

Uloha4

Regression Equation Section

| Independent Variable | Regression Coefficient b(i) | Standard Error Sb(i) |
|----------------------|-----------------------------|----------------------|
| Intercept | 15.814 | 0.748 |
| x1 | -1.015 | 0.063 |
| x2 | 0.065 | 0.003 |

Analysis of Variance Section

| Source | DF | Sum of Squares | Mean Square | F-Ratio | Prob | Level |
|--------|----|----------------|-------------|---------|--------|-------|
| Model | 2 | 102.77 | 51.38 | 207.6 | 0.0000 | |
| Error | 12 | 2.97 | 0.247 | | | |

Napište tvar regresního modelu pro tento příklad

Tvar regresního modelu se skládá z koeficientů (sloupec $b(i)$) pro **Intercept** a **nezávislé proměnné** (x_1, x_2).

Regresní rovnice:

$$\hat{y} = 15.814 - 1.015x_1 + 0.065x_2$$

Interpretace koeficientů:

- **Intercept** ($b_0 = 15.814$):
Očekávaná hodnota závislé proměnné (\hat{y}), pokud jsou obě nezávislé proměnné (x_1 a x_2) rovny nule.
- x_1 ($b_1 = -1.015$):
Zvýšení x_1 o jednu jednotku vede ke snížení \hat{y} o 1.015 jednotek, za předpokladu, že x_2 zůstane konstantní (ceteris paribus).
- x_2 ($b_2 = 0.065$):
Zvýšení x_2 o jednu jednotku vede ke zvýšení \hat{y} o 0.065 jednotek, za předpokladu, že x_1 zůstane konstantní (ceteris paribus).

Jaký byl rozsah výběru (kolik řádků v datové matici)?

Rozsah výběru (N) se vypočítá pomocí stupňů volnosti (DF) z tabulky:

- DF pro Model je $k = 2$ (počet nezávislých proměnných x_1, x_2).

- DF pro Error (rezidua) je $N - k - 1 = 12$.

Dosazením a řešením pro N :

$$N - 2 - 1 = 12$$

$$N - 3 = 12$$

$$N = 15$$

Rozsah výběru (počet řádků):

$$N = 15$$

Lze zamítnout některou z hypotéz, že regresní koeficienty jsou nulové?

Testování celého modelu (F-test):

F-Ratio: 207.6

Prob Level (p-hodnota): 0.0000

Protože je p -hodnota (0.0000) menší než jakákoli obvyklá hladina významnosti ($\alpha = 0.05$ nebo $\alpha = 0.01$), zamítáme nulovou hypotézu, že všechny regresní koeficienty (x_1, x_2) jsou nulové. Model je jako celek statisticky vysoce významný.

Testování jednotlivých koeficientů (t-test):

K ověření, zda lze zamítnout hypotézu, že daný regresní koeficient (β_i) je nulový ($H_0 : \beta_i = 0$), se používá t -test.

Hodnoty t -statistiky se počítají jako poměr koeficientu k jeho standardní chybě:

$$t_i = \frac{b_i}{S_{b_i}}$$

Výpočet kritické hodnoty:

```
from scipy.stats import t

ALPHA = 0.05

df = 12

prob = 1 - (ALPHA / 2)

t_critical = t.ppf(prob, df)

print(t_critical)
```

Zamítáme H_0 , pokud $|t_{\text{vypočítané}}| \geq 2.179$

| Koeficient | b(i) | Sb(i) | t-statistika | Interpretace |
|------------|--------|-------|--------------|----------------|
| Intercept | 15.814 | 0.748 | 21.1417 | Zamítáme H_0 |
| x_1 | -1.015 | 0.063 | -16.1111 | Zamítáme H_0 |
| x_2 | 0.065 | 0.003 | 21.6667 | Zamítáme H_0 |

Spočítejte hodnotu indexu determinace

Index determinace R^2 udává podíl variability závisle proměnné, který je vysvětlen modelem. Vypočítá se jako poměr součtu čtverců modelu SS_{Model} k celkovému součtu čtverců SS_{Total} .

$$SS_{\text{Total}} = SS_{\text{Model}} + SS_{\text{Error}}$$

$$SS_{\text{Total}} = 102.77 + 2.97 = 105.74$$

$$R^2 = \frac{SS_{\text{Model}}}{SS_{\text{Total}}} = \frac{102.77}{105.74}$$

Hodnota indexu determinace R^2 :

$$R^2 \approx 0.9719$$

Interpretace:

97.19% celkové variability závisle proměnné je vysvětleno regresním modelem – > jde o velmi silný model.

Statistický kontext

Princip testování nulovosti regresního koeficientu pomocí t -testu je založen na následujícím:

Hypotéza (H_0):

Nulová hypotéza tvrdí, že regresní koeficient (β_i) je roven nule ($H_0 : \beta_i = 0$). To by znamenalo, že nezávislá proměnná (x_i) nemá žádný lineární vliv na závisle proměnnou (y) v populaci.

Testová statistika (t):

Pro testování této hypotézy se používá t -statistika, která se vypočítá jako:

$$t = \frac{b_i}{S_{b_i}}$$

Kde b_i je odhadnutý koeficient (z výstupu regrese) a S_{b_i} je jeho standardní chyba.

Rozhodnutí:

Tato vypočtená t -hodnota se porovná s kritickou hodnotou t (při dané hladině významnosti α a příslušném počtu stupňů volnosti) nebo se použije p -hodnota (která často bývá uvedena v plném výstupu regrese). Pokud je absolutní hodnota t -statistiky vysoká (nebo p -hodnota nízká, typicky menší než 0.05), nulovou hypotézu zamítáme.