

Statystyka i analiza danych - laboratorium 8 - Korelacja i regresja

Paweł Misiorek

Instytut Informatyki
Politechnika Poznańska (PP)
Piotrowo 3, 60-965 Poznan, Poland
Email: pawel.misiorek@put.poznan.pl

25/28 kwietnia 2023

Kowariancja i korelacja

- Kowariancja, korelacja Pearsona - wzory ogólne (dla populacji):

- ▶ Kowariancja:

$$\sigma_{X,Y} = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$$

- ▶ Korelacja (współczynnik korelacji Pearsona):

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

- Kowariancja, korelacja Pearsona - estymatory (dla próby):

- ▶ Kowariancja:

$$S_{X,Y} = \frac{1}{n-1} ((X - \bar{X})(Y - \bar{Y}))$$

- ▶ Współczynnik korelacji Pearsona:

$$r_{X,Y} = \frac{S_{X,Y}}{S_X S_Y}$$

Test T na istotność korelacji, regresja jednej zmiennej

- Test na istotność korelacji - test T (o $n - 2$ stopniach swobody):
 - ▶ dla $H_0: \rho = 0$ i $H_1: \rho \neq 0$ (lub $\rho > 0$ lub $\rho < 0$)
 - ▶ dla statystyki zdefiniowanej jako:

$$T = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}$$

- Regresja liniowa jednej zmiennej $Y = \beta_0 + \beta_1 X$:
 - ▶ Wyznaczanie współczynników (metoda najmniejszych kwadratów):

$$\hat{\beta}_1 = \frac{S_{X,Y}}{S_X^2} = r_{X,Y} \frac{S_Y}{S_X}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Analiza regresji (następne laboratorium)

- Model regresji - regresja liniowa z szumem homoskedastycznym (ze stałą i niezależną wariancją)
- SST, SSR, SSE
- Współczynnik determinacji

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

- Test na istotność regresji (rozkład F Snedecora) - globalny test na istotność regresji wielorakiej
- Test na istotność pojedynczego parametru (test T)

- Test na istotność regresji (test F):
 - ▶ W globalnym teście na istotność regresji liniowej (w tym regresji wielorakiej) korzysta się ze statystyki

$$F = \frac{SSR/k}{SSE/(n - (k + 1))}$$

oraz z rozkładu F Snedecora z k i $n - k - 1$ stopniami swobody, gdzie k to liczba zmiennych wyjaśniających, a n to liczba obserwacji dla poszczególnych zmiennych.

- ▶ Uwaga 1: dla $k = 1$ test ten równoważny testowi T na istotność korelacji
- ▶ Uwaga 2: współczynnik β_0 nie wchodzi do układu hipotez.

- Test na istotność pojedynczego parametru β_i (test T):
 - ▶ W teście na istotność pojedynczego współczynnika liniowego dla regresji wielorakiej korzysta się ze statystyki

$$T = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$$

i rozkładu t-studenta o $n - k - 1$ stopni swobody (k - liczba zmiennych wyjaśniających).

- ▶ W przypadku regresji liniowej jednej zmiennej sprowadza się to do rozkładu t-studenta z $n - 2$ stopniami swobody - test ten de facto wtedy jest tożsamy na istotności korelacji Pearsona między badaną zmienną wyjaśniającą X_i , a zmienną wyjaśnianą Y .

Pytania proszę zadawać:

- w czasie konsultacji (zoom),
- na Slacku,
- albo mailowo.

