МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

РОЗРАХУНКОВО-ГРАФІЧНА РОБОТА

з курсу «Ймовірнісні моделі та статистичне

оцінювання в інформаційно-управляючих системах»

Викладач:
Вєчерковська
А.С.

Виконав:
студент 2 курсу групи
ІП-15 ФІОТ
Мєшков Андрій
Ігорович

Київ-2023

Розрахунково-графічна робота
Завдання 1

Завдання:

- За вибіркою з нормальної генеральної сукупності критерієм χ^2 перевірити гіпотезу про нормальний розподіл;
- Побудувати довірчі інтервали для: математичного сподівання при відомій дисперсії; математичного сподівання при невідомій дисперсії; для дисперсії;
- Задати структуру та параметри (кожному індивідуально) ідеальної багатовимірної лінійної регресії, реалізувати віртуальний активний експеримент і за його результатами методом найменших квадратів знайти оцінки її коефіцієнтів.

Хід розрахунків показаний у табличному файлі.

Хід роботи:

1. За вибіркою з нормальної генеральної сукупності критерієм χ^2 перевірити гіпотезу про нормальний розподіл;

Індивідуально було створена вибірка з нормальною генеральною сукупністю з n=60 елементів.

18	20	14	11	8	12
12	15	11	12	21	14
12	11	6	9	13	12
18	11	13	12	11	16
13	13	9	18	17	20
7	19	10	8	12	12
11	16	26	18	15	18
8	21	15	13	12	14
13	16	15	12	14	14
13	12	17	13	15	15

Для розрахунків був побудований варіаційний ряд. Були знайдені кількість інтервалів m=$\sqrt{n}$, та крок інтервалу $h = \frac{x_{max}-x_{min}}{m}$, де в чисельнику різниця максимального та мінімального елементів вибірки. Були побудовані інтервали [x_1, x_2),…,[x_m, x_{m+1}]. Були знайдені центри частинних інтервалів x_i^*=x_{i-1}+$\frac{h}{2} = \frac{x_m+x_{m+1}}{2}$.

Нульова гіпотеза H_0 – вибірка розподілена за нормальним законом, H_1 – вибірка не розподілена за нормальним законом, при рівні значущості $\alpha = 0{,}05$.

Обчислюємо вибіркову середню(x* average), $\overline{x_B} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$. Обчислюємо виправлену дисперсію s^2=$(\overline{x_B{}^2} - (\overline{x_B})^2) * \frac{n}{n-1}$, де $\overline{x_B{}^2} = \frac{1}{n}\sum_{i=1}^{k} x_i{}^2 n_i$. Знаходимо правлене вибіркове середнє квадратичне відхилення $\sigma = \sqrt{s^2}$.

$\overline{x_B}$ =13,9166667
$\overline{x_B{}^2}$ =207,354167
s^2= 13,9124294
σ =3,72993691

Нам необхідно знайти критерій χ^2-Пірсона, для цього з припущенням, що розподіл нормальним шукаємо теоретичні частоти n′ за таблицею. Маючи необхідні параметри, нормалізуємо випадкову величину X – переходимо до величини Z і обчислюємо інтервали (z_i;z_{i+1}), де $z_i = \frac{(x_i-\overline{x^*})}{\sigma^*}, z_{i+1} = \frac{(x_{i+1}-\overline{x^*})}{\sigma^*}, z_1 = -\infty, z_l = \infty$. Обчислюємо теоретичні ймовірності потрапляння

Z з інтервалами $(z_i;z_{i+1})$, $P_i = \Phi(z_{i+1}) - \Phi(z_i)$, де Φ(x) – функція Лапласа, значення якої ми знаходимо з таблиці. Знаходимо теоретичні частоти $n'_i=nP_i$.

Маючи теоретичні частоти знаходимо спостережуване значення критерію $\chi^2_{сп}=\sum_{i=1}^{l}\frac{(n_i-n'_i)^2}{n'_i}$.

Область прийняття нульової гіпотези визначаємо нерівністю: $\chi^2_{сп}= \chi^2_{кр}(\alpha, k)$, k=l-3=m-3, критичну точку знаходимо в таблиці.

За нашими даними $\chi^2_{сп}$= 15,3994748> $\chi^2_{кр}(0{,}05, 5)$=11,07, тому ми маємо підстави відкинути нульову гіпотезу.

2. Побудувати довірчі інтервали для: математичного сподівання при відомій дисперсії; математичного сподівання при невідомій дисперсії; для дисперсії;

Для обчислень ми будемо використовувати дані з попереднього завдання.

18	20	14	11	8	12
12	15	11	12	21	14
12	11	6	9	13	12
18	11	13	12	11	16
13	13	9	18	17	20
7	19	10	8	12	12
11	16	26	18	15	18
8	21	15	13	12	14
13	16	15	12	14	14
13	12	17	13	15	15

n=60

$\overline{x_B}$ =13,9166667

s^2= 13,9124294

σ =3,72993691

Також візьмемо надійність γ = 0,95.

При відомому параметрі σ довірчим для математичного сподівання є інтервал:

$$\left(\overline{x_B} - t\frac{\sigma}{\sqrt{n-1}}; \overline{x_B} + t\frac{\sigma}{\sqrt{n-1}}\right),$$

де t – розв'язок рівняння 2Ф(t)= γ. В нашому випадку t=1,96, n-1, бо маємо вибіркову дисперсію у розподілі Стьюдента, а не популяційну у нормальному розподілі.

Довірчий інтервал для математичного сподівання при відомій дисперсії (12,9648977; 14,8684357).

При невідомому параметрі σ довірчим для математичного сподівання припускається стандартний випадок нормального розподілу, де σ=1, тому можна скорегувати минулу формулу:

$$\left(\overline{x_B} - t\frac{1}{\sqrt{n-1}}; \overline{x_B} + t\frac{1}{\sqrt{n-1}}\right),$$

що дає не дуже точні результати: (13,6614964; 14,1718369).

При невідомому математичному сподіванні довірчий інтервал із надійністю γ для дисперсії D(X) нормального розподілу має вигляд:

$$\left(\frac{(n-1)s^2}{\chi_1^2(\gamma)}; \frac{(n-1)s^2}{\chi_2^2(\gamma)}\right),$$

Де додатні числа $\chi_1(\gamma), \chi_2(\gamma)$ визначаються з рівностей $F(\chi_1^2(\gamma))=\frac{1-\gamma}{2}$; $F(\chi_2^2(\gamma))=\frac{1+\gamma}{2}$;

Де F(x) — функція розподілу випадкової величини χ^2(хі-квадрат) з (n−1) ступенями свободи.

Для дисперсії довірчий інтервал: (9,99590016; 20,6957121).

3. Задати структуру та параметри (кожному індивідуально) ідеальної багатовимірної лінійної регресії, реалізувати віртуальний активний експеримент і за його результатами методом найменших квадратів знайти оцінки її коефіцієнтів.

Метод найменших квадратів для багатовимірної лінійної регресії використовується для знаходження оптимальних параметрів моделі, яка описує залежність між залежною змінною і декількома незалежними змінними. Основна ідея методу найменших квадратів полягає в мінімізації суми квадратів розбіжностей між спостережуваними значеннями залежної змінної і прогнозованими значеннями, обчисленими за допомогою лінійної функції з набором незалежних змінних. У багатовимірній лінійній регресії модель може бути виражена наступним чином:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$, де y - залежна змінна, $x_1, x_2, ..., x_k$ - незалежні змінні (фактори), $\beta_0, \beta_1, \beta_2, ..., \beta_k$ - параметри моделі, ε - помилка. Y=XB+E, як матричний вигляд

Метод найменших квадратів для багатовимірної лінійної регресії полягає в знаходженні таких значень параметрів β_0, β_1, β_2, ..., β_k, які мінімізують суму квадратів розбіжностей між фактичними значеннями залежної змінної і прогнозованими значеннями.

Реалізуємо віртуальний експеримент, для цього створимо дані x_1, x_2 з результатами y. Кожна сукупність має 50 елементів. Для нашого експерименту повинна виконуватись рівняння $y=\beta_0 + \beta_1 x_1 + \beta_2 x_2$, ε – можна припустити нульове значення для ідеального випадку.

Далі створимо розширену матриці предикторів, для цього додаємо одиничний масив як перший стовпець ($\beta_0= \beta_0 * x_0$, $x_0=1$). Транспонуємо матрицю і отримуємо X^T.

Суть нашого експерименту вирішити матричне рівняння $Y=X*B$, $X^TY=X^T*X*B$, $B=(X^T*X)^{-1} *X^T*Y$.

Маючи X^T, побудуємо матриці X^TX, X^TY, $(X^TX)^{-1}$, $(X^TX)^{-1}* X^TY$, використавши функції Excel та Google sheets для транспонування, обернення та множення матриць.

Знаходимо оцінки коефіцієнтів – це одномірний масив:

B=

295.3094267
-0.1468351926
-0.1098866551

Умовно маємо рівняння $y=295,31-0,15x_1-0,11x_2$.