ВСТУП

Система символьної математики *Mathematica* виникла в результаті тривалої еволюції застосування комп'ютерів для розв'язання математичних задач. Можливість застосування чисельних методів, символьні, графічні і числові обчислення, що виконуються в одному сеансі використання системи *Mathematica*, перетворюють її в зручний і потужний інструмент математичних досліджень [3, 7]. Вважається, що *Mathematica* – лідер серед систем символьної математики.

Система імітаційного моделювання GPSS World заснована на оригінальній мові комп'ютерного моделювання GPSS (General Purpose Simulation System – загальноцільова система моделювання). З основами побудови та принципами функціонування цієї системи можна ознайомитися в навчальних посібниках [2, 5, 6, 8]. Програмні засоби GPSS World використовуються в цьому посібнику для отримання випадкових вибірок під час формулювання задач математичної статистики.

Зупинимось на описі принципів функціонування основних блоків і операторів GPSS World, а також функцій системи *Mathematica*, які використовуються в цій книзі.

Почнемо з GPSS World.

```
Tm EQU 101
```

Задається значення часу моделювання в програмному середовищі GPSS World за допомогою змінної користувача з іменем Tm.

```
GENERATE Tm
TERMINATE 1
START 1
```

Завдяки цим блокам реалізується час моделювання, заданий змінною користувача Tm. У момент часу, заданий змінною користувача Tm (тобто в момент завершення моделювання), генерується транзакт, який прямує в наступний блок. Значення 1 операнда A блоку TERMINATE задає число одиниць, на яке цей блок при входженні транзакта зменшує значення лічильника завершень (тобто значення операнда A блоку START). Таким чином, значення лічильника завершень перетворюється в 0 і моделювання припиняється.

```
Dis   FUNCTION  RN50,D4
.1,1/.3,2/.6,3/1,4
```

Наведено опис функції Dis типу D (дискретна числова), що набуває чотири значення 1, 2, 3, 4 з імовірностями 0,1; 0,2; 0,3 і 0,4 відповідно. Номер генератора випадкових чисел – 50. У другому рядку наведено список накопичених частот і відповідних значень функції.

Gis TABLE Fn$Dis,0,1,50

Задаємо параметри таблиці з іменем Gis, в якій буде представлено розподіл випадкової величини Fn$Dis (значення функції Dis). Тут 0 – верхня межа першого інтервалу, 1 – довжина кожного інтервалу, 50 – очікуване число частотних інтервалів.

```
GENERATE 1
TABULATE Gis
TERMINATE
```

Через кожну одиницю часу блоком GENERATE генерується транзакт, який обслуговує таблицю Gis. Транзакт, що увійшов до блоку TABULATE, коректує статистику таблиці Gis і знищується після вхождення до блоку TERMINATE.

(Normal(50,150,4))

Задається випадкова величина, розподілена нормально з середнім значенням 150 і середнім квадратичним відхиленням 4. Номер генератора випадкових чисел 50.

(Uniform(50,50,55))

Задається випадкова величина, розподілена рівномірно на проміжку [50, 55]. Номер генератора випадкових чисел 50.

(Exponential(50,0,10))

Задається випадкова величина, розподілена згідно з показниковим законом з середнім значенням 10. Номер генератора випадкових чисел 50.

(Binomial(50,70,0.85))

Задається випадкова величина, розподілена згідно з біномним законом відповідно до кількості $n = 70$ незалежних випробувань з імовірністю появи події $p = 0,85$ в кожному випробуванні. Номер генератора випадкових чисел 50.

(Poisson(50,5))

Задається випадкова величина, розподілена згідно з пуассонівським законом з параметром $\lambda = 5$. Номер генератора випадкових чисел 50.

Z Variable (Exponential(50,0,1))

Задаємо арифметичну змінну з іменем Z, яка розподілена згідно з показниковим законом з середнім значенням 1.

Iks Variable (V$Z+V$Z)

Задаємо арифметичну змінну з іменем Iks, яка є сумою двох значень випадкової величини Z.

TABLE	MEAN	STD.DEV.	RANGE	RETRY	FREQUENCY	CUM.%
GIS	52.420	1.327		0		
			49.000 - 50.000		11	11.00
			50.000 - 51.000		17	28.00
			51.000 - 52.000		16	44.00
			52.000 - 53.000		31	75.00
			53.000 - 54.000		25	100.00

Наведено фрагмент стандартного звіту, отриманого в результаті реалізації програми GPSS World, що містить таблицю (гістограму) з іменем GIS. Таблиця задає розподіл випадкової величини із середнім значенням 52,420 і середнім квадратичним відхиленням 1,327. У стовпці FREQUENCY наведено частоти, що відповідають кожному проміжку. Наприклад, за час моделювання число потраплянь значень випадкової величини в проміжок (49, 50] становить 11, в проміжок (50, 51] – 17 і т. д. В останньому стовпці розраховані накопичені частоти у відсотках до сумарної кількості значень табульованого аргументу.

Наведемо перелік вбудованих функцій, які використовуються в даному посібнику і призначені для виконання операцій в системі *Mathematica*:

`Sqrt[x]` – обчислення квадратного кореня числа **`x`**;

`Times[x,y]` – обчислення добутку чисел **`x`** і **`y`**;

`Log[a,x]` – обчислення логарифма з основою **`a`** числа **`x`**;

`Binomial[n,k]` – обчислення біномних коефіцієнтів C_n^k;

`Gamma[z]` – ейлерова гамма-функція;

`Pi` – число π;

`N[k]` – представлення **`k`** у вигляді дійсного числа;

`Sum[f,i,{imin,imax}]` – обчислення суми **`f`** по **`i`** від **`imin`** до **`imax`**;

`Integrate[f,{z,0,t}]` – обчислення визначеного інтеграла $\int_0^t f(z)dz.$

`Solve[eq,var]` – розв’язує рівняння **`eq`** відносно змінної **`var.`** Рівняння має бути у вигляді **`lh==rh.`**

`FindRoot[lp==rp,{x,x0}]` – знаходить чисельний розв’язок рівняння **`lp=rp`** в околі точки **`x0.`**

`func[x_]:=g(x)` – визначення функції користувача з іменем **`func`** з використанням операції відкладеного присвоювання **`:=`** (тут **`g(x)`** – це явний вираз, який містить змінну **`x`**).

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №1
Числові характеристики статистичного розподілу вибірки
ЗАДАЧА 1.1
Числові характеристики дискретного статистичного розподілу вибірки

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 101  ; час моделювання
Rozp FUNCTION RNk,D5
.14,k/.35,k+1/.58,k+2/.8,k+3/1,k+4  ;випадкова величина
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, взяти за значення x_i праві кінці інтервалів $(z_{i-1}, z_i]$ і записати дискретний статистичний розподіл частот для отриманої вибірки та обчислити такі числові характеристики цього розподілу:

1) вибіркове середнє, 2) вибіркову дисперсію, 3) підправлену дисперсію, 4) вибіркове середнє квадратичне відхилення, 5) підправлене середнє квадратичне відхилення, 6) коефіцієнт варіації, 7) коефіцієнт асиметрії та 8) ексцес. Обчислення здійснювати за допомогою пакету *Mathematica*.

РОЗВ'ЯЗАННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
Rozp FUNCTION RN50,D5
.14,50/.35,51/.58,52/.8,53/1,54  ;випадкова величина
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE              MEAN    STD.DEV.       RANGE            RETRY  FREQUENCY  CUM.%
 GIS              52.420    1.327                             0
                                     49.000  -       50.000                 11     11.00
                                     50.000  -       51.000                 17     28.00
                                     51.000  -       52.000                 16     44.00
                                     52.000  -       53.000                 31     75.00
                                     53.000  -       54.000                 25    100.00
```

Користуючись гістограмою (стовпцями, виділеними жирним шрифтом), записуємо статистичний розподіл вибірки у вигляді дискретного статистичного розподілу частот:

x_i	50	51	52	53	54
n_i	11	17	16	31	25

Далі обчислимо числові характеристики статистичного розподілу вибірки.

1) Вибіркове середнє обчислюємо за формулою

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{n_1x_1 + n_2x_2 + ... + n_m x_m}{n}.$$

Тут $n = \sum_{i=1}^{m} n_i$ – обсяг вибірки. Для даного варіанта №50 маємо $m = 5,\ n = 100.$

2) Вибіркову дисперсію обчислюємо за формулою

$$D_{Â} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - (\overline{x})^2.$$

3) Підправлену дисперсію обчислюємо за формулою

$$s^2 = \frac{n}{n-1} D_{Â}.$$

4) Вибіркове середнє квадратичне відхилення обчислюємо за формулою

$$\sigma_B = \sqrt{D_{Â}}\,.$$

5) Підправлене середнє квадратичне відхилення обчислюємо за формулою

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

6) Коефіцієнт варіації обчислюємо за формулою

$$V = \frac{s}{\overline{x}} \cdot 100\% = \frac{1{,}9}{8{,}15} \cdot 100\% = 23{,}3\%.$$

7) Коефіцієнт асиметрії обчислюємо за формулою

$$A_B = \frac{m_3}{\sigma_B^3}.$$

Тут m_3 – центральний емпіричний момент 3-го порядку, $m_3 = \frac{1}{n}\sum_{i=1}^{m} (x_i - \overline{x})^3 n_i.$

8) Ексцесом E_B статистичного розподілу вибірки називається число, яке обчислюється за формулою:

$$E_B = \frac{m_4}{\sigma_B^4} - 3.$$

Тут m_4 – центральний емпіричний момент 4-го порядку, $m_4 = \frac{1}{n}\sum_{i=1}^{m} (x_i - \overline{x})^4 n_i.$

Наведемо повністю послідовність обчислень для розв'язання даної задачі з використанням програмного середовища *Mathematica*:

```
Варіанти (значення x[i]):
 x[1]:=50
 x[2]:=51
 x[3]:=52
 x[4]:=53
 x[5]:=54
Частоти:
 n[1]:=11
 n[2]:=17
 n[3]:=16
 n[4]:=31
 n[5]:=25
Вибіркове середнє:
 xav:=N[Sum[n[i]x[i],{i,1,5}]/100]
 xav
 52.42
Вибіркова дисперсія:
 Dv:=N[Sum[n[i](x[i]^2),{i,1,5}]/100-(xav)^2]
 Dv
 1.7436
Підправлена вибіркова дисперсія:
 skv:=100(Dv)/99
 skv
 1.76121
Підправлене середнє квадратичне відхилення:
 s:=Sqrt[skv]
 s
 1.32711
Вибіркове середнє квадратичне відхилення:
 sigv:=Sqrt[Dv]
 sigv
 1.32045
Коефіцієнт варіації:
 V:=100s/xav
 V
 2.53168
Центральний емпіричний момент 3-го порядку:
 m3:=N[Sum[n[i](x[i]-xav)^3,{i,1,5}]/100]
 m3
 -1.01102
Коефіцієнт асиметрії:
 Av:=(m3)/(sigv^3)
 Av
```

```
-0.439128
```

Центральний емпіричний момент 4-го порядку:

```
m4:=N[Sum[n[i](x[i]-xav)^4,{i,1,5}]/100]
m4
6.06198
```

Ексцес:

```
Ev:=(m4)/(sigv^4)-3
Ev
-1.00602
```

ЗАДАЧА 1.2
Інтервальні статистичні оцінки математичного сподівання довільно розподіленої випадкової величини

Знайти інтервал довіри для оцінки з надійністю γ невідомого математичного сподівання *a* ознаки *X* генеральної сукупності, розглянутої в задачі 1.1, для випадків:

1) γ=0,9; генеральне середнє квадратичне відхилення σ=1,332; обсяг вибірки n=100; значення вибіркового середнього $\overline{x}$ взяти з результатів розв'язання задачі 1.1;

2) γ=0,9; генеральне середнє квадратичне відхилення σ невідоме; обсяг вибірки n=100; значення вибіркового середнього $\overline{x}$ і підправленого середнього квадратичного відхилення *s* взяти з результатів розв'язання задачі 1.1.

Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай за результатами n незалежних спостережень за випадковою величиною X необхідно визначити інтервал довіри для невідомого параметра $a = E(X)$ (математичного сподівання).

Оскільки для математичного сподівання точковою оцінкою є вибіркове середнє $\overline{x}$, то для знаходження інтервалу довіри

$$\overline{x} - \delta < a < \overline{x} + \delta$$

необхідно розв'язати рівняння

$$\mathrm{P}\{\overline{x} - \delta < a < \overline{x} + \delta\} = \gamma$$

відносно точності оцінки δ, де $\gamma -$ задана надійність оцінки.

Якщо середнє квадратичне відхилення σ випадкової величини X відоме, $X -$ нормально розподілена випадкова величина або обсяг вибірки значний ($n > 30$), то можна записати, що

$$P\left\{|\,\overline{x} - a\,| < \frac{\sigma}{\sqrt{n}}t\right\} = 2\Phi(t) = \gamma, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx,$$

де $\Phi(t) -$ функція Лапласа. Тоді, якщо $t = t_\gamma -$ розв'язок рівняння

$$2\Phi(t) = \gamma\,, \tag{1.2}$$

то з надійністю γ інтервал

$$\overline{x} - \frac{\sigma}{\sqrt{n}}t_\gamma < a < \overline{x} + \frac{\sigma}{\sqrt{n}}t_\gamma \tag{1.3}$$

є інтервалом довіри для математичного сподівання a.

Якщо середнє квадратичне відхилення σ випадкової величини X невідоме, але обсяг вибірки значний ($n > 30$), то інтервал довіри можна записати у вигляді

$$\overline{x} - \frac{s}{\sqrt{n}} t_{\gamma} < a < \overline{x} + \frac{s}{\sqrt{n}} t_{\gamma}, \tag{1.4}$$

де $s-$ підправлене середнє квадратичне відхилення, знайдене за даними вибірки обсягу *n*.

РОЗВ’ЯЗАННЯ для k=50

Для випадків 1, 2 інтервали довіри шукаємо у вигляді (1.3) і (1.4) відповідно. Розв’язування рівняння (1.2) здійснюється за допомогою вбудованої функції **`Solve[eq,var],`** яка дає змогу розв’язати рівняння **`eq`** відносно змінної **`var.`** Рівняння (1.2) спочатку представляємо у вигляді

$$\frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz = \gamma.$$

Інтегрування здійснюється за допомогою вбудованої функції для обчислення визначеного інтеграла **`Integrate[f,{z,0,t}].`**

```
Розв’язування рівняння (1.2)
 Tgam[x_]:=Solve[2(Integrate[Exp[-
(z^2)/2],{z,0,t}]/(Sqrt[2Pi]))==x,t]
 Tgam[0.9]
 {t→1.64485}
 xav:=52.42
 sig:=1.332
 t:=1.6448536269514726`
 n:=100
Ліва межа інтервалу довіри для випадку 1:
 xav-(sig)t/Sqrt[n]
 52.2009
Права межа інтервалу довіри для випадку 1:
 xav+(sig)t/Sqrt[n]
 52.6391
Випадок 2
Значення t залишається з випадку 1.
 s:=1.3271066728834977
Ліва межа інтервалу довіри для випадку 2:
 xav-(s)t/Sqrt[n]
 52.2017
Права межа інтервалу довіри для випадку 2:
 xav+(s)t/Sqrt[n]
 52.6383
```

ЗАДАЧА 1.3
Інтервальні статистичні оцінки математичного сподівання нормально розподіленої випадкової величини для вибірки малого обсягу

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 21  ; час моделювання
Gis TABLE (Normal(k,100+k,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати інтервальний статистичний розподіл частот для отриманої вибірки з нормальної генеральної сукупності. Взявши за x_i середини інтервалів, записати дискретний статистичний розподіл частот та обчислити такі числові характеристики цього розподілу: 1) вибіркове середнє, 2) вибіркову дисперсію, 3) підправлену дисперсію, 4) вибіркове середнє квадратичне відхилення, 5) підправлене середнє квадратичне відхилення. Знайти інтервал довіри для оцінки з надійністю γ=0,95 невідомого математичного сподівання a нормально розподіленої ознаки X генеральної сукупності, вважаючи, що середнє квадратичне відхилення σ невідоме. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Якщо середнє квадратичне відхилення σ невідоме, обсяг вибірки невеликий ($n \le 30$), але $X-$ нормально розподілена випадкова величина, то інтервал довіри для математичного сподівання записують у вигляді (1.4), де значення $t_\gamma = t(\gamma, n)$ шукають як розв'язок рівняння

$$\mathrm{P}\{|T| < t_\gamma\} = 2\int_0^{t_\gamma} s(x,n)\,dx = \gamma. \tag{1.5}$$

Тут $T = \dfrac{\overline{x} - a}{s/\sqrt{n}} -$ випадкова величина, розподілена за законом Ст'юдента з $k = n-1$ ступенями вільності, який характеризується щільністю розподілу

$$s(x,n) = B_n\left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \qquad x \in (-\infty, +\infty),$$

де $B_n = \dfrac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)}$, $\Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt.$

РОЗВ'ЯЗАННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 21  ; час моделювання
Gis TABLE (Normal(50,150,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент, що містить гістограму:

```
TABLE    MEAN     STD.DEV.       RANGE        RETRY FREQUENCY   CUM.%
 GIS     150.144  3.477                         0
                           144.000 - 146.000          3       15.00
                           146.000 - 148.000          3       30.00
                           148.000 - 150.000          6       60.00
                           150.000 - 152.000          1       65.00
                           152.000 - 154.000          4       85.00
                           154.000 - 156.000          3      100.00
```

Записуємо інтервальний статистичний розподіл частот для отриманої вибірки:

$(z_{i-1}, z_i]$	[144,146]	(146,148]	(148,150]	(150,152]	(152,154]	(154,156]
n_i	3	3	6	1	4	3

Взявши за x_i середини проміжків, записуємо дискретний статистичний розподіл частот:

x_i	145	147	149	151	153	155
n_i	3	3	6	1	4	3

Числові характеристики розподілу обчислюємо за формулами (1.1), враховуючи, що $m = 6$, $n = 20$. Інтервал довіри шукаємо у вигляді (1.4), де значення $t_\gamma = t(\gamma, n)$ визначаємо як розв'язок рівняння (1.5). Розв'язування рівняння (1.5) здійснюється за допомогою вбудованої функції

`FindRoot[lp==rp,{x,x0}],`

яка дає змогу знайти чисельний розв'язок рівняння **`lp=rp`** в околі точки **`x0.`** Для визначення щільності розподілу Ст'юдента використовується вбудована функція **`Gamma[x].`**

```
Варіанти (значення x[i]):
 x[1]:=145
 x[2]:=147
 x[3]:=149
 x[4]:=151
 x[5]:=153
 x[6]:=155
Частоти:
 n[1]:=3
 n[2]:=3
 n[3]:=6
 n[4]:=1
 n[5]:=4
 n[6]:=3
Вибіркове середнє:
 xav:=N[Sum[n[i]x[i],{i,1,6}]/20]
 xav
 149.9
Вибіркова дисперсія:
 Dv:=N[Sum[n[i](x[i]^2),{i,1,6}]/20-(xav)^2]
 Dv
 10.99
Підправлена дисперсія:
 skv:=20(Dv)/19
 skv
 11.5684
Підправлене середнє квадратичне відхилення:
 s:=Sqrt[skv]
 s
 3.40124
Вибіркове середнє квадратичне відхилення:
 sigv:=Sqrt[Dv]
 sigv
 3.31512
Розподіл Ст'юдента:
 B[n_]:=Gamma[n/2]/(Sqrt[Pi(n-1)]Gamma[(n-1)/2])
 S[x_,n_]:=B[n](1+(x^2)/(n-1))^(-n/2)
Визначення t(gam,n)(розв'язування рівняння (1.5))
Tgamn[gam_,n_]:=FindRoot[2(Integrate[S[t,n],{t,0,x}])==gam
,{x,2}]
 Tgamn[0.95,20]
 {x→2.09302}
 t:=2.09302405440831l`
Ліва межа інтервалу довіри:
 xav-(s)t/Sqrt[20]
```

148.308

Права межа інтервалу довіри:

```
xav+(s)t/Sqrt[20]
```

151.492

ЗАДАЧА 1.4
Інтервали довіри для оцінки середнього квадратичного відхилення нормально розподіленої випадкової величини

Знайти інтервали довіри для оцінки з надійністю 1) γ=0,95; 2) γ=0,99 невідомого середнього квадратичного відхилення σ нормально розподіленої ознаки X генеральної сукупності за вибіркою із задачі 1.3. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Якщо X – нормально розподілена випадкова величина, то інтервал довіри для середнього квадратичного відхилення записують у вигляді

$$s(1-q) < \sigma < s(1+q), \qquad (1.7)$$

де s – підправлене середнє квадратичне відхилення, обчислене за даними вибірки обсягу n, q – розв'язок рівняння

$$\int_{\chi_1(q)}^{\chi_2(q)} R(x,n)\,dx = \gamma, \qquad \chi_1 = \sqrt{n-1}/(1+q), \;\; \chi_2 = \sqrt{n-1}/(1-q). \qquad (1.8)$$

Тут $R(x,n)$ – щільність розподілу випадкової величини $\chi = \sqrt{n-1}\,s/\sigma$,

$$R(x,n) = \frac{x^{n-2}e^{-x^2/2}}{2^{(n-3)/2}\Gamma((n-1)/2)}.$$

РОЗВ'ЯЗАННЯ для k=50

Інтервали довіри шукаємо у вигляді (1.7). Розв'язування рівняння (1.8) здійснюється за допомогою вбудованої функції

FindRoot[lp==rp,{x,x0}].

```
Розподіл Xi:
 R[x_,n_]:=(x^(n-2)Exp[-(x^2)/2])/((2^((n-3)/2))Gamma[(n-
1)/2])
Визначення q(gam,n) (Розв'язування рівняння (1.8))
 Qgaman[gam_,n_]:=FindRoot[Integrate[R[t,n],{t,Sqrt[n-
1]/(1+q),Sqrt[n-1]/(1-q)}]==gam,{q,0}]
 Qgaman[0.95,20]
 {q→0.37108}
 q:=0.3710804861306243`
 s:=3.401238164643787`
Випадок 1 (gam=0.95)
Ліва межа інтервалу довіри:
 s(1-q)
 2.13911
Права межа інтервалу довіри:
```

```
 s(1+q)
 4.66337
Випадок 2 (gam=0.99)
 Qgaman[0.99,20]
 {q →0.577745}
 q:=0.5777451743871259`
Ліва межа інтервалу довіри:
 s(1-q)
 1.43619
Права межа інтервалу довіри:
 s(1+q)
 5.36629
```

ЗАДАЧА 1.5
Вибіркове рівняння прямої регресії *Y* на *X*.

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 1001  ; час моделювання
Iks Variable (Uniform(k,-1,1))
GisX TABLE V$Iks,-1,0.2,500  ; гістограма для X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, і взявши за значення x_i середини інтервалів, записати дискретні статистичні розподіли частот для вибірки з генеральних сукупностей X, $Y=X^3$. Обчислити вибірковий коефіцієнт кореляції $r(X,Y)$, коефіцієнти α, β та записати вибіркове рівняння прямої регресії Y на X: $y=\alpha x+\beta$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)-$ результати n випробувань, які відповідають значенням випадкових величин X і Y. Вибірковий коефіцієнт кореляції між випадковими величинами X і Y обчислюємо за формулою

$$r(X,Y)=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\cdot\overline{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n\cdot\overline{x}^2\right)\left(\sum_{i=1}^{n}y_i^2-n\cdot\overline{y}^2\right)}}, \qquad (1.9)$$

де

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \qquad \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i.$$

Якщо $|r(X,Y)|\approx 1$, то можна розглядати вибіркове рівняння прямої регресії Y на X у вигляді $y=\alpha x+\beta$, де

$$\alpha=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}, \qquad \beta=\overline{y}-\alpha\overline{x}. \qquad (1.10)$$

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 1001  ; час моделювання
Iks Variable (Uniform(50,-1,1))
GisX TABLE V$Iks,-1,0.2,500  ; гістограма для X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE              MEAN    STD.DEV.       RANGE           RETRY FREQUENCY CUM.%
 GISX              0.010    0.573                           0
                                    -1.000  -      -0.800             83     8.30
                                    -0.800  -      -0.600            103    18.60
                                    -0.600  -      -0.400            113    29.90
                                    -0.400  -      -0.200            100    39.90
                                    -0.200  -      -0.000            105    50.40
                                    -0.000  -       0.200             90    59.40
                                     0.200  -       0.400            104    69.80
                                     0.400  -       0.600             93    79.10
                                     0.600  -       0.800            106    89.70
                                     0.800  -       1.000            103   100.00
```

Записуємо дискретний статистичний розподіли частот для вибірки з генеральних сукупностей *X*, *Y*=X^3:

x_i	–0,9	–0,7	–0,5	–0,3	–0,1	0,1	0,3	0,5	0,7	0,9
$y_i = x_i^3$	$(-0{,}9)^3$	$(-0{,}7)^3$	$(-0{,}5)^3$	$(-0{,}3)^3$	$(-0{,}1)^3$	$0{,}1^3$	$0{,}3^3$	$0{,}5^3$	$0{,}7^3$	$0{,}9^3$
n_i	83	103	113	100	105	90	104	93	106	103

Користуючись даними таблиці і враховуючи, що $y_i = x_i^3$, здійснюємо обчислення за формулами (1.9), (1.10).

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №2
Перевірка статистичних гіпотез про закон розподілу. Критерій згоди Пірсона

ЗАДАЧА 2.1
Перевірка статистичної гіпотези про нормальний закон розподілу

Реалізувати програму GPSS World, спочатку підставивши замість *k* свій № варіанту:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Normal(k,100+k,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати інтервальний статистичний розподіл частот для отриманої вибірки, обчислити теоретичні частоти для нормального розподілу з параметрами a=100+k (k – номер варіанту), σ=4, та для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу H_0: *випадкова величина X розподілена нормально*. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай висунуто статистичну гіпотезу H_0: *випадкова величина X розподілена згідно з законом A*. Для перевірки гіпотези здійснюють n випробувань і записують інтервальний статистичний розподіл частот отриманої вибірки:

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$	…	$[z_{m-1}, z_m)$
n_i	n_1	n_2	…	n_m

Оскільки перевіряється гіпотеза про те, що розподіл ознаки X генеральної сукупності описується певною (конкретною) функцією розподілу або, що те ж саме, щільністю розподілу $p(x)$, то для кожного інтервалу можна визначити теоретичні ймовірності потрапляння значень випадкової величини X в цей інтервал, а отже, й теоретичні частоти $n_i' = np_i$.

Для обчислення ймовірностей p_i використовують формули:

$$p_i = P\{z_{i-1} \le X < z_i\} = F(z_i) - F(z_{i-1}) = \int_{z_{i-1}}^{z_i} p(x)\,dx, \quad 2 \le i \le m. \tag{2.1}$$

Зазначимо, що для обчислення ймовірностей p_1 і p_m в формулах (2.1) приймають $z_0 = -\infty$ і $z_m = +\infty$. Тоді $\sum_{i=1}^{m} p_i = 1$. У випадку нормального розподілу з параметрами a і σ

$$\begin{gathered} p_1 = 0,5 + \Phi\left(\frac{z_1 - a}{\sigma}\right), \quad p_m = 0,5 - \Phi\left(\frac{z_{m-1} - a}{\sigma}\right), \\ p_i = \Phi\left(\frac{z_i - a}{\sigma}\right) - \Phi\left(\frac{z_{i-1} - a}{\sigma}\right), \qquad 2 \le i \le m-1, \end{gathered} \tag{2.2}$$

де $\Phi(x)-$ функція Лапласа.

Згідно з критерієм Пірсона для перевірки гіпотези H_0 вводиться випадкова величина (статистика)

$$K = \chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'},$$

де $m-$ число груп (інтервалів) в статистичному розподіл вибірки; n_i- емпірична частота ознаки X в i-й групі; $n_i' = np_i-$ теоретична частота; p_i- ймовірність того, що значення X належить i-й групі.

Відомо, що при $n \to \infty$ закон розподілу статистики K наближається до закону розподілу χ^2 з $k = m - r - 1$ ступенями вільності, де $r-$ кількість параметрів гіпотетичного розподілу A (наприклад, $r = 2$ для нормального розподілу, $r = 1$ для розподілу Пуассона, $r = 0$ для рівномірного розподілу).

Для критерію χ^2 будують правосторонню критичну область за правилом

$$\mathrm{P}\{\chi^2 > \chi^2_{cr}\} = \alpha. \tag{2.3}$$

За заданим рівнем значущості α і кількістю ступенів вільності k знаходять критичну точку $k_{cr} = \chi^2_{cr}(\alpha, k)$ як розв’язок рівняння (2.3).

Використовуючи щільність розподілу χ^2

$$f(x,k) = \frac{x^{(k/2)-1} e^{-x/2}}{2^{k/2}\Gamma(k/2)}, \qquad x > 0,$$

рівняння (2.3) можна записати у вигляді

$$1 - \int_0^{k_{cr}} f(x,k)dx = \alpha. \tag{2.4}$$

На основі даних вибірки обчислюють емпіричне значення критерію Пірсона

$$K_{emp} = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'}. \tag{2.5}$$

Порівнюємо значення K_{emp} і k_{cr}: Якщо $K_{emp} \ge k_{cr}$, то гіпотезу H_0 відхиляють; якщо ж $K_{emp} < k_{cr}$, то гіпотезу H_0 приймають.

Застосування критерію χ^2 вимагає дотримання наступних умов: 1) експериментальні дані повинні бути незалежними, тобто вибірка повинна бити випадковою; 2) обсяг вибірки повинен бити досить великим (практично не меншим, ніж 50 одиниць), а частота кожної групи – не менше 5. Якщо остання умова не виконується, то проводиться попереднє об'єднання малочисельних груп.

РОЗВ'ЯЗАННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Normal(50,150,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE               MEAN    STD.DEV.       RANGE            RETRY FREQUENCY CUM.%
 GIS             149.602    3.627                              0
                                  140.000  -      142.000              1     1.00
                                  142.000  -      144.000              7     8.00
                                  144.000  -      146.000              9    17.00
                                  146.000  -      148.000             15    32.00
                                  148.000  -      150.000             25    57.00
                                  150.000  -      152.000             11    68.00
                                  152.000  -      154.000             21    89.00
                                  154.000  -      156.000              9    98.00
                                  156.000  -      158.000              2   100.00
```

Користуючись гістограмою, записуємо статистичний розподіл вибірки у вигляді інтервального статистичного розподілу частот, попередньо об'єднавши інтервали, для яких $n_i < 5$, з сусідніми:

$(z_{i-1}, z_i]$	[140,144]	(144,146]	(146,148]	(148,150]	(150,152]	(152,154]	(154,158]
n_i	8	9	15	25	11	21	11
$n_i' = np_i$	6,68072	9,18481	14,9882	19,1462	19,1462	14,9882	15,8655

Оскільки кількість інтервалів $m = 7$ і для нормального розподілу кількість параметрів $r = 2$, то кількість ступенів вільності розподілу χ^2 становить: $k = m - r - 1 = 7 - 2 - 1 = 4$. Критичну точку знаходимо, розв'язавши рівняння (2.4), а емпіричне значення критерію обчислюємо за формулою (2.5):

$$k_{cr} = 9{,}48773, \quad K_{emp} = 9{,}42344.$$

Оскільки $K_{emp} < k_{cr}$, то гіпотезу H_0 приймаємо.

```
Розподіл Xi квадрат
 f[x_,k_]:=(x^(k/2-1)Exp[-x/2])/((2^(k/2))Gamma[k/2])
Знаходження критичної точки
 Xicr[alpha_,k_]:=FindRoot[1-
Integrate[f[t,k],{t,0,x}]==alpha,{x,20}]
 k=m-r-1=7-2-1=4
Значення критичної точки:
 Xicr[0.05,4]
 {x→9.48773}
Емпіричні частоти:
 n[1]:=8
 n[2]:=9
 n[3]:=15
 n[4]:=25
 n[5]:=11
 n[6]:=21
 n[7]:=11
Перевірка
 Sum[n[i],{i,1,7}]
 100
 a:=150
Значення z[i]:
 z[1]:=144
 z[2]:=146
 z[3]:=148
 z[4]:=150
 z[5]:=152
 z[6]:=154
Функція Лапласа
 Phi[x_]:=NIntegrate[Exp[-(z^2)/2],{z,0,x}]/(Sqrt[2Pi])
Визначення теоретичних частот
 np[1]:=(0.5+Phi[(z[1]-a)/4])100
 np[i_]:=(Phi[(z[i]-a)/4]-Phi[(z[i-1]-a)/4])100
 np[7]:=(0.5-Phi[(z[6]-a)/4])100
 np[1]
 6.68072
 np[2]
 9.18481
 np[3]
 14.9882
 np[4]
 19.1462
 np[5]
 19.1462
 np[6]
 14.9882
 np[7]
 15.8655
Перевірка
 Sum[np[i],{i,1,7}]/100
 1.
```

Емпіричне значення критерію

```
Kemp:=N[Sum[((n[i]-np[i])^2)/np[i],{i,1,7}]]
Kemp
9.42344
```

Kemp<Xicr, гіпотезу приймаємо.

ЗАДАЧА 2.2
Перевірка статистичної гіпотези про рівномірний закон розподілу

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 201  ; час моделювання
Gis TABLE (Uniform(k,k,k+5)),0,0.5,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати інтервальний статистичний розподіл частот для отриманої вибірки, обчислити теоретичні частоти для рівномірного розподілу на проміжку [k,k+5] (k – номер варіанту), та для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу H_0: *випадкова величина X розподілена рівномірно на проміжку* [k, k+5]. Обчислення здійснювати за допомогою пакету *Mathematica*.

Вказівка. У випадку рівномірного розподілу на проміжку [k,k+5] маємо $r = 0$, а ймовірності p_i обчислюємо за формулами

$$p_i = P\{z_{i-1} < X \le z_i\} = \int_{z_{i-1}}^{z_i} p(x)\,dx = \frac{z_i - z_{i-1}}{(k+5)-k} = \frac{z_i - z_{i-1}}{5}, \quad i = \overline{1,m}.$$

Довжини інтервалів $[z_{i-1}, z_i)$, для яких $n_i \ge 5$, для даної вибірки дорівнюють 0,5. Якщо ж для деякого інтервала $n_i < 5$, то його необхідно об'єднати з сусіднім. Обсяг вибірки визначаємо за формулою $n = \sum_{i=1}^{m} n_i$. Критичну точку знаходимо, розв'язавши рівняння (2.4), а емпіричне значення критерію обчислюємо за формулою (2.5).

ЗАДАЧА 2.3

Перевірка статистичної гіпотези про показниковий закон розподілу

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 1001  ; час моделювання
Gis TABLE (Exponential(k,0,10)),0,5,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати інтервальний статистичний розподіл частот для отриманої вибірки, обчислити теоретичні частоти для показникового розподілу з параметром λ=0,1, та для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу H_0: *випадкова величина X розподілена за показниковим законом з параметром* λ=0,1. Обчислення здійснювати за допомогою пакету *Mathematica*.

Вказівка. У випадку показникового розподілу з параметром λ маємо $r = 1$, а ймовірності p_i обчислюємо за формулами

$$p_i = P\{z_{i-1} < X \le z_i\} = \int_{z_{i-1}}^{z_i} \lambda e^{-\lambda x}\, dx = e^{-\lambda z_{i-1}} - e^{-\lambda z_i}, \quad i = \overline{1,m}; \;\; z_m = \infty.$$

Інтервали, для яких $n_i < 5$, необхідно об'єднати з сусідніми, щоб для кожного інтервала виконувалась умова $n_i \ge 5$. Наприклад, якщо зі стандартного звіту отримано гістограму

```
TABLE              MEAN    STD.DEV.       RANGE           RETRY FREQUENCY CUM.%
 GIS               9.837   10.132                           0
                                    0.000  -       5.000              401    40.10
                                    5.000  -      10.000              239    64.00
                                   10.000  -      15.000              146    78.60
                                   15.000  -      20.000               94    88.00
                                   20.000  -      25.000               50    93.00
                                   25.000  -      30.000               27    95.70
                                   30.000  -      35.000               11    96.80
                                   35.000  -      40.000               10    97.80
                                   40.000  -      45.000                5    98.30
                                   45.000  -      50.000                5    98.80
                                   50.000  -      55.000                4    99.20
                                   55.000  -      60.000                3    99.50
                                   60.000  -      65.000                3    99.80
                                   65.000  -      70.000                1    99.90
                                   70.000  -      75.000                1   100.00
```

то після об'єднання інтервалів отримуємо такий розподіл:

$(z_{i-1}, z_i]$	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
n_i	401	239	146	94	50	27	11
$(z_{i-1}, z_i]$	(35,40]	(40,45]	(45,50]	(50,60]	(60,75]	-	-
n_i	10	5	5	7	5	-	-

Критичну точку знаходимо, розв'язавши рівняння (2.4), а емпіричне значення критерію обчислюємо за формулою (2.5).

ЗАДАЧА 2.4
Перевірка статистичної гіпотези про біномний закон розподілу

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 71  ; час моделювання
Gis TABLE (Binomial(k,70,0.85)),0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати дискретний статистичний розподіл частот для отриманої вибірки, взявши за значення x_i **праві кінці** інтервалів $(z_{i-1}, z_i]$, обчислити теоретичні частоти для біномного розподілу з кількістю випробувань *n*=70 (рівному обсягу вибірки), імовірністю успішного випробування *p*=0,85, та для рівня значущості $\alpha = 0{,}001 \cdot k$ (*k* – номер варіанту) перевірити гіпотезу H_0: *випадкова величина X розподілена за біномним законом для n*=70 *і p*=0,85. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай висунуто статистичну гіпотезу H_0: *випадкова величина X розподілена згідно з біномним законом.* Для біномного розподілу кількість параметрів $r = 0$, а ймовірності p_i обчислюємо за формулами

$$p_i = \mathrm{P}\{z_{i-1} \le X \le z_i\} = \sum_{k=z_{i-1}}^{z_i} P_n(k), \quad p_i = \mathrm{P}\{X = z_i\} = P_n(z_i), \quad 1 \le i \le m; \tag{2.8}$$
$$P_n(k) = C_n^k p^k (1-p)^{n-k}.$$

Критичну точку знаходимо, розв'язавши рівняння (2.4), а емпіричне значення критерію обчислюємо за формулою (2.5).

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 71  ; час моделювання
Gis TABLE (Binomial(50,70,0.85)),0,1,500  ; гістограма
GENERATE 1
```

```
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE                  MEAN    STD.DEV.       RANGE            RETRY FREQUENCY  CUM.%
 GIS                 59.414    3.360                              0
                                    51.000  -       52.000               2      2.86
                                    52.000  -       53.000               3      7.14
                                    53.000  -       54.000               1      8.57
                                    54.000  -       55.000               2     11.43
                                    55.000  -       56.000               5     18.57
                                    56.000  -       57.000               8     30.00
                                    57.000  -       58.000               3     34.29
                                    58.000  -       59.000              10     48.57
                                    59.000  -       60.000               9     61.43
                                    60.000  -       61.000               8     72.86
                                    61.000  -       62.000               7     82.86
                                    62.000  -       63.000               5     90.00
                                    63.000  -       64.000               3     94.29
                                    64.000  -       65.000               3     98.57
                                    65.000  -       66.000               0     98.57
                                    66.000  -       67.000               0     98.57
                                    67.000  -       68.000               1    100.00
```

Записуємо дискретний статистичний розподіл частот, який після об'єднання значень з частотами $n_i < 5$ має вигляд:

x_i	[0,53]	[54,56]	[57,58]	59	60	61	62	63	[64,70]
n_i	5	8	11	10	9	8	7	5	7

У випадку гіпотетичного біномного розподілу $r = 0$. Кількість інтервалів $m = 9$, кількість ступенів вільності для розподілу χ^2 дорівнює $k = m - r - 1 = 9 - 0 - 1 = 8$, обсяг вибірки $n = 70$, а ймовірності p_i для біномного розподілу обчислюємо за формулами

$$p_i = P\{z_{i-1} \le X \le z_i\} = \sum_{k=z_{i-1}}^{z_i} P_n(k),\ \ p_i = P\{X = k\} = P_n(k),\ P_n(k) = C_n^k p^k (1-p)^{n-k}.$$

Перевірка гіпотези здійснюється за тим самим алгоритмом, що використовувався у задачах 2.1-2.3.

```
Розподіл Xi квадрат
 f[x_,k_]:=(x^(k/2-1)Exp[-x/2])/((2^(k/2))Gamma[k/2])
Знаходження критичної точки
 Xicr[alpha_,k_]:=FindRoot[1-
Integrate[f[t,k],{t,0,x}]==alpha,{x,20}]
 k=m-r-1=9-0-1=8
Значення критичної точки:
 Xicr[0.05,8]
 {x→15.5073}
Емпіричні частоти:
 n[1]:=5
 n[2]:=8
```

```
 n[3]:=11
 n[4]:=10
 n[5]:=9
 n[6]:=8
 n[7]:=7
 n[8]:=5
 n[9]:=7
Перевірка
 Sum[n[i],{i,1,9}]
 70
 p:=0.85
 P[k_]:=N[Binomial[70,k](p^k)(1-p)^(70-k)]
Теоретичні частоти:
 np[1]:=N[Sum[P[k],{k,0,53}]70]
 np[1]
 1.93422
 np[2]:=N[Sum[P[k],{k,54,56}]70]
 np[2]
 9.06796
 np[3]:=N[Sum[P[k],{k,57,58}]70]
 np[3]
 13.9185
 np[4]:=P[59]70
 np[4]
 8.9752
 np[5]:=P[60]70
 np[5]
 9.32424
 np[6]:=P[61]70
 np[6]
 8.66186
 np[7]:=P[62]70
 np[7]
 7.12508
 np[8]:=P[63]70
 np[8]
 5.12704
 np[9]:=N[Sum[P[k],{k,64,70}]70]
 np[9]
 5.86591
Перевірка
 Sum[np[i],{i,1,9}]/70
 1.
 Kemp:=N[Sum[((n[i]-np[i])^2)/np[i],{i,1,9}]]
 Kemp
 6.00052
Kemp<Xicr, гіпотезу приймаємо.
```

ЗАДАЧА 2.5
Перевірка статистичної гіпотези про закон розподілу Пуассона

Реалізувати програму GPSS World, спочатку підставивши замість k свій

№ варіанту:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Poisson(k,5)),0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати дискретний статистичний розподіл частот для отриманої вибірки, взявши за значення x_i **праві кінці** інтервалів $(z_{i-1}, z_i]$, обчислити теоретичні частоти для розподілу Пуассона з параметром λ=5. Для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу H_0: *випадкова величина X розподілена за законом Пуассона з параметром* λ=5. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай висунуто статистичну гіпотезу H_0: *випадкова величина X розподілена згідно з законом Пуассона.* Для пуассонівського розподілу кількість параметрів $r = 1$, а ймовірності p_i обчислюємо за формулами

$$p_i = \mathrm{P}\{z_{i-1} \le X \le z_i\} = \sum_{k=z_{i-1}}^{z_i} P_n(k), \quad p_i = \mathrm{P}\{X = z_i\} = P_n(z_i),\ 1 \le i \le m-1;$$
$$p_m = \mathrm{P}\{z_m \le X < \infty\} = 1 - \sum_{i=0}^{m-1} P_n(z_i); \qquad P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.9}$$

Критичну точку знаходимо, розв’язавши рівняння (2.4), а емпіричне значення критерію обчислюємо за формулою (2.5).

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Poisson(50,5)),0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE                MEAN    STD.DEV.       RANGE            RETRY FREQUENCY CUM.%
```

```
GIS                4.990   2.163                        0
                                0.000  -     1.000          7    7.00
                                1.000  -     2.000          9   16.00
                                2.000  -     3.000          6   22.00
                                3.000  -     4.000         21   43.00
                                4.000  -     5.000         16   59.00
                                5.000  -     6.000         14   73.00
                                6.000  -     7.000         12   85.00
                                7.000  -     8.000         11   96.00
                                8.000  -     9.000          4  100.00
```

Записуємо дискретний статистичний розподіл частот, який після об'єднання значень з частотами $n_i < 5$ має вигляд:

x_i	$[0,1]$	2	3	4	5	6	7	$[8,\infty)$
n_i	7	9	6	21	16	14	12	15

У випадку гіпотетичного розподілу Пуассона маємо $r = 1$. Кількість інтервалів $m = 8$, кількість ступенів вільності для розподілу χ^2 дорівнює $k = m - r - 1 = 8 - 1 - 1 = 6$, обсяг вибірки $n = 100$, а ймовірності p_i для розподілу Пуассона обчислюємо за формулами

$$p_i = P\{z_{i-1} \le X \le z_i\} = \sum_{k=z_{i-1}}^{z_i} P_n(k), \quad p_i = P\{X = k\} = P_n(k), \; P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \; \lambda = 5;$$

$$p_8 = P\{8 \le X < \infty\} = 1 - \sum_{k=0}^{7} P_n(k).$$

Перевірка гіпотези здійснюється за тим самим алгоритмом, що використовувався у задачах 2.1-2.4.

```
Розподіл Xi квадрат
 f[x_,k_]:=(x^(k/2-1)Exp[-x/2])/((2^(k/2))Gamma[k/2])
Знаходження критичної точки
 Xicr[alpha_,k_]:=FindRoot[1-
Integrate[f[t,k],{t,0,x}]==alpha,{x,20}]
 k=m-r-1=8-1-1=6
Значення критичної точки:
 Xikr[0.05,6]
 {x→12.5916}
Емпіричні частоти:
 n[1]:=7
 n[2]:=9
 n[3]:=6
 n[4]:=21
 n[5]:=16
 n[6]:=14
 n[7]:=12
 n[8]:=15
Перевірка
Sum[n[i],{i,1,8}]
 100
Розподіл Пуассона
 Poisson[k_]:=N[(5^k)Exp[-5]/(k!)]
Теоретичні частоти:
```

```
np[1]:=N[Sum[Poisson[k],{k,0,1}]100]
np[1]
4.04277
np[2]:=N[Poisson[2]100]
np[2]
8.42243
np[3]:=N[Poisson[3]100]
np[3]
14.0374
np[4]:=N[Poisson[4]100]
np[4]
17.5467
np[5]:=N[Poisson[5]100]
np[5]
17.5467
np[6]:=N[Poisson[6]100]
np[6]
14.6223
np[7]:=N[Poisson[7]100]
np[7]
10.4445
np[8]:=N[(1-Sum[Poisson[k],{k,0,7}])100]
np[8]
13.3372
```

Перевірка

```
Sum[np[i],{i,1,8}]
100.
Kemp:=N[Sum[((n[i]-np[i])^2)/np[i],{i,1,8}]]
Kemp
8.08617
```

Kemp<Xicr, гіпотезу приймаємо.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №3
Перевірка параметричних статистичних гіпотез

ЗАДАЧА 3.1
Перевірка статистичної гіпотези про значення математичного сподівання нормального закону розподілу за відомої дисперсії

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Normal(k,100+k,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, записати інтервальний статистичний розподіл частот для отриманої вибірки обсягу n=100 з нормальної генеральної сукупності. Взявши за x_i середини інтервалів, записати відповідний дискретний розподіл частот і за даними вибірки обчислити вибіркове середнє $\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$. Вважаючи, що середнє квадратичне відхилення розподілу відоме σ=4, для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: a = a_0 = 100 + k$ (k – номер варіанту) *про рівність математичного сподівання нормально розподіленої випадкової величини X гіпотетичному значенню* a_0 за наявності альтернативної гіпотези $H_1: a \neq a_0$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай випадкова величина X нормально розподілена з невідомими математичним сподіванням $a = E(X)$, але відомою дисперсією $\sigma^2 = D(X)$. Необхідно на основі вибірки перевірити гіпотезу $H_0: a = a_0$ про рівність математичного сподівання a гіпотетичному числу a_0. Припускаємо, що відомі такі величини: дані вибірки обсягу n, середнє квадратичне відхилення $\sigma = \sqrt{D(X)}$, гіпотетичне значення математичного сподівання a_0, рівень значущості α.

Вибіркове середнє $\overline{x}$ для вибірки з нормального розподілу з параметрами (a, σ^2) має нормальний розподіл з параметрами $(a, \sigma^2 / n)$, тому за умови істинності гіпотези H_0 (коли $E(\overline{x}) = a_0$) випадкова величина

$$U = \frac{(\overline{x} - a_0)\sqrt{n}}{\sigma},$$

яку беремо за критерій перевірки гіпотези H_0, також розподілена нормально з параметрами (0, 1).

Справді,

$$E(U) = \frac{\sqrt{n}}{\sigma} E(\overline{x} - a_0) = \frac{\sqrt{n}}{\sigma}\left(E(\overline{x}) - a_0\right) = 0;$$

$$D(U) = \frac{n}{\sigma^2} D(\overline{x} - a_0) = \frac{n}{\sigma^2} D(\overline{x}) = \frac{n}{\sigma^2} \cdot \frac{\sigma^2}{n} = 1.$$

Отже, щільність розподілу випадкової величини U має вигляд:

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

Тому

$$\mathrm{P}\{U \in (0, z)\} = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx = \Phi(z).$$

Якщо альтернативною є гіпотеза $H_1 : a \neq a_0$, то розглядають двосторонню симетричну критичну область, для якої критичну точку шукають зі співвідношення:

$$\mathrm{P}\{U > u_{cr}\} = \alpha / 2.$$

Оскільки $\mathrm{P}\{0 < U < \infty\} = \Phi(\infty) = 0,5,$ то

$$\mathrm{P}\{0 < U < \infty\} = \mathrm{P}\{0 < U < u_{cr}\} + \mathrm{P}\{U > u_{cr}\} = 0,5,$$

тобто

$$\Phi(u_{cr}) + \frac{\alpha}{2} = \frac{1}{2}$$

або

$$\Phi(u_{cr}) = \frac{1-\alpha}{2}. \qquad (3.1)$$

Отже, у випадку альтернативної гіпотези $H_1 : a \neq a_0$, перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$U_{емр} = \frac{(\overline{x} - a_0)\sqrt{n}}{\sigma}; \qquad (3.2)$$

2) знаходимо критичну точку u_{cr} як розв’язок рівняння (3.1);

3) робимо висновок про висунуту гіпотезу: якщо $|U_{åmp}| < u_{cr}$, то гіпотезу H_0 приймаємо; якщо $|U_{ài\,ï}| > u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Якщо альтернативною є гіпотеза $H_1: a > a_0$, то розглядаємо правосторонню критичну область, для якої критичну точку шукаємо зі співвідношення:

$$\mathrm{P}\{U > u_{cr}\} = \alpha.$$

Тоді

$$\Phi(u_{cr}) + \alpha = \frac{1}{2},$$

тобто

$$\Phi(u_{cr}) = \frac{1-2\alpha}{2}. \tag{3.3}$$

Якщо $U_{åmp} < u_{cr}$, то гіпотезу H_0 приймаємо; якщо $U_{åmp} > u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a > a_0$.

Якщо альтернативною є гіпотеза $H_1: a < a_0$, то розглядаємо лівосторонню критичну область, для якої $P\{U < -u_{cr}\} = \alpha$. Критичну точку u_{cr} шукаємо як розв'язок рівняння (3.3). Якщо $U_{åmp} > -u_{cr}$, то гіпотезу H_0 приймаємо; якщо $U_{åmp} < -u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a < a_0$.

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
Gis TABLE (Normal(50,150,4)),0,2,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

TABLE	MEAN	STD.DEV.	RANGE			RETRY	FREQUENCY	CUM.%
GIS	149.602	3.627				0		
			140.000	-	142.000		1	1.00
			142.000	-	144.000		7	8.00
			144.000	-	146.000		9	17.00
			146.000	-	148.000		15	32.00
			148.000	-	150.000		25	57.00
			150.000	-	152.000		11	68.00
			152.000	-	154.000		21	89.00
			154.000	-	156.000		9	98.00

```
156.000  -     158.000          2   100.00
```

Користуючись гістограмою, записуємо дискретний статистичний розподіл частот, попередньо взявши за x_i середини інтервалів (об'єднання інтервалів не здійснюється):

x_i	141	143	145	147	149	151	153	155	157
n_i	1	7	9	15	25	11	21	9	2

Для k=50 гіпотеза H_0 має вигляд $a = a_0 = 150$. Обсяг вибірки обчислюємо за формулою $n = \sum_{i=1}^{m} n_i$.

Оскільки альтернативною є гіпотеза $H_1 : a \neq a_0$, то розглядаємо двосторонню симетричну критичну область, для якої критичну точку шукаємо з рівняння (3.1). Розв'язування рівняння (3.1) здійснюється за допомогою вбудованої функції **Solve[eq,var].** Обчислюємо емпіричне значення критерію за формулою (3.2). Оскільки $u_{cr} = 1{,}95996$, $U_{åmp} = -1$, то $|U_{åmp}| < u_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=141
 x[2]:=143
 x[3]:=145
 x[4]:=147
 x[5]:=149
 x[6]:=151
 x[7]:=153
 x[8]:=155
 x[9]:=157
Емпіричні частоти:
 n[1]:=1
 n[2]:=7
 n[3]:=9
 n[4]:=15
 n[5]:=25
 n[6]:=11
 n[7]:=21
 n[8]:=9
 n[9]:=2
Вибіркове середнє
 xav:=N[Sum[x[i]n[i],{i,1,9}]/100]
 xav
 149.6
 a0:=150
 n:=100
```

```
 sig:=4
```

Емпіричне значення критерію

```
 Uemp:=(xav-a0)Sqrt[n]/sig
 Uemp
 -1.
```

Розв'язування рівняння для критичної точки

```
 Ucr[beta_]:=Solve[Integrate[Exp[-
(z^2)/2],{z,0,u}]/(Sqrt[2Pi])==beta,u]
 beta:=(1-alpha)/2
 alpha:=0.05
```

Значення критичної точки

```
 Ukr[beta]
{{u→1.95996}}
```

|Uemp|<Ucr, гіпотезу приймаємо.

ЗАДАЧА 3.2
Перевірка статистичної гіпотези про значення математичного сподівання нормального закону розподілу за невідомої дисперсії

Розглядаємо дискретний статистичний розподіл частот для нормально розподіленої випадкової величини, одержаний під час розв'язання задачі 3.1. Вважаємо, що середнє квадратичне відхилення розподілу невідоме. За даними вибірки обчислити підправлене середнє квадратичне відхилення s і для рівня значущості $\alpha = 0,001 \cdot k$ ($k -$номер варіанту) перевірити гіпотезу $H_0: a = a_0 = 100 + k$ ($k -$номер варіанту) *про рівність математичного сподівання нормально розподіленої випадкової величини X гіпотетичному значенню* a_0 у двох випадках: 1) за наявності альтернативної гіпотези $H_1: a \neq a_0$; 2) за наявності альтернативної гіпотези $H_1: a < a_0$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай випадкова величина X нормально розподілена з невідомими математичним сподіванням $a = E(X)$ і невідомою дисперсією $\sigma^2 = D(X)$. Необхідно на основі вибірки перевірити гіпотезу $H_0: a = a_0$ про рівність математичного сподівання a гіпотетичному числу a_0. Припускаємо, що відомі такі величини: дані вибірки обсягу n, гіпотетичне значення математичного сподівання a_0, рівень значущості α.

Для перевірки гіпотези H_0 використовуємо статистику

$$T = \frac{(\overline{x} - a_0)\sqrt{n}}{s}.$$

Тут $\overline{x} -$ вибіркове середнє, а $s -$ підправлене середнє квадратичне відхилення. У випадку істинності гіпотези H_0 випадкова величина T розподілена за законом Ст'юдента з $k = n - 1$ ступенями вільності, який характеризується щільністю розподілу

$$s(x,n) = B_n \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \qquad x \in (-\infty, +\infty),$$

де $B_n = \dfrac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)}$, $\quad \Gamma(x) = \int\limits_0^{\infty} t^{x-1} e^{-t} dt.$

У випадку альтернативної гіпотези $H_1: a \neq a_0$, перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$T_{\text{åmp}} = \frac{(\overline{x} - a_0)\sqrt{n}}{s}; \tag{3.4}$$

2) знаходимо критичну точку t_{cr} як розв'язок рівняння

$$\mathrm{P}\{|T| > t_{cr}\} = \alpha; \tag{3.5}$$

3) робимо висновок про висунуту гіпотезу: якщо $|T_{\text{åmp}}| < t_{cr}$, то гіпотезу H_0 приймаємо; якщо $|T_{\text{åì ï}}| > t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Рівняння (3.5) можна записати у вигляді, зручному для розв'язування в програмному середовищі *Mathematica.*

$$1 - 2\int_0^{t_{cr}} s(x,n)dx = \alpha. \tag{3.6}$$

Якщо альтернативною є гіпотеза $H_1: a > a_0$, то розглядаємо правосторонню критичну область, для якої критичну точку шукаємо з рівняння:

$$\mathrm{P}\{T > t_{cr}\} = \alpha,$$

яке еквівалентне такому

$$1 - 2\int_0^{t_{cr}} s(x,n)dx = 2\alpha. \tag{3.7}$$

Якщо $T_{\text{åmp}} < t_{cr}$, то гіпотезу H_0 приймаємо; якщо $T_{\text{åmp}} > t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a > a_0$.

Якщо альтернативною є гіпотеза $H_1: a < a_0$, то розглядаємо лівосторонню критичну область, для якої $\mathrm{P}\{T < -t_{cr}\} = \alpha$. Критичну точку t_{cr} шукаємо як розв'язок рівняння (3.7). Якщо $T_{\text{åmp}} > -t_{cr}$, то гіпотезу H_0 приймаємо; якщо $T_{\text{åmp}} < -t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a < a_0$.

ПОЯСНЕННЯ для k=50

Користуємось дискретним статистичним розподілом частот, одержаним під час розв'язання задачі 3.1:

x_i	141	143	145	147	149	151	153	155	157
n_i	1	7	9	15	25	11	21	9	2

Враховуючи, що $\overline{x} = 149{,}6$, обчислюємо підправлене середнє квадратичне відхилення за формулою

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{m} n_i (x_i - \overline{x})^2}, \tag{3.8}$$

$s = 3{,}64595,$ а потім емпіричне значення критерію згідно з (3.4): $T_{åmp} = -1{,}09711.$

1) Оскільки альтернативною є гіпотеза $H_1: a \neq a_0$, то розглядаємо двосторонню симетричну критичну область, для якої критичну точку шукаємо з рівняння (3.6). Розв'язування рівняння (3.6) здійснюється за допомогою вбудованої функції

FindRoot[lp==rp,{x,x0}],

Оскільки $t_{cr} = 1{,}98422$, то $|T_{åmp}| < t_{cr}$, і гіпотезу H_0 приймаємо.

2) У випадку альтернативної гіпотези $H_1: a < a_0$ розглядаємо лівосторонню критичну область, а критичну точку знаходимо, розв'язавши рівняння (3.7). Оскільки $t_{cr} = 1{,}66039$, то $T_{åmp} > -t_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=141
 x[2]:=143
 x[3]:=145
 x[4]:=147
 x[5]:=149
 x[6]:=151
 x[7]:=153
 x[8]:=155
 x[9]:=157
Емпіричні частоти:
 n[1]:=1
 n[2]:=7
 n[3]:=9
 n[4]:=15
 n[5]:=25
 n[6]:=11
 n[7]:=21
 n[8]:=9
 n[9]:=2
Вибіркове середнє
 xav:=N[Sum[x[i]n[i],{i,1,9}]/100]
 xav
 149.6
Обчислення s:
 skv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,9}]/99]
 skv
 13.2929
 s:=Sqrt[skv]
```

```
 s
 3.64595
 a0:=150
 n:=100
```
Емпіричне значення критерію
```
 Temp:=(xav-a0)Sqrt[n]/s
 Temp
 -1.09711
```
Розподіл Ст’юдента
```
 B[n_]:=Gamma[n/2]/(Sqrt[Pi(n-1)]Gamma[(n-1)/2])
 S[x_,n_]:=B[n](1+(x^2)/(n-1))^(-n/2)
```
Випадок 1: Розв’язування рівняння (3.6)
```
 Tcr2[alpha_,k_]:=FindRoot[1-
2Integrate[S[t,k+1],{t,0,x}]==alpha,{x,1}]
 k=n-1=99
 Tcr2[0.05,99]
 {x→1.98422}
```
|Temp|<Tcr, гіпотезу приймаємо.

Випадок 2: Розв’язування рівняння (3.7)
```
 Tcr1[alpha_,k_]:=FindRoot[1-
2Integrate[S[t,k+1],{t,0,x}]==2alpha,{x,1}]
 k=n-1=99
 Tcr1[0.05,99]
 {x→1.66039}
```
Temp>-Tcr, гіпотезу приймаємо.

ЗАДАЧА 3.3

Перевірка статистичної гіпотези про значення математичного сподівання довільного закону розподілу у випадку великого обсягу вибірки та невідомої дисперсії

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 1001  ; час моделювання
Rozp FUNCTION RNk,D10
.1,k+1/.2,k+2/.3,k+3/.4,k+4/.5,k+5/.6,k+6/.7,k+7/.8,k+8/.
9,k+9/1,k+10  ;випадкова величина
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, взяти за значення x_i **праві кінці** інтервалів $(z_{i-1}, z_i]$ і записати дискретний статистичний розподіл частот для отриманої вибірки. За даними вибірки обчислити вибіркове середнє $\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$ та підправлене середнє квадратичне відхилення s за формулою (3.8). Вважаючи, що середнє квадратичне відхилення розподілу невідоме, для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: a = a_0 = k + 5{,}5$ (k – номер варіанту) *про рівність математичного сподівання випадкової величини X гіпотетичному значенню* a_0 у двох випадках: 1) за наявності альтернативної гіпотези $H_1: a \neq a_0$; 2) за наявності альтернативної гіпотези $H_1: a > a_0$.. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай випадкова величина X довільно розподілена з невідомими математичним сподіванням $a = E(X)$ і невідомою дисперсією $\sigma^2 = D(X)$. Необхідно на основі вибірки перевірити гіпотезу $H_0: a = a_0$ про рівність математичного сподівання a гіпотетичному числу a_0. Припускаємо, що відомі такі величини: дані вибірки обсягу $n > 30$, гіпотетичне значення математичного сподівання a_0, рівень значущості α.

Для перевірки гіпотези H_0 використовуємо статистику

$$U = \frac{(\overline{x} - a_0)\sqrt{n}}{s},$$

де $\overline{x}$ – вибіркове середнє, а s – підправлене середнє квадратичне відхилення. Правило перевірки гіпотези і рівняння для відшукання критичних точок залишаються такими, що й в задачі 3.1 (см. п. 3.1.1), але емпіричне значення критерію шукаємо за зміненою формулою

$$U_{emp} = \frac{(\overline{x} - a_0)\sqrt{n}}{s}, \tag{3.9}$$

спочатку обчисливши s згідно з (3.8).

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 1001  ; час моделювання
Rozp FUNCTION RN50,D10
.1,51/.2,52/.3,53/.4,54/.5,55/.6,56/.7,57/.8,58/.9,59/1,60
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмою:

```
TABLE                 MEAN    STD.DEV.       RANGE              RETRY  FREQUENCY  CUM.%
 GIS                 55.463    2.872                               0
                                     50.000  -       51.000             107    10.70
                                     51.000  -       52.000             100    20.70
                                     52.000  -       53.000              88    29.50
                                     53.000  -       54.000             111    40.60
                                     54.000  -       55.000              88    49.40
                                     55.000  -       56.000             117    61.10
                                     56.000  -       57.000              99    71.00
                                     57.000  -       58.000              99    80.90
                                     58.000  -       59.000              89    89.80
                                     59.000  -       60.000             102   100.00
```

Користуючись гістограмою, записуємо дискретний статистичний розподіл частот, попередньо взявши за x_i праві кінці інтервалів:

x_i	51	52	53	54	55	56	57	58	59	60
n_i	107	100	88	111	88	117	99	99	89	102

Для k=50 гіпотеза H_0 має вигляд $a = a_0 = 55{,}5$. Враховуючи, що $\overline{x} = 55{,}463$, обчислюємо підправлене середнє квадратичне відхилення за формулою (3.8): $s = 3{,}64595$, а потім емпіричне значення критерію згідно з (3.9): $U_{åmp} = -0{,}407434$.

1) Оскільки альтернативною є гіпотеза $H_1: a \neq a_0$, то розглядаємо двосторонню симетричну критичну область, для якої критичну точку шукаємо з рівняння (3.1). Розв’язування рівняння (3.1) здійснюється за допомогою вбудованої функції **`Solve[eq,var].`** Оскільки $u_{cr} = 1{,}95996$, то $|U_{åmp}| < u_{cr}$, і гіпотезу H_0 приймаємо.

2) У випадку альтернативної гіпотези $H_1: a > a_0$ розглядаємо правостороню критичну область, а критичну точку знаходимо, розв’язавши рівняння (3.3). Оскільки $u_{cr} = 1{,}64485$, то $U_{åmp} < u_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=51
 x[2]:=52
 x[3]:=53
 x[4]:=54
 x[5]:=55
 x[6]:=56
 x[7]:=57
 x[8]:=58
 x[9]:=59
 x[10]:=60
Емпіричні частоти:
 n[1]:=107
 n[2]:=100
 n[3]:=88
 n[4]:=111
 n[5]:=88
 n[6]:=117
 n[7]:=99
 n[8]:=99
 n[9]:=89
 n[10]:=102
Перевірка
 Sum[n[i],{i,1,10}]
 1000
Вибіркове середнє
 xav:=N[Sum[x[i]n[i],{i,1,10}]/1000]
 xav
 55.463
 xav:=55.463`
Обчислення s:
 skv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,10}]/999]
 s:=Sqrt[skv]
 s
 2.87174
```

```
a0:=55.5
```

Емпіричне значення критерію

```
Uemp:=(xav-a0)Sqrt[1000]/s
Uemp
-0.407434
```

Розв'язування рівняння для критичної точки

```
Ucr[beta_]:=Solve[Integrate[Exp[-
(z^2)/2],{z,0,u}]/(Sqrt[2Pi])==beta,u]
```

Випадок 1

```
beta:=(1-alpha)/2
alpha:=0.05
```

Значення критичної точки

```
Ucr[beta]
{{u→1.95996}}
```

|Uemp|<Ucr, гіпотезу приймаємо.

Випадок 2

```
beta:=(1-2(alpha))/2
```

Значення критичної точки

```
Ukr[beta]
{{u→1.64485}}
```

Uemp<Ucr, гіпотезу приймаємо.

ЗАДАЧА 3.4
Перевірка статистичної гіпотези про рівність математичних сподівань двох нормальних генеральних сукупностей, дисперсії яких відомі

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 101  ; час моделювання
GisX TABLE (Normal(k,100+k,3)),0,2,500  ; гістограма X
GisY TABLE (Normal(k,100+k,4)),0,2,500  ; гістограма Y
GENERATE 1
TABULATE GisX
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограми з одержаного стандартного звіту, записати дискретні статистичні розподіли частот для випадкових величин X і Y, взявши за значення x_i, y_i середини інтервалів $(z_{i-1}, z_i]$. Обчислити вибіркові середні $\overline{x} = \frac{1}{n_x}\sum_{i=1}^{m_x} n_{xi}x_i$, $\overline{y} = \frac{1}{n_y}\sum_{i=1}^{m_y} n_{yi}y_i$. Вважаючи, що середні квадратичні відхилення розподілів X і Y відомі ($\sigma_x = 3,\ \sigma_y = 4$), а математичні сподівання a_x, a_y невідомі, для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: a_x = a_y$ *про рівність математичних сподівань двох нормальних генеральних сукупностей з відомими дисперсіями* за наявності альтернативної гіпотези $H_1: a_x \neq a_y$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай генеральні сукупності X і Y розподілені нормально, причому їхні дисперсії σ_x^2, σ_y^2 відомі. За незалежними вибірками, обсяги яких відповідно рівні n_x і n_y, отриманими з цих генеральних сукупностей, знайдено вибіркові середні

$$\overline{x} = \frac{1}{n_x}\sum_{i=1}^{m_x} n_{xi}x_i, \qquad \overline{y} = \frac{1}{n_y}\sum_{i=1}^{m_y} n_{yi}y_i.$$

Тут n_{xi}, n_{yi} – емпіричні частоти варіант x_i, y_i відповідно, m_x, m_y – кількість інтервалів для вибірок з генеральних сукупностей X і Y.

Необхідно за заданим рівнем значущості α перевірити гіпотезу $H_0 : a_x = a_y$, яка полягає в тому, що генеральні середні $E(X) = a_x$ і $E(Y) = a_y$ рівні між собою.

За критерій перевірки гіпотези H_0 приймемо випадкову величину

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}},$$

яка за умови істинності гіпотези H_0 розподілена нормально з параметрами $E(Z) = 0,\ \sqrt{D(Z)} = 1.$

У випадку альтернативної гіпотези $H_1 : a_x \neq a_y$ перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$Z_{emp} = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}. \tag{3.10}$$

2) знаходимо критичну точку z_{cr} як розв'язок рівняння

$$\Phi(z_{cr}) = \frac{1-\alpha}{2}; \tag{3.11}$$

3) робимо висновок про висунуту гіпотезу: якщо $|Z_{\text{åmp}}| < z_{cr}$, то гіпотезу H_0 приймаємо; якщо $|Z_{\text{åì ï}}| > z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Якщо альтернативною є гіпотеза $H_1 : a_x > a_y$, то розглядаємо правосторонню критичну область, для якої критичну точку шукаємо з рівняння:

$$\Phi(z_{cr}) = \frac{1-2\alpha}{2}. \tag{3.12}$$

Якщо $Z_{\text{åmp}} < z_{cr}$, то гіпотезу H_0 приймаємо; якщо $Z_{\text{åmp}} > z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1 : a_x > a_y$.

Якщо альтернативною є гіпотеза $H_1 : a_x < a_y$, то розглядаємо лівосторонню критичну область, для якої $\mathrm{P}\{Z < -z_{cr}\} = \alpha$. Критичну точку z_{cr} шукаємо як розв'язок рівняння (3.12). Якщо $Z_{\text{åmp}} > -z_{cr}$, то гіпотезу H_0 приймаємо; якщо $Z_{\text{åmp}} < -z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1 : a_x < a_y$.

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
GisX TABLE (Normal(50,150,3)),0,2,500  ; гістограма X
GisY TABLE (Normal(50,150,4)),0,2,500  ; гістограма Y
GENERATE 1
TABULATE GisX
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагмент з гістограмами:

```
TABLE              MEAN    STD.DEV.       RANGE            RETRY FREQUENCY CUM.%
 GISX            149.601    2.735                            0
                                     140.000  -      142.000            1     1.00
                                     142.000  -      144.000            1     2.00
                                     144.000  -      146.000           10    12.00
                                     146.000  -      148.000           15    27.00
                                     148.000  -      150.000           31    58.00
                                     150.000  -      152.000           24    82.00
                                     152.000  -      154.000           13    95.00
                                     154.000  -      156.000            4    99.00
                                     156.000  -      158.000            1   100.00
 GISY            149.768    3.890                            0
                                     140.000  -      142.000            1     1.00
                                     142.000  -      144.000            6     7.00
                                     144.000  -      146.000           14    21.00
                                     146.000  -      148.000           12    33.00
                                     148.000  -      150.000           19    52.00
                                     150.000  -      152.000           15    67.00
                                     152.000  -      154.000           20    87.00
                                     154.000  -      156.000            9    96.00
                                     156.000  -      158.000            3    99.00
                                     158.000  -      160.000            1   100.00
```

Користуючись гістограмами, записуємо дискретні статистичні розподіли частот, попередньо взявши за x_i, y_i середини інтервалів:

x_i	141	143	145	147	149	151	153	155	157
n_{xi}	1	1	10	15	31	24	13	4	1

y_i	141	143	145	147	149	151	153	155	157	159
n_{yi}	1	6	14	12	19	15	20	9	3	1

Оскільки альтернативною є гіпотеза $H_1 : a_x \neq a_y$, то розглядаємо двосторонню симетричну критичну область, для якої критичну точку шукаємо з рівняння (3.11). Розв'язування рівняння (3.11) здійснюється за допомогою вбудованої функції **`Solve[eq,var]`.** Обчислюємо вибіркові середні

($\overline{x} = 149{,}48,\ \overline{y} = 149{,}74$) та емпіричне значення критерію за формулою (3.10). Оскільки $z_{cr} = 1{,}95996,\ Z_{\mathring{a}mp} = -0{,}52$, то $|Z_{\mathring{a}mp}| < z_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=141
 x[2]:=143
 x[3]:=145
 x[4]:=147
 x[5]:=149
 x[6]:=151
 x[7]:=153
 x[8]:=155
 x[9]:=157
Емпіричні частоти:
 n[1]:=1
 n[2]:=1
 n[3]:=10
 n[4]:=15
 n[5]:=31
 n[6]:=24
 n[7]:=13
 n[8]:=4
 n[9]:=1
Перевірка
 Sum[n[i],{i,1,9}]
 100
Вибіркове середнє X
 xav:=N[Sum[x[i]n[i],{i,1,9}]/100]
 xav
 149.48
Варіанти (значення y[i]):
 y[1]:=141
 y[2]:=143
 y[3]:=145
 y[4]:=147
 y[5]:=149
 y[6]:=151
 y[7]:=153
 y[8]:=155
 y[9]:=157
 y[10]:=159
Емпіричні частоти:
 m[1]:=1
 m[2]:=6
 m[3]:=14
```

```
 m[4]:=12
 m[5]:=19
 m[6]:=15
 m[7]:=20
 m[8]:=9
 m[9]:=3
 m[10]:=1
```

Перевірка

```
 Sum[m[i],{i,1,10}]
 100
```

Вибіркове середнє Y

```
 yav:=N[Sum[y[i]m[i],{i,1,10}]/100]
 yav
 149.74
 sigxkv:=9
 sigykv:=16
```

Емпіричне значення критерію

```
 Zemp:=(xav-yav)/Sqrt[(sigxkv+sigykv)/100]
 Zemp
 -0.52
```

Розв'язування рівняння для критичної точки

```
 Zcr[beta_]:=Solve[Integrate[Exp[-
(x^2)/2],{x,0,z}]/(Sqrt[2Pi])==beta,z]
 beta:=(1-alpha)/2
 alpha:=0.05
 Zcr[beta]
 {{z→1.95996}}
```

|Zemp|<Zcr, гіпотезу приймаємо.

ЗАДАЧА 3.5
Перевірка статистичної гіпотези про рівність математичних сподівань двох нормальних генеральних сукупностей, дисперсії яких невідомі і однакові (випадок малих незалежних вибірок)

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 21  ; час моделювання
Z Variable (Exponential(k,0,1))
Iks Variable
(84+k+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z
+V$Z+V$Z+V$Z)
GisX TABLE (V$Iks),0,4,500  ; гістограма X
GisY TABLE (Normal(k,100+k,4)),0,4,500  ; гістограма Y
GENERATE 1
TABULATE GisX
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограми з одержаного стандартного звіту, записати дискретні статистичні розподіли частот для випадкових величин X і Y, взявши за значення x_i, y_i середини інтервалів $(z_{i-1}, z_i]$. Обчислити вибіркові середні і підправлені вибіркові дисперсії:

$$\overline{x} = \frac{1}{n_x}\sum_{i=1}^{m_x} n_{xi}x_i, \qquad \overline{y} = \frac{1}{n_y}\sum_{i=1}^{m_y} n_{yi}y_i,$$

$$s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{m_x} n_{xi}(x_i - \overline{x})^2, \quad s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{m_y} n_{yi}(y_i - \overline{y})^2.$$

Вважаючи, що середні квадратичні відхилення розподілів X і Y невідомі і однакові, для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0 : a_x = a_y$ *про рівність математичних сподівань двох нормальних генеральних сукупностей з невідомими і однаковими дисперсіями* за наявності альтернативної гіпотези $H_1 : a_x \neq a_y$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай генеральні сукупності X і Y розподілені нормально, причому їхні дисперсії σ_x^2, σ_y^2 невідомі, але однакові. Необхідно за заданим рівнем значущості α перевірити гіпотезу $H_0 : a_x = a_y$, яка полягає в тому, що генеральні середні $E(X) = a_x$ та $E(Y) = a_y$ рівні між собою.

За критерій перевірки гіпотези H_0 приймемо випадкову величину

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}} \cdot \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}},$$

яка за умови істинності гіпотези H_0 має розподіл Ст'юдента з $k = n_x + n_y - 2$ ступенями вільності і щільністю розподілу

$$s(x,k) = B_k \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \qquad x \in (-\infty, +\infty),$$

де $B_k = \dfrac{\Gamma((k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)}, \quad \Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt.$

У випадку альтернативної гіпотези $H_1 : a_x \neq a_y$ перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$T_{emp} = \frac{\overline{x} - \overline{y}}{\sqrt{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}} \cdot \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}, \qquad (3.13)$$

де

$$\overline{x} = \frac{1}{n_x} \sum_{i=1}^{m_x} n_{xi} x_i, \qquad \overline{y} = \frac{1}{n_y} \sum_{i=1}^{m_y} n_{yi} y_i,$$

$$s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{m_x} n_{xi} (x_i - \overline{x})^2, \quad s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{m_y} n_{yi} (y_i - \overline{y})^2;$$

2) знаходимо критичну точку z_{cr} як розв'язок рівняння

$$\mathrm{P}\{|T| > t_{cr}\} = \alpha; \qquad (3.14)$$

3) робимо висновок про висунуту гіпотезу: якщо $|T_{åmp}| < t_{cr}$, то гіпотезу H_0 приймаємо; якщо $|T_{åmp}| > t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Рівняння (3.14) запишемо у вигляді, зручному для розв'язування в програмному середовищі *Mathematica.*

$$1 - 2\int\limits_0^{t_{cr}} s(x,k)dx = \alpha. \qquad (3.15)$$

Якщо альтернативною є гіпотеза $H_1 : a_x > a_y$, то розглядаємо правосторонню критичну область, для якої критичну точку шукаємо з рівняння

$$P\{T > t_{cr}\} = \alpha,$$

еквівалентного рівності

$$1 - 2\int_0^{t_{cr}} s(x,k)dx = 2\alpha. \qquad (3.16)$$

Якщо $T_{\mathring{a}mp} < t_{cr}$, то гіпотезу H_0 приймаємо; якщо $T_{\mathring{a}mp} > t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1 : a_x > a_y$.

Якщо альтернативною є гіпотеза $H_1 : a_x < a_y$, то розглядаємо лівосторонню критичну область, для якої $\mathrm{P}\{T < -t_{cr}\} = \alpha$. Критичну точку t_{cr} шукаємо як розв'язок рівняння (3.16). Якщо $T_{\mathring{a}mp} > -t_{cr}$, то гіпотезу H_0 приймаємо; якщо $T_{\mathring{a}mp} < -t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1 : a_x < a_y$.

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 21  ; час моделювання
Z Variable (Exponential(50,0,1))
Iks Variable
(84+50+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$
Z+V$Z+V$Z+V$Z)
GisX TABLE (V$Iks),0,4,500  ; гістограма X
GisY TABLE (Normal(50,150,4)),0,4,500  ; гістограма Y
GENERATE 1
TABULATE GisX
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту вибираємо фрагменти з гістограмами:

TABLE	MEAN	STD.DEV.	RANGE			RETRY	FREQUENCY	CUM.%
GISX	150.431	5.043				0		
			140.000	-	144.000		3	15.00
			144.000	-	148.000		5	40.00
			148.000	-	152.000		3	55.00
			152.000	-	156.000		7	90.00
			156.000	-	160.000		1	95.00
			160.000	-	164.000		1	100.00
GISY	149.494	4.976				0		
			136.000	-	140.000		1	5.00
			140.000	-	144.000		1	10.00
			144.000	-	148.000		6	40.00
			148.000	-	152.000		4	60.00

```
            152.000  -      156.000              7    95.00
            156.000  -      160.000              1   100.00
```

Користуючись гістограмами, записуємо дискретні статистичні розподіли частот, попередньо взявши за x_i, y_i середини інтервалів:

x_i	142	146	150	154	158	162
n_{xi}	3	5	3	7	1	1

y_i	138	142	146	150	154	158
n_{yi}	1	1	6	4	7	1

Емпіричне значення критерію шукаємо за формулою

$$T_{\text{åì ï}} = \frac{\overline{x} - \overline{y}}{\sqrt{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}} \cdot \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}.$$

Оскільки альтернативною є гіпотеза $H_1 : a_x \neq a_y$, то розглядаємо двосторонню симетричну критичну область, для якої критичну точку шукаємо з рівняння (3.15). Розв’язування рівняння (3.15) здійснюється за допомогою вбудованої функції **FindRoot[lp==rp,{x,x0}]**.

Обчислюємо вибіркові середні ($\overline{x} = 150{,}2$, $\overline{y} = 149{,}6$), підправлені вибіркові дисперсії ($s_x^2 = 31{,}1158$, $s_y^2 = 25{,}0947$) і емпіричне значення критерію за формулою (3.13). Оскільки $t_{cr} = 2{,}02439$, $T_{\text{åmp}} = 0{,}357896$, то $|T_{\text{åmp}}| < t_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=142
 x[2]:=146
 x[3]:=150
 x[4]:=154
 x[5]:=158
 x[6]:=162
Емпіричні частоти:
 n[1]:=3
 n[2]:=5
 n[3]:=3
 n[4]:=7
 n[5]:=1
 n[6]:=1
Вибіркове середнє X
 xav:=N[Sum[x[i]n[i],{i,1,6}]/20]
 xav
 150.2
Варіанти (значення y[i]):
 y[1]:=138
```

```
 y[2]:=142
 y[3]:=146
 y[4]:=150
 y[5]:=154
 y[6]:=158
Емпіричні частоти:
 m[1]:=1
 m[2]:=1
 m[3]:=6
 m[4]:=4
 m[5]:=7
 m[6]:=1
Вибіркове середнє Y
 yav:=N[Sum[y[i]m[i],{i,1,6}]/20]
 yav
 149.6
Підправлена вибіркова дисперсія X
 sxkv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,6}]/19]
 sxkv
 31.1158
Підправлена вибіркова дисперсія Y
 sykv:=N[Sum[(m[i])((y[i]-yav)^2),{i,1,6}]/19]
 sykv
 25.0947
Емпіричне значення критерію
 Temp:=(xav-yav)Sqrt[Times[400,38]]/Sqrt[19(sxkv+sykv)40]
 Temp
 0.357896
Розподіл Ст'юдента з n ступенями вільності
 B[n_]:=Gamma[(n+1)/2]/(Sqrt[Pi(n)]Gamma[n/2])
 S[x_,n_]:=B[n](1+(x^2)/n)^(-(n+1)/2)
Розв'язування рівняння для критичної точки
 Tcr2[alpha_,k_]:=FindRoot[1-
2Integrate[S[t,k],{t,0,x}]==alpha,{x,1}]
 k=20+20-2=38
 Tcr2[0.05,38]
 {x→2.02439}
|Temp|<Tcr, гіпотезу приймаємо.
```

ЗАДАЧА 3.6

Перевірка статистичної гіпотези про рівність математичних сподівань двох нормальних генеральних сукупностей, дисперсії яких невідомі (випадок великих вибірок)

Реалізувати дві програми GPSS World (кожну в окремому файлі), спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 201  ; час моделювання
Z Variable (Exponential(k,0,1))
Iks Variable
(84+k+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z
+V$Z+V$Z+V$Z)
GisX TABLE (V$Iks),0,4,500  ; гістограма X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

```
Tmod EQU 301  ; час моделювання
Z Variable (Exponential(k,0,5/4))
Y Variable
(80+k+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z
+V$Z+V$Z+V$Z)
GisY TABLE (V$Y),0,4,500  ; гістограма Y
GENERATE 1
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограми з одержаних стандартних звітів, записати дискретні статистичні розподіли частот для випадкових величин X і Y, взявши за значення x_i, y_i середини інтервалів $(z_{i-1}, z_i]$. Обчислити вибіркові середні і підправлені вибіркові дисперсії. Вважаючи, що середні квадратичні відхилення розподілів X і Y невідомі, для рівня значущості $\alpha = 0,001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0 : a_x = a_y$ *про рівність математичних сподівань двох нормальних генеральних сукупностей з невідомими дисперсіями* за наявності альтернативної гіпотези $H_1 : a_x \neq a_y$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай генеральні сукупності X і Y розподілені нормально, причому їхні математичні сподівання a_x, a_y і дисперсії σ_x^2, σ_y^2 невідомі. За незалежними вибірками з цих генеральних сукупностей, обсяги яких відповідно рівні $n_x \geq 100$ і $n_y \geq 100$, знайдено вибіркові середні $\overline{x}$ і $\overline{y}$. Необхідно за заданим рівнем значущості α перевірити гіпотезу $H_0: a_x = a_y$, яка полягає в тому, що генеральні середні $E(X) = a_x$ і $E(Y) = a_y$ рівні між собою.

Оскільки обсяги вибірок великі, то за критерій перевірки гіпотези H_0 приймемо випадкову величину

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}},$$

де s_x^2, s_y^2 – підправлені вибіркові дисперсії, обчислені за вибірками з генеральних сукупностей X і Y відповідно.

У випадку альтернативної гіпотези $H_1: a_x \neq a_y$ перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$Z_{emp} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}}. \qquad (3.17)$$

2) знаходимо критичну точку z_{cr} як розв'язок рівняння (3.11);

3) робимо висновок про висунуту гіпотезу: якщо $|Z_{\text{åmp}}| < z_{cr}$, то гіпотезу H_0 приймаємо; якщо $|Z_{\text{åì ï}}| > z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Якщо альтернативною є гіпотеза $H_1: a_x > a_y$, то розглядаємо правосторонню критичну область, для якої критичну точку шукаємо з рівняння (3.12). Якщо $Z_{\text{åmp}} < z_{cr}$, то гіпотезу H_0 приймаємо; якщо $Z_{\text{åmp}} > z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a_x > a_y$.

Якщо альтернативною є гіпотеза $H_1: a_x < a_y$, то розглядаємо лівосторонню критичну область, для якої $\mathrm{P}\{Z < -z_{cr}\} = \alpha$. Критичну точку z_{cr} шукаємо як розв’язок рівняння (3.12). Якщо $Z_{\text{åmp}} > -z_{cr}$, то гіпотезу H_0 приймаємо; якщо $Z_{\text{åmp}} < -z_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: a_x < a_y$.

ПОЯСНЕННЯ для k=50

Програми записуємо у вигляді:

```
Tmod EQU 201  ; час моделювання
Z Variable (Exponential(50,0,1))
Iks Variable
(84+50+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$
Z+V$Z+V$Z+V$Z)
GisX TABLE (V$Iks),0,4,500  ; гістограма X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

```
Tmod EQU 301  ; час моделювання
Z Variable (Exponential(50,0,5/4))
Y Variable
(80+50+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$
Z+V$Z+V$Z+V$Z)
GisY TABLE (V$Y),0,4,500  ; гістограма Y
GENERATE 1
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартних звітів вибираємо фрагменти з гістограмами:

TABLE	MEAN	STD.DEV.	RANGE			RETRY	FREQUENCY	CUM.%
GISX	150.385	4.322				0		
			140.000	-	144.000		8	4.00
			144.000	-	148.000		56	32.00
			148.000	-	152.000		71	67.50
			152.000	-	156.000		46	90.50
			156.000	-	160.000		11	96.00
			160.000	-	164.000		8	100.00

TABLE	MEAN	STD.DEV.	RANGE			RETRY	FREQUENCY	CUM.%
GISY	150.124	5.203				0		
			136.000	-	140.000		1	0.33
			140.000	-	144.000		29	10.00
			144.000	-	148.000		77	35.67
			148.000	-	152.000		99	68.67
			152.000	-	156.000		56	87.33
			156.000	-	160.000		24	95.33
			160.000	-	164.000		10	98.67
			164.000	-	168.000		4	100.00

Користуючись гістограмами, записуємо дискретні статистичні розподіли частот, попередньо взявши за x_i, y_i середини інтервалів:

x_i	142	146	150	154	158	162
n_{xi}	8	56	71	46	11	8

y_i	138	142	146	150	154	158	162	166
n_{yi}	1	29	77	99	56	24	10	4

Для даного варіанту

$$n_x = \sum_{i=1}^{m_x} n_{xi} = 200, \quad n_y = \sum_{i=1}^{m_y} n_{yi} = 300, \quad m_x = 6, \quad m_y = 8.$$

Емпіричні частоти для X позначимо через n[i], для Y - через m[i]. Обчислимо вибіркові середні і підправлені дисперсії.

```
x[1]:=142
x[2]:=146
x[3]:=150
x[4]:=154
x[5]:=158
x[6]:=162
n[1]:=8
n[2]:=56
n[3]:=71
n[4]:=46
n[5]:=11
n[6]:=8
xav:=N[Sum[x[i]n[i],{i,1,6}]/200]
xav
150.4
y[1]:=138
y[2]:=142
y[3]:=146
y[4]:=150
y[5]:=154
y[6]:=158
y[7]:=162
y[8]:=166
m[1]:=1
m[2]:=29
m[3]:=77
m[4]:=99
m[5]:=56
m[6]:=24
m[7]:=10
m[8]:=4
yav:=N[Sum[y[i]m[i],{i,1,8}]/300]
yav
150.16
sxkv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,6}]/199]
sxkv
```

```
 19.9397
 sykv:=N[Sum[(m[i])((y[i]-yav)^2),{i,1,8}]/299]
 sykv
 27.1583
Емпіричне значення критерію
 Zemp:=(xav-yav)/Sqrt[(sxkv/200)+(sykv/300)]
 Zemp
 0.550271
Розв'язування рівняння для критичної точки
 Zcr[beta_]:=Solve[Integrate[Exp[-
(x^2)/2],{x,0,z}]/(Sqrt[2Pi])==beta,z]
 beta:=(1-alpha)/2
 alpha:=0.05
 Zcr[beta]
{{z→1.95996}}
 Zkr=1.9599639845400543`
 1.95996
 Оскільки  |Zemp|=0.5502705508536286`<Zkr=1.9599639845400543`, то
гіпотезу H_0 приймаємо.
```

ЗАДАЧА 3.7

Перевірка статистичної гіпотези про рівність дисперсій двох нормальних генеральних сукупностей

Реалізувати дві програми GPSS World (кожну в окремому файлі), спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 31  ; час моделювання
GisX TABLE (Normal(k,200+k,4)),0,2,500  ; гістограма X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

```
Tmod EQU 21  ; час моделювання
GisY TABLE (Normal(k,100+k,4)),0,2,500  ; гістограма Y
GENERATE 1
TABULATE GisY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограми з одержаних стандартних звітів, записати дискретні статистичні розподіли частот для випадкових величин X і Y, взявши за значення x_i, y_i середини інтервалів $(z_{i-1}, z_i]$. Обчислити вибіркові середні і підправлені вибіркові дисперсії за формулами (3.2). Вважаючи, що дисперсії розподілів X і Y невідомі, для рівня значущості $\alpha = 0,001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: \sigma_x^2 = \sigma_y^2$ *про рівність дисперсій двох нормальних генеральних сукупностей* за наявності альтернативної гіпотези $H_1:\ \sigma_x^2 \neq \sigma_y^2$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай генеральні сукупності X і Y розподілені нормально. За незалежними вибірками з обсягами із цих сукупностей, відповідно рівними n_x і n_y, знайдено підправлені вибіркові дисперсії s_x^2 і s_y^2. Необхідно за підправленими дисперсіями для заданого рівня значущості α перевірити статистичну гіпотезу, яка полягає в тому, що генеральні дисперсії сукупностей X і Y рівні між собою: $H_0: \sigma_x^2 = \sigma_y^2$.

За критерій перевірки висунутої гіпотези приймемо відношення більшої підправленої дисперсії до меншої, тобто випадкову величину

$$F = \frac{s_1^2}{s_2^2},$$

якщо $s_1^2 \geq s_2^2$. Величина F за умови істинності гіпотези H_0 має розподіл Фішера-Снедекора зі ступенями вільності $k_1 = n_1 - 1$ і $k_2 = n_2 - 1$. Тут $n_1 -$ обсяг вибірки, за якою обчислено більшу підправлену дисперсію, $n_2 -$ обсяг вибірки, за якою знайдено меншу підправлену дисперсію. Щільність цього розподілу

$$f(x, k_1, k_2) = \begin{cases} 0, & x \leq 0, \\ C_0 \dfrac{x^{(k_1-2)/2}}{(k_2 + k_1 x)^{(k_1+k_2)/2}}, & x > 0, \end{cases}$$

де

$$C_0 = \frac{\Gamma\left(\dfrac{k_1 + k_2}{2}\right) k_1^{k_1/2} k_2^{k_2/2}}{\Gamma(k_1/2)\Gamma(k_2/2)}.$$

У випадку альтернативної гіпотези $H_1: \sigma_x^2 \neq \sigma_y^2$ перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$F_{emp} = \frac{s_1^2}{s_2^2}, \tag{3.18}$$

де $s_1^2 = s_x^2$, $s_2^2 = s_y^2$, якщо $s_x^2 \geq s_y^2$, або $s_1^2 = s_y^2$, $s_2^2 = s_x^2$, якщо $s_y^2 \geq s_x^2$,

$$s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{m_x} n_{xi}(x_i - \overline{x})^2, \quad s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{m_y} n_{yi}(y_i - \overline{y})^2;$$

2) знаходимо критичну точку f_{cr} як розв'язок рівняння

$$\mathrm{P}\{F > f_{cr}\} = \alpha / 2; \tag{3.19}$$

3) робимо висновок про висунуту гіпотезу: якщо $F_{\mathring{a}mp} < f_{cr}$, то гіпотезу H_0 приймаємо; якщо $F_{\mathring{a}mp} > f_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Рівняння (3.19) запишемо у вигляді, зручному для розв'язування в програмному середовищі *Mathematica.*

$$1 - \int_0^{f_{cr}} f(x, k_1, k_2)dx = \alpha / 2. \tag{3.20}$$

Якщо альтернативною є гіпотеза $H_1 : \sigma_x^2 > \sigma_y^2$, то критичну точку шукаємо з рівняння:

$$\mathrm{P}\{F > f_{cr}\} = \alpha,$$

яке еквівалентне такому

$$1 - \int_0^{f_{cr}} f(x, k_1, k_2)dx = \alpha.$$

Якщо $F_{\mathring{a}mp} < f_{cr}$, то гіпотезу H_0 приймаємо; якщо $F_{\mathring{a}mp} > f_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1 : \sigma_x^2 > \sigma_y^2$.

ПОЯСНЕННЯ для k=50

Програми записуємо у вигляді:

```
Tmod EQU 31  ; час моделювання
GisX TABLE (Normal(50,250,4)),0,2,500  ; гістограма X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1

Tmod EQU 21  ; час моделювання
GisY TABLE (Normal(50,150,4)),0,2,500  ; гістограма Y
GENERATE 1
TABULATE GisY
TERMINATE
GENERATE Tmod
```

```
TERMINATE 1
START 1
```

Зі стандартних звітів вибираємо фрагменти з гістограмами:

```
TABLE               MEAN    STD.DEV.       RANGE            RETRY FREQUENCY CUM.%
 GISX             249.828    3.293                            0
                                  244.000  -      246.000                5    16.67
                                  246.000  -      248.000                5    33.33
                                  248.000  -      250.000                8    60.00
                                  250.000  -      252.000                2    66.67
                                  252.000  -      254.000                7    90.00
                                  254.000  -      256.000                3   100.00

TABLE               MEAN    STD.DEV.       RANGE            RETRY FREQUENCY CUM.%
 GISY             150.144    3.477                            0
                                  144.000  -      146.000                3    15.00
                                  146.000  -      148.000                3    30.00
                                  148.000  -      150.000                6    60.00
                                  150.000  -      152.000                1    65.00
                                  152.000  -      154.000                4    85.00
                                  154.000  -      156.000                3   100.00
```

Користуючись гістограмами, записуємо дискретні статистичні розподіли частот, попередньо взявши за x_i, y_i середини інтервалів:

x_i	245	247	249	251	253	255
n_{xi}	5	5	8	2	7	3

y_i	145	147	149	151	153	155
n_{yi}	3	3	6	1	4	3

Обчислюємо вибіркові середні ($\overline{x} = 249{,}667,\ \overline{y} = 149{,}9$) і підправлені вибіркові дисперсії ($s_x^2 = 10{,}8506,\ s_y^2 = 11{,}5684$). Оскільки $s_y^2 \geq s_x^2$, то приймаємо $s_1^2 = s_y^2,\ s_2^2 = s_x^2$ і обчислюємо емпіричне значення критерію за формулою (3.18).

Оскільки альтернативною є гіпотеза $H_1 : \sigma_x^2 \neq \sigma_y^2$, то критичну точку шукаємо з рівняння (3.20). Розв’язування рівняння (3.20) здійснюється за допомогою вбудованої функції **FindRoot[lp==rp,{x,x0}].** Оскільки

$$f_{cr} = 2{,}23127, \quad F_{åmp} = 1{,}06616,$$

то $F_{åmp} < f_{cr}$, і гіпотезу H_0 приймаємо.

```
Варіанти (значення x[i]):
 x[1]:=245
 x[2]:=247
 x[3]:=249
 x[4]:=251
 x[5]:=253
 x[6]:=255
Емпіричні частоти:
 n[1]:=5
```

```
 n[2]:=5
 n[3]:=8
 n[4]:=2
 n[5]:=7
 n[6]:=3
```

Вибіркове середнє X

```
 xav:=N[Sum[x[i]n[i],{i,1,6}]/30]
 xav
 249.667
```

Варіанти (значення y[i]):

```
 y[1]:=145
 y[2]:=147
 y[3]:=149
 y[4]:=151
 y[5]:=153
 y[6]:=155
```

Емпіричні частоти:

```
 m[1]:=3
 m[2]:=3
 m[3]:=6
 m[4]:=1
 m[5]:=4
 m[6]:=3
```

Вибіркове середнє Y

```
 yav:=N[Sum[y[i]m[i],{i,1,6}]/20]
 yav
 149.9
```

Підправлена вибіркова дисперсія X

```
 sxkv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,6}]/29]
 sxkv
 10.8506
```

Підправлена вибіркова дисперсія Y

```
 sykv:=N[Sum[(m[i])((y[i]-yav)^2),{i,1,6}]/19]
 sykv
 11.5684
```

Емпіричне значення критерію

```
 Femp:=sykv/sxkv
 Femp
 1.06616
```

Розподіл Фішера-Снедекора

```
C0[a_,b_]:=(Gamma[(a+b)/2]((a)^((a)/2))((b)^((b)/2)))/(Ga
mma[(a)/2]Gamma[(b)/2])
 f[x_,k1_,k2_]:=(C0[k1,k2])(x^((k1-
2)/2))/((k2+(k1)x)^((k1+k2)/2))
```

Розв’язування рівняння для критичної точки

```
 Fcr[beta_,k1_,k2_]:=FindRoot[1-
Integrate[f[t,k1,k2],{t,0,x}]==beta,{x,2}]
 Fcr[0.025,19,29]
 {x→2.23127}
Femp<Fcr, гіпотезу приймаємо.
```

ЗАДАЧА 3.8
Перевірка статистичної гіпотези про значення дисперсії нормального закону розподілу

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 21  ; час моделювання
Z Variable (Exponential(k,0,1))
Iks Variable
(84+k+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z+V$Z
+V$Z+V$Z+V$Z)
GisX TABLE (V$Iks),0,4,500  ; гістограма X
GENERATE 1
TABULATE GisX
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, взяти за значення x_i середини інтервалів $(z_{i-1}, z_i]$ і записати дискретний статистичний розподіл частот для отриманої вибірки. За даними вибірки обчислити вибіркове середнє $\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$ та підправлене середнє квадратичне відхилення s за формулою (3.1). Вважаючи, що дисперсія розподілу невідома, для рівня значущості $\alpha = 0{,}001 \cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 16$ *про рівність дисперсії випадкової величини X гіпотетичному значенню* σ_0^2 у двох випадках: 1) за наявності альтернативної гіпотези $H_1: \sigma^2 \neq \sigma_0^2$; 2) за наявності альтернативної гіпотези $H_1: \sigma^2 < \sigma_0^2$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай генеральна сукупність X розподілена нормально, причому генеральна дисперсія хоч і невідома, але є підстави припускати, що вона дорівнює гіпотетичному значенню σ_0^2.

Припустимо, що з генеральної сукупності отримано вибірку обсягу n і за її даними знайдено підправлену вибіркову дисперсію s^2. Необхідно для заданого рівня значущості α перевірити статистичну гіпотезу, яка полягає в тому, що генеральна дисперсія даної сукупності дорівнює гіпотетичному значенню σ_0^2: $H_0 : \sigma^2 = \sigma_0^2$.

За критерій перевірки висунутої гіпотези приймемо випадкову величину

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

розподілену за законом χ^2 з $k = n-1$ ступенями вільності.

Якщо альтернативною є гіпотеза $H_1 : \sigma^2 > \sigma_0^2$, то розглядаємо правосторонню критичну область і критичну точку шукаємо з рівняння

$$\mathrm{P}\{\chi^2 > \chi_{cr}^2\} = \alpha. \tag{3.21}$$

Використовуючи щільність розподілу χ^2

$$f(x,k) = \frac{x^{(k/2)-1} e^{-x/2}}{2^{k/2}\Gamma(k/2)}, \quad x > 0,$$

рівняння (3.21) можна записати у вигляді

$$1 - \int_0^{\chi_{cr}^2} f(x,k)dx = \alpha.$$

На основі даних вибірки обчислюємо емпіричне значення критерію

$$\chi_{emp}^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{3.22}$$

і робимо висновок про висунуту гіпотезу: якщо $\chi_{emp}^2 < \chi_{cr}^2$, то гіпотезу H_0 приймаємо; якщо $\chi_{emp}^2 > \chi_{cr}^2$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Якщо альтернативною є гіпотеза $H_1 : \sigma^2 \neq \sigma_0^2$, то розглядаємо двосторонню критичну область, і критичні точки шукаємо з рівнянь

$$\mathrm{P}\{\chi^2 < \chi_{l.cr}^2\} = \mathrm{P}\{\chi^2 > \chi_{r.kr}^2\} = \frac{\alpha}{2}. \tag{3.23}$$

Рівняння (3.23) можна записати у вигляді

$$1 - \int_0^{\chi_{l.cr}^2} f(x,k)dx = 1 - \alpha/2, \qquad 1 - \int_0^{\chi_{r.cr}^2} f(x,k)dx = \alpha/2. \tag{3.24}$$

Якщо $\chi^2_{l.cr} < \chi^2_{emp} < \chi^2_{r.cr}$, то гіпотезу H_0 приймаємо; якщо $\chi^2_{emp} \notin (\chi^2_{l.cr}, \chi^2_{r.cr})$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: \sigma^2 \neq \sigma_0^2$.

Якщо альтернативною є гіпотеза $H_1: \sigma^2 < \sigma_0^2$, то розглядаємо лівосторонню критичну область, і критичну точку шукаємо з рівняння

$$\mathrm{P}\{\chi^2 < \chi^2_{cr}\} = \alpha. \tag{3.25}$$

Рівняння (3.25) можна записати у вигляді

$$1 - \int_0^{\chi^2_{cr}} f(x,k)dx = 1 - \alpha. \tag{3.26}$$

Якщо $\chi^2_{emp} > \chi^2_{cr}$, то гіпотезу H_0 приймаємо; якщо $\chi^2_{emp} < \chi^2_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: \sigma^2 < \sigma_0^2$.

ПОЯСНЕННЯ для k=50

Зі стандартного звіту GPSS World вибираємо гістограму:

```
TABLE   MEAN STD.DEV.  RANGE            RETRY    REQUENCY   CUM.%
GISX 150.371 4.720                        0
                       144.000 - 148.000                6      30.00
                       148.000 - 152.000                9      75.00
                       152.000 - 156.000                3      90.00
                       156.000 - 160.000                0      90.00
                       160.000 - 164.000                2     100.00
```

Взявши за x_i середини інтервалів, записуємо дискретний розподіл частот:

x_i	146	150	154	158	162
n_i	6	9	3	0	2

Обчислюємо вибіркове середнє ($\overline{x} = 150{,}6$), підправлену вибіркову дисперсію ($s^2 = 22{,}3579$) та емпіричне значення критерію за формулою (3.22).

1) Оскільки альтернативною є гіпотеза $H_1: \sigma^2 \neq \sigma_0^2$, то критичні точки шукаємо з рівнянь (3.24). Оскільки

$$\chi^2_{emp} = 26{,}55, \quad \chi^2_{l.cr} = 8{,}90652, \quad \chi^2_{r.cr} = 32{,}8523,$$

то $\chi^2_{l.cr} < \chi^2_{emp} < \chi^2_{r.cr}$ і гіпотезу H_0 приймаємо.

2) Оскільки альтернативною є гіпотеза $H_1: \sigma^2 < \sigma_0^2$, то критичну точку шукаємо з рівняння (3.26). Оскільки $\chi^2_{cr} = 10{,}117$, то $\chi^2_{emp} > \chi^2_{cr}$ і гіпотезу H_0 приймаємо.

Розв'язування рівнянь (3.24) і (3.26) здійснюється за допомогою вбудованої функції **`FindRoot[lp==rp,{x,x0}]`**.

Варіанти (значення x[i]):

```
x[1]:=146
x[2]:=150
x[3]:=154
x[4]:=158
x[5]:=162
```

Емпіричні частоти:

```
n[1]:=6
n[2]:=9
n[3]:=3
n[4]:=0
n[5]:=2
```

Вибіркове середнє

```
xav:=N[Sum[x[i]n[i],{i,1,5}]/20]
xav
150.6
```

Підправлена вибіркова дисперсія

```
skv:=N[Sum[(n[i])((x[i]-xav)^2),{i,1,5}]/19]
skv
22.3579
sg0kv:=16
```

Емпіричне значення критерію

```
Xikvemp:=(19)skv/sg0kv
Xikvemp
26.55
```

Розподіл Xi квадрат

```
f[x_,k_]:=(x^(k/2-1)Exp[-x/2])/((2^(k/2))Gamma[k/2])
alpha:=0.05
1-alpha/2
0.975
```

Розв'язування рівняння для критичної точки

```
Xicr[alpha_,k_]:=FindRoot[1-
Integrate[f[t,k],{t,0,x}]==alpha,{x,20}]
```

Ліва критична точка (випадок 1)

```
Xicr[0.975,19]
{x→8.90652}
```

Права критична точка (випадок 1)

```
alpha/2
0.025
Xicr[0.025,19]
{x→32.8523}
```

8.906516481987993`<(Xikvemp=26.549999999999997`)<32.85232 68617297`, гіпотезу приймаємо (Випадок 1)

```
alpha:=0.05
1-alpha
```

```
 0.95
```
Критична точка (випадок 2)

Xicr[0.95,19]

```
 {x→10.117}
```
Xikvemp>Xicr, гіпотезу приймаємо (випадок 2).

ЗАДАЧА 3.9
Перевірка статистичної гіпотези про рівність нулю вибіркового коефіцієнта кореляції між нормальними генеральними сукупностями

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 5001  ; час моделювання
Iks Variable (Normal(50+k,30,8))
Y Variable (Normal(50+k,40,10))
GisX TABLE V$Iks,0,10,500  ; гістограма X
GisY TABLE V$Y,0,10,500  ; гістограма Y
GisXY TABLE ((V$Iks)#(V$Y)),0,500,500 ; гістограма XY
GENERATE 1
TABULATE GisX
TABULATE GisY
TABULATE GisXY
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограми з одержаного стандартного звіту, і взявши за значення $\overline{x}$, $\overline{y}$, $\overline{xy}$, σ_{Bx} і σ_{By} ті значення, які автоматично обчислені в гістограмах GPSS World, обчислити вибірковий коефіцієнт кореляції $r(X,Y)$ за формулою:

$$r(X,Y)=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_{Bx}\sigma_{By}}.$$

Для рівня значущості α=0,001 перевірити гіпотезу $H_0: r(X,Y)=0$ за наявності альтернативної гіпотези $H_1:\ r(X,Y)\neq 0.$

Теоретичні відомості

Нехай двовимірна генеральна сукупність (X,Y) розподілена нормально. З цієї сукупності отримано вибірку обсягу n і за її даними знайдено вибірковий коефіцієнт кореляції $r(X,Y)$, який виявився відмінним від нуля. Оскільки вибірка відібрана випадково, то ще неможливо зробити висновок, що коефіцієнт кореляції генеральної сукупності $\rho(X,Y)$ також відмінний від нуля. Виникає необхідність для заданого рівня значущості α перевірити статистичну гіпотезу $H_0 : \rho(X,Y) = 0$ *про рівність нулю генерального коефіцієнта кореляції* за наявності альтернативної гіпотези $H_1 : \rho(X,Y) \neq 0$.

Якщо гіпотеза H_0 буде прийнята, то вибірковий коефіцієнт кореляції незначущий, а X і Y некорельовані, тобто не зв'язані лінійною залежністю.

За критерій перевірки гіпотези H_0 приймемо випадкову величину

$$T = \frac{r(X,Y)\sqrt{n-2}}{\sqrt{1-r^2(X,Y)}}.$$

Величина T за умови істинності гіпотези H_0 має розподіл Ст'юдента з $k = n-2$ ступенями вільності, який характеризується щільністю розподілу

$$s(x,k) = B_k \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \qquad x \in (-\infty, +\infty),$$

де $B_k = \dfrac{\Gamma((k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)}, \quad \Gamma(x) = \int\limits_0^{\infty} t^{x-1} e^{-t} dt.$

Перевірку гіпотези H_0 здійснюємо за схемою: 1) обчислюємо емпіричне значення критерію за формулою

$$T_{emp} = \frac{r(X,Y)\sqrt{n-2}}{\sqrt{1-r^2(X,Y)}}, \tag{3.27}$$

де

$$r(X,Y) = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y}}{\sigma_{\hat{A}x}\sigma_{\hat{A}y}} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma_{\hat{A}x}\sigma_{\hat{A}y}},$$
$$\sigma_{\hat{A}x}^2 = \frac{1}{n}\sum_{i=1}^{m_x} n_{xi}(x_i - \overline{x})^2, \quad \sigma_{\hat{A}y}^2 = \frac{1}{n}\sum_{i=1}^{m_y} n_{yi}(y_i - \overline{y})^2. \tag{3.28}$$

2) знаходимо критичну точку z_{cr} як розв'язок рівняння

$$\mathrm{P}\{|T| > t_{cr}\} = \alpha; \tag{3.29}$$

3) робимо висновок про висунуту гіпотезу: якщо $|T_{\mathring{a}mp}| < t_{cr}$, то гіпотезу H_0 приймаємо; якщо $|T_{\mathring{a}mp}| > t_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Рівняння (3.29) запишемо у вигляді, зручному для розв'язування в програмному середовищі *Mathematica.*

$$1-2\int_{0}^{t_{cr}} s(x,k)dx=\alpha. \tag{3.30}$$

ПОЯСНЕННЯ для k=0

Зі стандартного звіту GPSS World вибираємо гістограми:

```
TABLE   MEAN  STD.DEV.  RANGE            RETRY    REQUENCY   CUM.%
GISX 29.955   7.912                        0
                     0.000 -  10.000                  30      0.60
                    10.000 -  20.000                 504     10.68
                    20.000 -  30.000                2011     50.90
                    30.000 -  40.000                1952     89.94
                    40.000 -  50.000                 471     99.36
                    50.000 -  60.000                  32    100.00
GISY 40.083  10.158                        0
                         _ -   0.000                   1      0.02
                     0.000 -  10.000                   8      0.18
                    10.000 -  20.000                 124      2.66
                    20.000 -  30.000                 640     15.46
                    30.000 -  40.000                1655     48.56
                    40.000 -  50.000                1778     84.12
                    50.000 -  60.000                 669     97.50
                    60.000 -  70.000                 114     99.78
                    70.000 -  80.000                  11    100.00
GISXY 1199.776 444.196                     0
                     0.000 - 500.000                 186      3.72
                   500.000 -1000.000                1558     34.88
                  1000.000 -1500.000                2072     76.32
                  1500.000 -2000.000                 945     95.22
                  2000.000 -2500.000                 207     99.36
                  2500.000 -3000.000                  27     99.90
                  3000.000 -3500.000                   5    100.00
```

З гістограм отримуємо таку інформацію (виділено жирним шрифтом):

$\overline{x}=29{,}995,\ \ \overline{y}=40{,}083,\ \ \overline{xy}=1199{,}776,\ \ \sigma_{Bx}=7{,}912,\ \ \sigma_{By}=10{,}158,\ \ n=5000.$

За формулою (3.28) знаходимо $r(X,Y)=-0{,}0113259$. Обчислюємо T_{emp} за формулою (3.27). Критичну точку шукаємо з рівняння (3.30). Розв'язування рівняння (3.30) здійснюється за допомогою вбудованої функції

FindRoot[lp==rp,{x,x0}].

Оскільки

$$t_{cr}=3{,}29247,\qquad T_{\mathring{a}mp}=-0{,}800754,$$

то $|T_{\mathring{a}mp}|<t_{cr}$, і гіпотезу H_0 приймаємо.

Значення середніх:

xyav:=1199.776

xav:=29.955

```
 yav:=40.083
 sigx:=7.912
 sigy:=10.158
Вибірковий коефіцієнт кореляції
 r:=(xyav-(xav)(yav))/((sigx)(sigy))
 r
 -0.0113259
Емпіричне значення критерію
 Temp:=(r)Sqrt[4998/(1-r^2)]
 Temp
 -0.800754
Розподіл Ст'юдента
 B[n_]:=Gamma[(n+1)/2]/(Sqrt[Pi(n)]Gamma[n/2])
 S[x_,n_]:=B[n](1+(x^2)/n)^(-(n+1)/2)
Розв'язування рівняння для критичної точки
 Tcr2[alpha_,k_]:=FindRoot[1-
2Integrate[S[t,k],{t,0,x}]==alpha,{x,1}]
 Tcr2[0.001,4998]
 {x→3.29247}
|Temp|<Tcr, гіпотезу приймаємо.
```

ЗАДАЧА 3.10
Перевірка статистичної гіпотези про значення ймовірності появи події в незалежних випробуваннях за схемою Бернуллі

Реалізувати програму GPSS World, спочатку підставивши замість k свій № варіанту:

```
Tmod EQU 101  ; час моделювання
Rozp FUNCTION RNk,D6
.2,6/.6,5/.7,4/.8,3/.9,2/1,1  ;випадкова величина
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Використовуючи гістограму з одержаного стандартного звіту, взяти за значення відносної частоти події $A=\{X=6\}$ в n випробуваннях за схемою Бернуллі відношення m/n, де $m-$частота події A, взята з останнього рядка гістограми, $n-$обсяг вибірки (сума відносних частот гістограми). Для рівня значущості $\alpha = 0{,}001\cdot k$ (k – номер варіанту) перевірити гіпотезу $H_0: p = p_0 = 0{,}2$ *про рівність імовірності* $p = P(A)$ *гіпотетичному значенню* p_0 у двох випадках: 1) за наявності альтернативної гіпотези H_1: $p \neq p_0$; 2) за наявності альтернативної гіпотези $H_1: \; p < p_0$. Обчислення здійснювати за допомогою пакету *Mathematica*.

Теоретичні відомості

Нехай за достатньо великою кількістю n незалежних випробувань, в кожному з яких імовірність p появи події стала, але невідома, знайдено відносну частоту m/n. Є підстави вважати, що невідома ймовірність дорівнює гіпотетичному значенню p_0. Необхідно за заданим рівнем значущості α перевірити статистичну гіпотезу $H_0: p = p_0$.

За критерій перевірки гіпотези H_0 приймемо випадкову величину

$$U = \left(\frac{m}{n} - p_0\right)\sqrt{\frac{n}{p_0(1-p_0)}}\,.$$

Величина U за умови істинності гіпотези H_0 розподілена наближено нормально з параметрами $E(U) = 0,\ \sigma(U) = 1$.

Емпіричне значення критерію обчислюємо за формулою

$$U_{emp} = \left(\frac{m}{n} - p_0\right)\sqrt{\frac{n}{p_0(1-p_0)}}\,. \tag{3.31}$$

У випадку альтернативної гіпотези $H_1: \; p \neq p_0$ знаходимо критичну точку u_{cr} як розв'язок рівняння

$$\Phi(u_{cr}) = \frac{1-\alpha}{2}. \tag{3.32}$$

Якщо $|U_{\text{åmp}}| < u_{cr}$, то гіпотезу H_0 приймаємо; якщо $|U_{\text{ài ï}}| > u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи H_1.

Якщо альтернативною є гіпотеза $H_1: p > p_0$, то критичну точку шукаємо з рівняння:

$$\Phi(u_{cr}) = \frac{1-2\alpha}{2}. \tag{3.33}$$

Якщо $U_{\text{åmp}} < u_{cr}$, то гіпотезу H_0 приймаємо; якщо $U_{\text{åmp}} > u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: p > p_0$.

Якщо альтернативною є гіпотеза $H_1: p < p_0$, то критичну точку шукаємо як розв'язок рівняння (3.33). Якщо $U_{\text{åmp}} > -u_{cr}$, то гіпотезу H_0 приймаємо; якщо $U_{\text{åmp}} < -u_{cr}$, то відхиляємо гіпотезу H_0 на користь альтернативи $H_1: p < p_0$.

ПОЯСНЕННЯ для k=50

Програму записуємо у вигляді:

```
Tmod EQU 101  ; час моделювання
Rozp FUNCTION RN50,D6
.2,6/.6,5/.7,4/.8,3/.9,2/1,1  ;випадкова величина
Gis TABLE Fn$Rozp,0,1,500  ; гістограма
GENERATE 1
TABULATE Gis
TERMINATE
GENERATE Tmod
TERMINATE 1
START 1
```

Зі стандартного звіту GPSS World вибираємо гістограму:

```
TABLE   MEAN  STD.DEV.   RANGE           RETRY    REQUENCY   CUM.%
GIS    4.230   1.575                       0
                     0.000 - 1.000                   10       10.00
                     1.000 - 2.000                   11       21.00
                     2.000 - 3.000                    4       25.00
                     3.000 - 4.000                   13       38.00
                     4.000 - 5.000                   45       83.00
                     5.000 - 6.000                   17      100.00
```

З гістограми отримуємо:

$$m = 17,\quad n = 10 + 11 + 4 + 13 + 45 + 17 = 100,\quad m/n = 0{,}17.$$

Обчислюємо емпіричне значення критерію за формулою (3.31): $U_{emp} = -0{,}75$.

1) Оскільки альтернативною є гіпотеза $H_1: p \neq p_0$, то критичну точку шукаємо з рівнянь (3.32). Оскільки $u_{cr} = 1{,}95996$, то $|U_{åmp}| < u_{cr}$ і гіпотезу H_0 приймаємо.

2) Оскільки альтернативною є гіпотеза $H_1: p < p_0$, то критичну точку шукаємо з рівняння (3.33). Оскільки $u_{cr} = 1{,}64485,$ то $U_{åmp} > -u_{cr}$ і гіпотезу H_0 приймаємо.

Розв'язування рівнянь (3.32) і (3.33) здійснюється за допомогою вбудованої функції **Solve[eq,var]**.

```
 p0:=0.2
 n:=100
Емпіричне значення критерію
 Uemp:=(0.17-p0)Sqrt[n/(p0(1-p0))]
 Uemp
 -0.75
Розв'язування рівняння для критичної точки
 Ucr[beta_]:=Solve[Integrate[Exp[-
(z^2)/2],{z,0,u}]/(Sqrt[2Pi])==beta,u]
Випадок 1
```

```
 beta:=(1-alpha)/2
 alpha:=0.05
```
Значення критичної точки
```
 Ucr[beta]
 {{u→1.95996}}
```
|Uemp|<Ucr, гіпотезу приймаємо.

Випадок 2
```
 beta:=(1-2(alpha))/2
 beta
 0.45
```
Значення критичної точки
```
 Ucr[beta]
 {{u→1.64485}}
```
Uemp>-Ucr, гіпотезу приймаємо.