таблицю. Для контролю підраховуємо кількість її елементів (обсяг). Потім копіюємо її для подальшої обробки.
2. Варіаційний ряд утворюємо, впорядкувавши вибірку за зростанням. Для цього вибираємо в меню Дані — Сортування за зростанням — Поточне вибране.
3. Записуємо обсяг вибірки $n = 14$ (можна використати значок Σ — кількість), підраховуємо частоти усіх варіант та складаємо таблицю дискретного варіаційного ряду.

X	1	2	3	4	5	Σ
n	2	3	3	5	1	14

4. Знаходимо відносні частоти, поділивши частоти на обсяг вибірки.

X	1	2	3	4	5	Σ
n	2	3	3	5	1	14
ω	0,143	0,214	0,214	0,357	0,071	1

5. Додаючи послідовно відносні частоти, знаходимо значення правобічної границі емпіричної функції розподілу для кожної варіанти вибірки (або накопичені відносні частоти).

X	1	2	3	4	5
n	2	3	3	5	1
ω	0,143	0,214	0,214	0,357	0,071
$F_{14}^*(x+)$	0,143	0,357	0,571	0,929	1

6. Зсунувши значення правобічної границі вправо, знайдемо саму емпіричну функцію розподілу.

X	1	2	3	4	5
n	2	3	3	5	1
ω	0,143	0,214	0,214	0,357	0,071
$F_{14}^*(x+)$	0,143	0,357	0,571	0,929	1
$F_{14}^*(x)$	0	0,143	0,357	0,571	0,929

7. Тепер можна зобразити графік емпіричної функції розподілу. Він має східчастий вигляд.
8. Обсяг $n = 14$; розмах — різниця між найбільшою та найменшою варіантами $R = 5 - 1 = 4$. Вибіркова мода — це варіанта з найбільшою частотою. Найбільша частота 5 у варіанти 4. Отже, мода $M_0 = 4$.

9. У вибірковій медіані емпірична функція розподілу переходить через 0,5. Із таблиці видно, що
- $$F_{14}^*(3) = 0,357 \leq 0,5 \leq 0,571 = F_{14}^*(3+).$$
- Отже, вибіркова медіана $Md=3$.
10. Під вибіркою вставляємо значок Σ — Середнє значення.
Знаходимо вибіркове середнє $\bar{x}=3$.
11. Під варіаційним рядом вставляємо статистичну функцію STDEV.P. Отримуємо $S \approx 1,1952$.
12. Незсунену оцінку дисперсії знаходимо функцією VAR, вибравши діапазон чисел, що містять саму вибірку. Маємо $S_e^2 \approx 1,5385$.
13. Ймовірність події оцінюємо за її відносною частотою. Підраховуємо кількість елементів вибірки (легше варіаційного ряду), які більші від 3, та ділимо на обсяг. Простіше додати відносні частоти варіант, більших від 3.
 $P(X>3) \approx v_{14}(X>3) = (5+1)/14 \approx 0,4286$.

§ 4. Надійні інтервали

Мета роботи: Навчитися знаходити інтервальні оцінки параметрів нормального розподілу та ймовірності події.

Вступна інформація

Для відшукування надійності вказаного інтервалу ймовірності події використовується статистична функція BINOM.DIST=BINOMDIST. Її ж можна використати для уточнення меж приблизно знайдених надійних інтервалів. Слід мати на увазі, що оцінити ймовірність події можна із точністю

не кращою, ніж $\frac{1}{2n}$, де n — обсяг вибірки.

Квантили t -розподілу Стьюдента знаходяться за допомогою функції T.INV. Також для знаходження квантиля

побудови двобічного надійного інтервалу надійності γ можна набрати $=TINV(1-\gamma, n-1)$, де n — обсяг вибірки, $n-1$ — кількість ступенів вільності. Зауважте, що $T.INV.2T=TINV$.

Безпосередньо відхилення від вибіркового середнього в надійному інтервалі для математичного сподівання можна знайти, набравши $=CONFIDENCE.T(1-\gamma; S_e; n)$, де n — обсяг вибірки.

Функція Т-розподілу — статистична функція T.DIST;

$T.DIST.RT=1-T.DIST$;

$T.DIST.2T=2*T.DIST.RT$ визначена тільки при додатних значеннях 1-го аргумента;

$TDIST=T.DIST.2T$, якщо останній аргумент $=2$ і

$TDIST=T.DIST.RT$, якщо останній аргумент $=1$, визначена тільки при додатних значеннях 1-го аргумента.

Квантили хі-квадрат розподілу Хельмерта_Пірсона обчислює функція $CHISQ.INV=CHISQINV$. Їх також можна знайти $=CHISQ.INV.RT(1-\gamma; n)$ та $=CHIINV(1-\gamma; n)$ (тут n — кількість ступенів вільності, а не обсяг вибірки). Функція хі-квадрат-розподілу — статистична функція CHISQ.DIST.

Виконання роботи № 4

Задача № 92. $n=250, k=32, \gamma=0,95$.

Будемо оцінювати ймовірність виготовлення нестандартної деталі з точністю $\frac{1}{2n} = 0,002$. Тому будемо шукати межі надійного інтервалу з 3-ма знаками після коми. Спочатку варто знайти

надійний інтервал приблизно за формулою

$$\left(v_n - \sqrt{\frac{v_n(1-v_n)}{n}} u_{(1+\gamma)/2}, v_n + \sqrt{\frac{v_n(1-v_n)}{n}} u_{(1+\gamma)/2} \right).$$

У ній v_n — відносна частота. Вона дорівнює $= \frac{k}{n} = 0,128$.

Далі знаходимо рівень квантиля $(1+\gamma)/2=0,975$. Сам квантиль знаходимо, використавши статистичну функцію $=\text{NORMSINV}(0,975) \approx 1,960$. Підставивши знайдені числа у формулу знаходимо приблизний надійний інтервал $(0,087; 0,169)$. Тепер перевіримо, наскільки він точний. Для перевірки використовуємо функцію $=\text{BINOM.DIST}(31;250;0,087;1)$. ($31=k-1$.) Отримуємо приблизно 0,982. Це число трохи більше від 0,975. Спробуємо трохи збільшити нижню межу. Крок збільшення

$$\frac{1}{2n} = 0,002$$

$=\text{BINOM.DIST}(31;250;0,089;1) \approx 0,976$. Зрозуміло, що при дальшому збільшенні ймовірність стане вже меншою від 0,975. Тому шукана нижня межа 0,089.

Аналогічно знаходимо верхню межу. $=\text{BINOM.DIST}(32;250;0,169;1) \approx 0,046 > 0,025 = (1-\gamma)/2$. Тому верхню межу обов'язково потрібно збільшувати.

$$=\text{BINOM.DIST}(32;250;0,171;1) \approx 0,039 > 0,025;$$

$$=\text{BINOM.DIST}(32;250;0,173;1) \approx 0,033 > 0,025;$$

$$=\text{BINOM.DIST}(32;250;0,175;1) \approx 0,027 > 0,025.$$

Але різниця стала зовсім невеликою. Тому на наступному

кроці збільшуємо лише на 0,001.

$$= \text{BINOM.DIST}(32;250;0,176;1) \approx 0,0246 < 0,025.$$

Таким чином, верхня межа надійного інтервалу приблизно 0,176.

Отже, шуканий двобічний інтервал надійності 0,95 ймовірності p виготовлення нестандартної деталі (0,089; 0,176).

Верхню оцінку p знаходимо спочатку приблизно за допомогою однієї бічної надійного інтервалу

$$\left(0, v_n + \sqrt{\frac{v_n(1-v_n)}{n}} u_{\gamma} \right). \quad \gamma \text{ — це надійність } 0,95. \text{ Знаходимо}$$

$= \text{NORMSINV}(0,95) \approx 1,645$. Далі обчислюємо приблизну праву межу надійного інтервалу 0,163.

$$= \text{BINOM.DIST}(32;250;0,163;1) \approx 0,076 > 0,05 = 1-\gamma.$$

Потрібно збільшувати нижню межу.

$$= \text{BINOM.DIST}(32;250;0,165;1) \approx 0,065 > 0,05;$$

$$= \text{BINOM.DIST}(32;250;0,167;1) \approx 0,055 > 0,05;$$

$$= \text{BINOM.DIST}(32;250;0,169;1) \approx 0,046 < 0,05.$$

Таким чином, із надійністю 0,95, $p < 0,169$.

Відповідь на останнє запитання задачі можна знайти наступним чином: $= 1 - \text{BINOM.DIST}(32;250;0,15;1) \approx 0,811$. Отже, із надійністю 0,811 можна стверджувати, що $p < 0,15$.

Задача № 93. Вибірка 8, 11, 11, 10, $\gamma = 0,95$.

1. Під вибіркою вставляємо значок Σ — Середнє значення.

Знаходимо вибіркоче середнє $\bar{x} = 10$.

2. Незсунену оцінку дисперсії знаходимо функцією VAR, вибравши діапазон чисел, що містять саму вибірку. Маємо $S_e^2 = 2$.
3. Знаходимо $S_e = \sqrt{2}$, набравши =SQRT(2) $\approx 1,414$.
4. Знаходимо квантиль рівня $(1+\gamma)/2=0,975$ t -розподілу Стьюдента з $n-1=3$ ступенями вільності, набравши =T.INV(0,975;3) або =TINV(0,05;3). Отримаємо $\approx 3,182$.
5. Тепер можемо знайти двобічний надійний інтервал для математичного сподівання за формулами

$$\left(\bar{x} - \frac{S_e}{\sqrt{n}} t_{(1+\gamma)/2}, \bar{x} + \frac{S_e}{\sqrt{n}} t_{(1+\gamma)/2} \right).$$

Отримаємо $10 \pm 2,250$, тобто (7,75; 12,25).

6. Для побудови надійного інтервалу для дисперсії відшукаємо квантилі хі-квадрат розподілу Хельмерта-Пірсона рівнів $(1+\gamma)/2=0,975$ та $(1-\gamma)/2=0,025$ з $n-1=3$ ступенями вільності. =CHISQINV(0,975;3) $\approx 9,3484$;
=CHISQINV(0,025;3) $\approx 0,2158$.
7. Знаходимо двобічний надійний інтервал для дисперсії за формулами $\left(\frac{(n-1)S_e^2}{k_{(1+\gamma)/2}}, \frac{(n-1)S_e^2}{k_{(1-\gamma)/2}} \right)$. Отримаємо (0,64; 27,80).
8. =T.DIST.RT($(7-\bar{x}) * \sqrt{n} / S_e$; 3) $\approx 0,9880$.

§ 5. Параметричні статистичні критерії

Мета роботи: Навчитися виявляти наявність різниці середніх у випадках зв'язаних та незв'язаних вибірках та наявність різниці дисперсій.

Вступна інформація

Виявляти наявність різниці середніх потрібно за допомогою t-критерію Стюдента навіть якщо розподіл не є нормальним, а наявність різниці дисперсій — за допомогою F-критерію Фішера. Для перевірки рівності (чи нерівності) дисперсії заданому числу, використовуємо факт, що випадкова величина $\frac{(n-1)S_e^2}{\sigma^2}$ має хі-квадрат розподіл Хельмерта-Пірсона з $n - 1$ ступенями вільності.

Квантили F-розподілу Беренса-Фішера-Снедокора можна знайти за допомогою статистичної функції F.INV. Але для застосування F-критерію Фішера зручнішими є функції F.INV.RT=F.INV, оскільки $F.INV(1-\alpha; m; n) = F.INV(\alpha; m; n)$.

Перевірити гіпотезу про рівність дисперсій за двома незв'язними вибірками можна, використавши статистичну функцію F.TEST. Якщо результат буде більшим від вказаного рівня значущості, то немає підстав відкидати гіпотезу про рівність дисперсій за F-критерієм Фішера.

Набравши =T.TEST(Дані 1; Дані 2; 2; 1), можна перевірити гіпотезу про рівність математичних сподівань при загальній альтернативі за двома зв'язними вибірками. Якщо результат буде більшим від вказаного рівня значущості, то немає підстав стверджувати, що математичні сподівання різні. У випадку односторонньої альтернативи слід набирати =T.TEST(Дані 1; Дані 2; 1; 1), але альтернативна гіпотеза приймається лише тоді,

коли значення отриманої функції менше від заданого рівня значущості і відношення між середніми таке ж, яке очікується між математичними сподіваннями.

Перевірити гіпотезу про рівність математичних сподівань при загальній альтернативі за двома незв'язними вибірками у припущенні, що дисперсії рівні, можна, набравши =T.TEST(Дані 1; Дані 2; 2; 2). Якщо ж альтернатива одностороння, то набираємо =T.TEST(Дані 1; Дані 2; 1; 2). Якщо виявиться, що отриманий результат менший від заданого рівня значущості, то приймається альтернативна гіпотеза за умови, що відношення між середніми таке ж, яке очікується між математичними сподіваннями.

Якщо ж дисперсії незв'язних вибірок не рівні, то використовуються функції =T.TEST(Дані 1; Дані 2; 2; 3), =T.TEST(Дані 1; Дані 2; 1; 3) при загальній та односторонній альтернативах відповідно.

Виконання роботи № 5

Задача № 94. $n=10$, $a=1$; $\alpha=0,05$. Одна вибірка. Двобічний критерій для перевірки математичного сподівання.

1. Вносимо вибірку в електронну таблицю.
2. Знаходимо вибіркоче середнє значення, вибравши Σ — Середнє значення. Отримуємо 0,986.
3. Знаходимо S_s за допомогою статистичної функції STDEV. Отримуємо приблизно 0,008055.
4. Знаходимо модуль (абсолютне значення) відхилення вибіркового середнього від норми =ABS(0,986-1)=0,014.
5. Множимо його на квадратний корінь з обсягу вибірки

- *SQRT(10) та ділимо на S_g . В результаті отримуємо t-статистику $\approx 2,355$.
- Знаходимо двобічний квантиль t-розподілу Стьюдента $=T.INV.2T(\alpha; n-1)$, підставивши $\alpha=0,05, n-1=9$. Отримали приблизно 2,262.
 - Оскільки t-статистика $2,355 > 2,262$, то робимо висновок, що є систематична похибка. Середнє значення величини X відрізняється від стандартного.

Задача № 95. $n = 10, m=50; \sigma^2 \leq 0,01; \alpha=0,05$. Одна вибірка. Двобічний критерій для перевірки математичного сподівання та одnobічний для перевірки дисперсії.

- Копіюємо розв'язок попередньої задачі, замінивши вибірку та 1 на 50.
- Оскільки t-статистика $1,051 < 2,262$, то робимо висновок, що немає підстав відкидати основну гіпотезу, що маса пакунка з природним барвником відрізняється від 50 кг.
- Побудуємо одnobічний надійний інтервал для дисперсії $\left(\frac{(n-1)S_g^2}{k_\gamma}, +\infty \right)$, де $\gamma=1-\alpha$. Для цього знаходимо квантиль хі-квадрат розподілу з $n-1=9$ ступенями вільності рівня $\gamma=1-\alpha$
 $k_\gamma = \text{CHISQ.INV.RT}(\alpha; n-1) = \text{CHISQ.INV.RT}(0,05; 9) \approx 16,92$.
- Підносимо S_g до квадрату. Результат множимо на $n-1=9$ та ділимо на квантиль 16,92. Отримуємо надійний інтервал $(0,0151; +\infty)$.
- Оскільки число 0,01 не належить до цього надійного інтервалу, то гіпотеза, що $\sigma^2 \leq 0,01 \text{ кг}^2$ відкидається на рівні значущості $\alpha=0,05$. З надійністю 0,95 можна стверджувати, що дисперсія більша від $0,015 \text{ кг}^2$.

Задача № 96. $n_1 = 3$, $n_2 = 3$; $\alpha = 0,01$. У цій задачі, на відміну від попередніх, є дві незв'язні вибірки. Крім того, в цьому випадку слід використовувати однобічний критерій для перевірки рівності математичних сподівань, бо сказано, що другий спосіб є удосконаленим, значить очікується, що він має бути міцнішим.

1-й спосіб розв'язування

1. Заносимо дані до електронної таблиці. Знаходимо вибіркові середні значення спротиву тиску кожної з двох марок бетону та незсунені оцінки дисперсій (VAR).
2. Оскільки обсяги вибірок рівні, то t -статистику можна обчислювати за формулою $\hat{t}_{2n-2} = \sqrt{n} \frac{\bar{y} - \bar{x}}{\sqrt{S_x^2 + S_y^2}}$. Тут $n=3$,
 U — середнє значення міцності удосконаленої марки бетону.
3. Знаходимо однобічний квантиль t -розподілу Стьюдента $=T.INV(1-\alpha; 2n-2)$, підставивши $1-\alpha=1-0,01=0,99$, $2n-2=4$. Отримали приблизно 3,747.
4. Оскільки t -статистика 2,294 менша від квантиля 3,747, то даних експерименту недостатньо, щоб стверджувати, що міцність бетону удосконаленої марки вища.

2-й спосіб розв'язування

1. Заносимо дані до електронної таблиці. Знаходимо вибіркові середні значення спротиву тиску кожної з двох марок бетону.
2. Вибіркове середнє спротиву тиску удосконаленої марки бетону вище, ніж звичайної. Залишилося перевірити чи

ймовірність помилки 1-го роду є меншою від заданого рівня значущості. Щоб з'ясувати, яку форму критерію Стьюдента можна використати, спочатку перевіримо рівність дисперсій.

3. Набираємо $=F.TEST(\text{Дані 1}; \text{Дані 2}) \approx 0,2368 > 0,05$. Отже, немає підстав стверджувати, що дисперсії відрізняються.
4. $=T.TEST(\text{Дані 1}; \text{Дані 2}; 1; 2) \approx 0,0417 > 0,01$. Даних недостатньо, щоб стверджувати, що міцність бетону удосконаленої марки вища.

Зауваження

1. Якби був заданий рівень значущості 0,05, то була б прийнята альтернативна гіпотеза, що удосконалена марка бетону міцніша.
2. Все-таки незсунені оцінки дисперсій сильно відрізняються. Очевидно, що критерій Фішера не дозволив зробити висновок, що дисперсії різні через малі обсяги вибірок. Тому спробуємо застосувати t -критерій Стьюдента, не припускаючи рівності дисперсій.
3. $=T.TEST(\text{Дані 1}; \text{Дані 2}; 1; 3) \approx 0,061$. Це число більше навіть від 0,05. Вибірки такого малого обсягу ($n=3$) не дозволили зробити висновок, що міцність удосконаленої марки бетону вища.

Задача № 97. $n = 10$, $\alpha=0,05$. Вибірки зв'язні, двобічний критерій для перевірки рівності математичних сподівань, бо сіялки рівноправні.

1-й спосіб розв'язування

1. Заносимо дані до електронної таблиці та віднімаємо від

останнього рядка попередній. Далі працюватимемо з однією отриманою вибіркою різниць — так само, як у задачах 94, 95. Цього разу перевіряємо середнє значення 0.

2. Оскільки $3,214 > 2,262$, то на рівні значущості 0,05 є різниця між сіялками. При застосуванні сіялки 1-го типу урожай вищий.

2-й спосіб розв'язування

1. Заносимо дані до електронної таблиці.
2. =T.TEST(Дані 1; Дані 2; 2; 1) $\approx 0,0106$. Це менше від заданого рівня значущості $\alpha=0,05$. Урожай таки залежить від типу сіялки.
3. Знайшовши вибіркові середні урожаю, робимо висновок, що при застосуванні сіялки 1-го типу урожай вищий.

Задача № 98. $S_X^2 = 385,52$; $n_X = 64$; $S_Y^2 = 636,24$; $n_Y = 17$, $\alpha=0,05$. Двобічний критерій перевірки рівності дисперсій.

1. Ділимо більшу незсунену дисперсію на меншу. Отримаємо F-статистику приблизно 1,650.
2. Оскільки більшою є незсунена оцінка дисперсії випадкової величини Y , то кількість ступенів вільності чисельника $n_Y - 1 = 17 - 1 = 16$.
3. Тоді кількість ступенів вільності знаменника, $n_X - 1 = 64 - 1 = 63$.
4. Оскільки ми вибрали більшу незсунену оцінку дисперсії, рівень значущості ділимо на 2: $= 0,05 / 2 = 0,025$.

5. Знаходимо квантиль F-розподілу Беренса-Фішера-Снедекора $=F.INV.RT(0,025;16;63) \approx 2,022$.
6. Оскільки F-статистика $1,650 < 2,022$, то на рівні значущості 0,05 немає підстав стверджувати, що дисперсії X та Y відрізняються між собою.

§ 6. Перевірка статистичних гіпотез про дискретний розподіл та про відсутність зв'язку

Мета роботи: Навчитися перевіряти статистичні гіпотези про заданий дискретний розподіл та про незалежність дискретних випадкових величин за критерієм хі-квадрат Пірсона.

Вступна інформація. Для можливості застосування критерію хі-квадрат Пірсона потрібно прослідкувати, щоб теоретичні частоти були не меншими від 5. В іншому випадку деякі з проміжків потрібно об'єднати. Хі-квадрат статистику знаходимо за

формулою $\sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}$, де n_i — емпіричні частоти, n'_i —

теоретичні частоти. Кількість ступенів вільності знаходимо за формулою $k - 1 - s$, де k — кількість проміжків, s — кількість оцінюваних параметрів. При перевірці гіпотези про відсутність зв'язку між випадковими величинами кількість ступенів вільності дорівнює $(k - 1) * (l - 1)$, де k — кількість рядків, а l — кількість стовпчиків.

Виконання роботи № 6

Задача № 99. $\alpha=0,05$, Пуассона, геометричний, біномний.

Оскільки всі вказані розподіли є параметричними, потрібно насамперед оцінити їхні параметри. Для цього оцінюємо

математичне сподівання.

1. Вносимо дані в електронну таблицю. Обчислюємо обсяг вибірки N , як суму частот: $N=50$.
2. Знаходимо вибіркове середнє значення, помноживши рядки, знайшовши суми та поділивши суму добутків на обсяг вибірки N . Отримуємо $\bar{x}=1,74$. Це число й буде оцінкою параметра a розподілу Пуассона.

0	1	2	3	4	5	6	7	Σ	\bar{x}
15	12	9	6	4	2	1	1	50	
0	12	18	18	16	10	6	7	87	1,74

3. Оскільки математичне сподівання геометричного розподілу дорівнює $1/p-1$, то оцінку параметра p геометричного розподілу знаходимо з рівняння $1/p-1=1,74$. Отже, для геометричного розподілу $p \approx 0,3650$.
4. Математичне сподівання біномного розподілу дорівнює np . Найбільш правдоподібно, що кількість випробувань n дорівнює максимальній кількості успіхів, тобто $n=7$. Тоді оцінкою p біномного розподілу буде $1,74/7 \approx 0,2486$.
5. Знаходимо закон розподілу Пуассона з параметром $a=1,74$. Для цього використовуємо статистичну функцію $=\text{POISSON.DIST}(x; 1,74; 0)$, де x — можливі значення розподілу Пуассона від 0 до 6. Усі інші значення від 7 до нескінченності об'єднуємо в одну комірку, а його ймовірність знаходимо, віднявши від 1 суму решти ймовірностей.
6. Обчислюємо теоретичні частоти, помноживши ймовірності на обсяг вибірки $N=50$.

7. Об'єднуємо комірки з теоретичними частотами, меншими від 5.

0	1	2	3	4 і більше	Σ
8,776	15,27	13,29	7,705	4,96	50

8. У 5-й комірці теоретична частота менша від 5. Об'єднуємо її з попередньою. Випишемо також емпіричні частоти.

0	1	2	3 і більше	Σ
8,78	15,27	13,29	12,67	50
15	12	9	14	50

9. Тепер знайдемо різниці між теоретичними та емпіричними частотами, піднесемо їх до квадрату та поділимо на теоретичні частоти.

x_i	0	1	2	3 і більше	Σ
n'_i	8,78	15,27	13,29	12,67	50
n_i	15	12	9	14	50
$n_i - n'_i$	6,22	-3,27	-4,29	1,33	0
$(n_i - n'_i)^2$	38,74	10,69	18,36	1,77	
$\frac{(n_i - n'_i)^2}{n'_i}$	4,41	0,70	1,38	0,13	6,64

10. Сума в останньому рядку 6,64 є значенням χ^2 -квadrat статистики. Її потрібно порівняти з квантилем χ^2 -квadrat розподілу рівня $1-0,05=0,95$. Було оцінено 1 параметр a . Отже, потрібно знайти квантиль з двома ступенями вільності.

11. $=\text{CHISQ.INV}(0,95;2) \approx 5,99$.

12. Оскільки $6,64 > 5,99$, то гіпотеза про розподіл Пуассона відкидається.

13. Знаходимо геометричний розподіл за формулою $(1-p)^k p$, $0 \leq k \leq 6$. Ймовірність останньої комірки “7 і більше” знаходимо, віднявши від 1 суму ймовірностей у перших 7 комірках.

0	1	2	3	4	5	6	7 і >
0,365	0,232	0,147	0,093	0,059	0,038	0,024	0,042

14. Обчислюємо теоретичні частоти, помноживши ймовірності на обсяг вибірки $N=50$.

0	1	2	3	4	5	6	7 і >
0,365	0,232	0,147	0,093	0,059	0,038	0,024	0,042
18,25	11,59	7,36	4,67	2,97	1,88	1,20	2,08

15. Об'єднуємо комірку $x=3$ з коміркою $x=4$, а також 3 останні.

0	1	2	3 або 4	5 і >
18,25	11,59	7,36	7,64	5,16

16. Підраховуємо емпіричні частоти та обчислюємо хі-квадрат статистику.

x_i	0	1	2	3 або 4	5 і >	Σ
n'_i	18,25	11,59	7,36	7,64	5,16	50
n_i	15	12	9	10	4	50
$n_i - n'_i$	-3,25	0,41	1,64	2,36	-1,16	0
$(n_i - n'_i)^2$	10,55	0,17	2,69	5,57	1,35	

$$\frac{(n_i - n'_i)^2}{n'_i} \quad 0,58 \quad 0,01 \quad 0,37 \quad 0,73 \quad 0,26 \quad 1,95$$

17. Був оцінений 1 параметр. $=\text{CHISQ.INV}(0,95;3) \approx 7,81$.

18. Оскільки $1,95 < 7,81$, то немає підстав відкидати гіпотезу про геометричний розподіл на рівні значущості 0,05.

19. Тепер знаходимо закон розподілу Бернуллі з параметрами $n=7, p=0,2486$ та знаходимо теоретичні частоти.

x_i	0	1	2	3	4	5	6	7	Σ
p_i	0,133	0,311	0,311	0,173	0,058	0,012	0,01	0	1
n'_i	6,67	15,57	15,57	8,65	2,88	0,58	0,06	0	50

20. Щоб не було теоретичних частот, менших від 5, об'єднуємо 5 останніх комірок. Далі знаходимо хі-квадрат статистику.

x_i	0	1	2	3 і більше	Σ
n'_i	6,67	15,57	15,57	12,18	50
n_i	15	12	9	14	50
$n_i - n'_i$	8,33	-3,57	-6,57	1,82	0
$(n_i - n'_i)^2$	69,32	12,77	43,21	3,31	
$\frac{(n_i - n'_i)^2}{n'_i}$	10,39	0,82	2,77	0,27	14,25

21. Був оцінений 1 параметр. $=\text{CHISQ.INV}(0,95;2) \approx 5,99$.

Оскільки $14,25 > 5,99$, то гіпотеза про біномний розподіл відкидається.

Задача № 101. $\alpha=0,05$, гіпотеза про відсутність зв'язку.

1. Знаходимо суми чисел у рядках та у стовпчиках.

34	62	28	124
27	28	20	75
57	105	52	214
118	195	100	413

2. Знаходимо теоретичні частоти, помноживши суму в рядку на суму у стовпчику та поділивши на загальну суму.

35,43	58,55	30,02	124
21,43	35,41	18,16	75
61,14	101,04	51,82	214
118	195	100	413

3. Обчислюємо χ^2 -квadrat статистику.

0,058	0,204	0,136	0,398
1,449	1,551	0,186	3,186
0,281	0,155	0,001	0,436
1,787	1,910	0,324	4,020

4. Кількість ступенів вільності $(3-1)*(3-1)=4$.
 5. $=\text{CHISQ.INV}(0,95;4) \approx 9,488$.
 6. Оскільки $4,02 < 9,488$, то немає підстав відкидати гіпотезу про відсутність зв'язку між однобічним розвитком рук та очей (на рівні значущості 0,05).

§ 7. Перевірка гіпотези про неперервний розподіл за критеріями χ^2 -квadrat та Колмогорова

Мета роботи: Навчитися перевіряти гіпотезу про нормальний розподіл за критеріями χ^2 -квadrat та Колмогорова, а також про рівномірний і показниковий розподіли.

Вступна інформація. Перед застосуванням χ^2 -критерію Пірсона для перевірки гіпотези про неперервний розподіл

потрібно розподілити дані по проміжках. Проміжки краще вибирати так, щоб ймовірності потрапити до них були однаковими. Тоді будуть рівними і теоретичні частоти. Оскільки нормальний розподіл має 2 параметри, то при їх оцінюванні кількість ступенів вільності χ^2 -критерію Пірсона буде $k-3$, де k — кількість інтервалів групування. Отже, має бути не менше 4 інтервалів. Крім того, теоретичні частоти інтервалів не мають бути меншими 5. Тому мінімальний обсяг вибірки, при якому можна застосовувати χ^2 -критерій Пірсона, $4*5=20$.

При $20 \leq n \leq 40$ варто розподіляти дані на 4 проміжки, при $41 \leq n \leq 59$ — на 5 проміжків, при $60 \leq n \leq 100$ — на 6 проміжків, а при $n > 100$ кількість проміжків обчислювати за формулою $k = 1 + [\log_2 n]$, де n — обсяг вибірки.

Для перевірки складної гіпотези про нормальний розподіл за критерієм Колмогорова слід використовувати статистику Стефенса $\lambda^n = D_n \left(\sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}} \right)$ і таблицю критичних точок Ліліфорса

Рівень значущості α	0,15	0,1	0,05	0,025	0,01
Критична точка	0,775	0,819	0,895	0,955	1,035

Для обчислення квантилів показникового розподілу використовуємо математичну функцію $= -\bar{x} * \text{LN}(1-\alpha)$.

Виконання роботи № 7

Задача № 100. $\alpha = 0,01$, $n = 58$, нормальний розподіл.

1. Виберемо 5 проміжків. Але спочатку потрібно оцінити параметри.
2. Заносимо дані до електронної таблиці.
3. Знаходимо вибіркове середнє, скориставшись значком Σ та статистичною функцією =STDEV.S.
4. Отримуємо $\bar{x} \approx 4,781$; $S_{\bar{x}} \approx 0,0153$.
5. Оскільки проміжків буде 5, знаходимо квінтилі стандартного нормального розподілу. $=\text{NORM.S.INV}(0,2) \approx -0,8416$; $=\text{NORM.S.INV}(0,4) \approx -0,2533$; $=\text{NORM.S.INV}(0,6) \approx 0,2533$; $=\text{NORM.S.INV}(0,8) \approx 0,8416$.
6. Множимо ці квінтилі на вибіркове середнє квадратичне відхилення та додаємо вибіркове середнє. В результаті ми отримаємо квінтилі розподілу, що перевіряється. Вони будуть кінцями проміжків.
7. Отже, маємо такі 5 проміжків: $(-\infty; 4,7679)$; $(4,7679; 4,7769)$; $(4,7769; 4,7847)$; $(4,7847; 4,7837)$; $(4,7837; +\infty)$. Їх теоретичні частоти $=58/5=11,6$.
8. Підраховуємо кількість ступенів вільності $5-2-1=2$.
9. Знаходимо квантиль $=\text{CHISQ.INV.RT}(0,01;2) \approx 9,21$.
10. Підрахуємо емпіричні частоти. Для цього спочатку скопіюємо вибірку та перетворимо її на варіаційний ряд, вибравши в меню Дані — Сортування за зростанням —

Поточне вибране.

11. Для підрахунку емпіричних частот зручно виділити відповідні проміжки чисел різними кольорами.

12. Складаємо таблицю відповідностей емпіричних та теоретичних частот.

10	12	11	15	10
11,6	11,6	11,6	11,6	11,6

13. Обчислюємо χ^2 -квдрат статистику.

					Σ
10	12	11	15	10	58
11,6	11,6	11,6	11,6	11,6	58
1,6	-0,4	0,6	-3,4	1,6	0
2,56	0,16	0,36	11,56	2,56	
0,221	0,014	0,031	0,997	0,221	1,483

14. Оскільки $1,483 < 9,21$, то немає підстав відхилити гіпотезу про нормальний розподіл вимірів Міллікеном заряду електрона на рівні значущості $0,01$.

Задача № 100(б). $\alpha = 0,01, n = 58$, перевірити гіпотезу про нормальний розподіл за критерієм Колмогорова.

1. Знайдемо емпіричну функцію розподілу. Для цього починаємо з 0 та додаємо $k/58$ при кожній наступній варіанті варіаційного ряду з частотою k .

2. Знаходимо правобічну границю емпіричної функції розподілу, зсунувши її значення на одну клітинку вгору та

закінчивши таблицю одиницею.

3. Знайдемо значення нормального розподілу, який перевіряється, у кожній варіанті вибірки. Для цього зручно спочатку набрати стовпчики однакових значень з емпіричним середнім та S_e . Далі використовуємо функцію NORM.DIST з останнім аргументом 1.
4. Обчислюємо абсолютні значення різниць теоретичної та емпіричної функцій розподілу, а також теоретичної функції розподілу та правобічної границі емпіричної функції розподілу. Для цього використовуємо математичну функцію ABS.
5. Знаходимо максимальне відхилення за допомогою статистичної функції MAX. Отримуємо $D_{58} = 0,0629$.
6. Обчислюємо статистику Стефенса за формулою
$$\lambda^n = D_n \left(\sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}} \right) \approx 0,486.$$
7. З таблиці критичних точок Ліліфорса вибираємо точку 1,035, яка відповідає рівню значущості 0,01.
8. Оскільки $0,486 < 1,035$, то за критерієм Колмогорова немає підстав відкидати гіпотезу про нормальний розподіл на рівні значущості 0,01.

Задача № 100(в). $\alpha = 0,01, n = 58$, перевірити гіпотезу про рівномірний розподіл за критерієм хі-квадрат Пірсона.

1. Копіюємо вибірку та утворюємо варіаційний ряд.

2. Оцінюємо параметри рівномірного розподілу — найменшу та найбільшу варіанти 4,74 та 4,81.
3. Знаходимо довжину цього відрізка та ділимо її на 5. $h=0,014$.
4. Записуємо відрізки поділу [4,74; 4,754], [4,754; 4,768], [4,768, 4,782], [4,782; 4,796], [4,796, 4,81].
5. Усі теоретичні частоти дорівнюють $58/5=11,6$.
6. За варіаційним рядом підраховуємо емпіричні частоти та заносимо їх у таблицю.

3	8	19	19	9
11,6	11,6	11,6	11,6	11,6

7. Обчислюємо хі-квадрат статистику

					Σ
3	8	19	19	9	58
11,6	11,6	11,6	11,6	11,6	58
8,6	3,6	-7,4	-7,4	2,6	0
73,96	12,96	54,76	54,76	6,76	
6,38	1,12	4,72	4,72	0,58	17,51

8. Як і у випадку перевірки нормального розподілу, було оцінено 2 параметри. Тому кількість ступенів вільності $5-1-2=2$.
9. Знаходимо квантиль $=\text{CHISQ.INV.RT}(0,01;2) \approx 9,21$.
10. Оскільки $17,51 > 9,21$, то немає гіпотеза про рівномірний розподіл вимірів Міллікеном заряду електрона на рівні значущості 0,01 відкидається.

§ 8. Кореляційний аналіз

Мета роботи: Навчитися перевіряти значущість зв'язку між випадковими величинами та оцінювати його тісноту,

будувати надійний інтервал для коефіцієнта лінійної кореляції Браве-Пірсона, записувати рівняння лінійної регресії.

Вступна інформація. Якщо розподіли обох випадкових величин нормальні, то тіснота їхнього зв'язку характеризується коефіцієнтом кореляції Браве-Пірсона. В інших випадках цей коефіцієнт характеризує тільки тісноту їх лінійного зв'язку. За допомогою t -розподілу Стюдента можна перевірити наявність зв'язку між нормально розподіленими випадковими величинами, а також і в інших випадках, якщо обсяг вибірки великий. Тоді

статистика $\hat{t} = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$ матиме t -розподіл із $n-2$

ступенями вільності. Абсолютне значення цієї статистики слід порівнювати з $T.INV.2T(\alpha, n-2)$, де α — рівень значущості.

Коефіцієнт кореляції обчислюється за допомогою статистичних функцій PEARSON або CORREL.

Для знаходження рівняння лінійної регресії використовуються статистичні функції SLOPE (нахил) та INTERCEPT (відрізок).

Щоб побудувати надійний інтервал для коефіцієнта кореляції, використовується перетворення Фішера (статистична функція FISHER) та обернене перетворення (статистична функція FISHERINV).

Виконання роботи № 8

Задача № 102. $\alpha = 0,05, \gamma = 0,95, n = 5$.

1. Заносимо дані до електронної таблиці.
2. За допомогою функції PEARSON обчислюємо коефіцієнт кореляції. Отримуємо $\rho \approx 0,906$.

3. За допомогою статистичної функції SLOPE знаходимо кутовий коефіцієнт рівняння прямої регресії приблизно 1,643.
4. За допомогою статистичної функції INTERSEPT знаходимо відрізок, який відтинає пряма регресії на осі (OY) приблизно 0,643.
5. Отже, рівняння прямої лінії регресії $y_x \approx 1,643x + 0,643$.
6. Для побудови надійного інтервалу для коефіцієнта кореляції, знайдемо його перетворення Фішера за допомогою статистичної функції FISHER. Воно приблизно дорівнює 1,507.
7. Далі потрібно знайти квантиль стандартного нормального розподілу рівня $\frac{1+\gamma}{2} = 0,975$. Для цього використовуємо статистичну функцію NORM.S.INV. Шуканий квантиль приблизно дорівнює 1,96.
8. Тоді ділимо цей квантиль на $\sqrt{n-3}$, де $n=5$ — обсяг вибірки. Отримуємо приблизно 1,386.
9. Віднімаємо та додаємо до перетворення Фішера знайдену величину. Дістанемо інтервал (0,121; 2,892).
10. Від кінців цього інтервалу знаходимо обернене перетворення Фішера за допомогою статистичної функції. Отримаємо (0,120; 0,994) — надійний інтервал для коефіцієнта кореляції.
11. Для перевірки гіпотези про наявність зв'язку знаходимо t -

статистику за формулою $\hat{t} = \frac{|r|}{\sqrt{1-r^2}} \sqrt{n-2}$. Вона приблизно дорівнює 3,715.

12. Знаходимо квантиль t-розподілу Стьюдента рівня $\frac{1-\alpha}{2}$ з $n-2 = 3$ ступенями вільності. $=T.INV.2T(0,05; 3) \approx 3,182$.

13. Оскільки $3,715 > 3,182$, то на рівні значущості 0,05 можна стверджувати, що між випадковими величинами X та Y є кореляційний зв'язок.

14. Оскільки кореляційне відношення може бути більшим від модуля коефіцієнта кореляції, то з надійністю 0,975 можна стверджувати, що кореляційне відношення може бути в межах від 0,121 до 1. Це не дозволяє оцінити силу зв'язку між X та Y . Обсяг вибірки виявився надто малим, щоб зробити висновок.

Задача № 103. $\alpha = 0,01, \gamma = 0,99, n = 5$.

1. Заносимо дані до електронної таблиці. Обчисливши суми частот в рядках і стовпчиках, знайдемо обсяг вибірки $n = 50$, дискретний варіаційний ряд кількості отворів у пилинках фуксії (X) та інтервальний варіаційний ряд діаметрів цих пилинок (Y).

x_i		0	1	2	3	4						Σ
n_i	10	10	10	10	10	10						50
y_i	10	15	20	25	30	35	40	45	50	55	60	Σ
n_j	3	10	6	1	4	5	4	4	6	4	3	50

2. Знайдемо \overline{xy} , перемноживши частоти із значеннями X та Y і поділивши їх на обсяг вибірки. Отримуємо 88,2.
3. Знаходимо вибіркові середні, помноживши варіанти на їх частоти, додавши ці добутки та поділивши на обсяг вибірки. $\overline{x} = 2$ та $\overline{y} = 33,2$.
4. Знаходимо $\overline{x^2} = 6$ та $\overline{y^2} = 1354$.
5. Знаходимо

$$S_x^2 = \overline{x^2} - (\overline{x})^2 = 2 \text{ та } S_y^2 = \overline{y^2} - (\overline{y})^2 = 251,76.$$
6. Добувши квадратні корені з цих виразів, отримаємо вибіркові середні квадратичні відхилення. $S_x \approx 1,414$; $S_y \approx 15,867$.
7. Обчислюємо вибірковий коефіцієнт кореляції за формулою

$$r_{xy} = \frac{\overline{xy} - \overline{x}\overline{y}}{S_x S_y} \approx 0,9715.$$
8. Кутовий коефіцієнт рівняння прямої регресії знаходимо, помноживши вибірковий коефіцієнт кореляції на S_y та поділивши на S_x . Отримуємо 10,9.
9. Знаходимо відрізок, який відтинає пряма регресії на осі (OY) , віднявши від Y добуток кутового коефіцієнта на X . Отримуємо 11,4.
10. Отже, вибіркове рівняння прямої лінії регресії $y_x = 10,9X + 11,4$.

11. Далі будемо надійний інтервал для коефіцієнта кореляції так само, як у попередній задачі тільки надійності 0,99. Отримуємо (0,941; 0,986).
12. Таким чином, з надійністю 0,995, кореляційне відношення не менше від 0,941.
13. Так само, як у попередній задачі, перевіряємо гіпотезу про наявність зв'язку на рівні значущості 0,01. Оскільки $28,4 > 2,68$, то на рівні значущості 0,05 можна стверджувати, що між випадковими величинами X та Y є кореляційний зв'язок (значущий).
14. Оскільки коефіцієнт кореляції додатний, то зв'язок прямий.
15. Оскільки коефіцієнт кореляції більший від 0,9, то зв'язок дуже сильний та близький до лінійного, бо різниця між кореляційним відношенням та коефіцієнтом лінійної кореляції Брауе-Пірсона не може бути великою (вони обидва знаходяться в межах від 0,941 до 1 з надійністю 0,995.)
16. Висновок: зв'язок між діаметром пилинок фуксії та кількістю отворів у них для виходу пилок трубок кореляційний, значущий, дуже сильний, прямий, близький до лінійного.

§ 9. Дисперсійний аналіз

Мета роботи: Навчитися перевіряти наявність впливу фактора на залежну випадкову величину за допомогою дисперсійного аналізу.

Вступна інформація. Дисперсійний аналіз можна застосовувати не тільки для вивчення нормально розподілених випадкових величин, але й тоді, коли дисперсії в групах однакові або ж приблизно однакові обсяги груп.

Квантили F-розподілу знаходяться за допомогою статистичної функції F.INV, а сама функція F-розподілу F.DIST. Але для знаходження квантилів за заданим рівнем значущості зручніше використовувати статистичні функції FINV=F.INV.RT, а рівень значущості, якому відповідає знайдена F-статистика знаходить функція FDIST (без крапки). Таким чином, крапка в цих випадках змінює функцію. Другий аргумент усіх цих функцій — кількість ступенів вільності чисельника, а 3-й — кількість ступенів вільності знаменника.

Виконання роботи № 9

Задача № 104. 4 рівні, обсяги 8; 9; 4; 5.

Оскільки загальний обсяг вибірки $8+9+4+5=26$, то повна сума квадратів SS має $26-1=25$ ступенів вільності.

Міжгрупова сума квадратів SS_{fac} має $4-1=3$ ступені вільності.

Решту $25-3=22$ ступенів вільності має залишкова дисперсія.

Цю кількість можна підрахувати безпосередньо так:

$$(8-1)+(9-1)+(4-1)+(5-1)=7+8+3+4=22.$$

Задача № 105.

	df	SS	MS	F
Між групами	5	16	3,2	1,0667
Всередині груп	18	54	3	
Всього	23	70	3,043	

Задача № 106. 5-ма методами з значення фактора — вибірка зв'язана.

1. Знаходимо вибіркові середні значення результативної випадкової величини при кожному значенні фактора $\overline{Y}_A, \overline{Y}_B, \overline{Y}_C$, а також у кожному вимірі $\overline{Y}_X, \overline{Y}_Y, \overline{Y}_Z, \overline{Y}_U, \overline{Y}_V$ та загальне вибіркове середнє \overline{Y} .
2. Знаходимо повну суму квадратів, помноживши 15 на вибіркову дисперсію усієї вибірки (знайдена за допомогою статистичної функції VAR.P).
3. Знаходимо міжгрупову суму квадратів, помноживши 15 на $\text{VAR.P}(\overline{Y}_A, \overline{Y}_B, \overline{Y}_C)$.
4. Знаходимо міжвимірну суму квадратів, помноживши 15 на $\text{VAR.P}(\overline{Y}_X, \overline{Y}_Y, \overline{Y}_Z, \overline{Y}_U, \overline{Y}_V)$.
5. Знаходимо залишкову суму квадратів, віднявши від повної міжгрупову та міжвимірну.
6. Знаходимо середні квадрати, поділивши суми квадратів на кількість ступенів вільності (для міжгрупової — на 2, для міжвимірної — на 4, а для залишкової — на $2 \cdot 4 = 8$). Отримуємо приблизно 1877; 284 та 184 відповідно.
7. Обчислюємо F-статистики, поділивши на залишкові середні квадрати міжгрупові та міжвимірні. Отримуємо приблизно 10,216 та 1,545 відповідно.
8. Знаходимо квантили F-розподілу $=F.INV.RT(0,05;2;8) \approx 4,459$;
 $=F.INV.RT(0,05;4;8) \approx 3,838$.

9. Оскільки $1,545 < 3,838$, то результати вимірювання результативної ознаки не залежать від того, який метод використовувався на рівні значущості 0,05.
10. Оскільки $10,216 > 4,459$, то результативна випадкова величина залежить від значення, яке набуває фактор: А, В чи С.
11. Зауважимо, що $=F.INV.RT(0,05;2;8) \approx 8,649$, але F-статистика 10,216 більша і за цей квантиль. Отже, навіть на рівні значущості 0,01 основна гіпотеза відхиляється. Фактор таки впливає на результативну ознаку.

Література

1. Теоретико-ймовірнісні та стохастичні методи в економіці та фінансовій математиці / Леоненко М. М., Мішура Ю. С., Пархоменко В. М., Ядренко М. Й. К. : Інформтехніка, 1995. 380 с. URL: <https://book.cc/book/3130768/9a5575> (дата звернення 28.02.2022).
2. Карташов М. В. Імовірність, процеси, статистика : посібник. К. : Видавничо-поліграфічний центр «Київський університет», 2008. 494 с.
3. Турчин В. М. Математична статистика : посібник. К. : Академія, 1999. 240 с.