

# Linearna regresija

## ZADATAK A

U svrhu istraživanja veze između kvantitativnih vještina i vladanja znanjima iz društvenih disciplina, sprovedeno je sljedeće istraživanje. Prikupljeni su podaci o bodovima postignutim na kvalifikacijskim testovima.  $X$  predstavlja bodove s kvantitativnog dijela, a  $Y$  s društvenog dijela kvalifikacijskog testa. (G. K. Bhattacharyya, R. A. Johnson, Statistical Concepts and Methods, Wiley, 1977.) Podaci se nalaze u tablici:

$X$	41	39	53	67	61	67	46	50	55	72	63	59
$Y$	29	19	30	27	28	27	22	29	24	33	25	20
$X$	53	62	65	48	32	64	59	54	52	64	51	62
$Y$	28	22	27	22	27	28	30	29	21	36	20	29
$X$	56	38	52	40	65	61	64	64	53	51	58	65
$Y$	34	21	25	24	32	29	27	26	24	25	34	28

- (a) Prikažite podatke u Kartezijevom koordinatnom sustavu.
- (b) Za svaku varijablu posebno ispitajte je li normalno distribuirana, i to upotrebom tri kriterija: grafičkog, koji se sastoji od grafa normalnih vjerojatnosti (Q-Q plot), Lillieforsove inačice Kolmogorov-Smirnovljevog testa i Pearsonovog  $\chi^2$ -testa.
- (c) Ako pretpostavimo da podaci dolaze iz bivarijantne normalne distribucije s funkcijom gustoće

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right) \right\},$$

procijenite parametre modela  $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$  i  $\rho$  te njihove 95% pouzdane intervale.

- (d) Na istom grafu na kojem ste prikazali podatke iz (a), nacrtajte graf funkcije gustoće s prilagođenim (procijenjenim) vrijednostima iz (c) pomoću izohipsa koje obuhvaćaju područja s vjerojatnostima (volumenom) od 6/36 do 31/36 s korakom od 5/36.

- (e) Izračunajte Pearsonov koeficijent korelacije i sprovedite test da li su  $X$  i  $Y$  pozitivno korelirane. Nadalje, testirajte

$$H_0 : \rho = 0.5$$

$$H_1 : \rho \neq 0.5.$$

- (f) Koristeći se disjunktним područjima dobivenim u (d) dijelu zadatka kao razredima, sprovedite Pearsonov  $\chi^2$ -test da podaci dolaze iz bivarijatne normalne razdiobe.

## ZADATAK B

Taylorova jednadžba za vijek trajanja alata ( $y$ ) kao funkcija vremena rezanja (cutting time) ( $x$ ) glasi  $y = \alpha x^\beta$ . U članku "The Effect of Experimental Error on the Determination of Optimum Metal Cutting Conditions" (J. Eng. for Industry, 1967., str. 315-322) se navodi da ta ovisnost nije deterministička i da parametri  $\alpha$  i  $\beta$  moraju biti procijenjeni iz podataka. Podaci o mjerenju vijeka trajanja 12 karbidnih alata se nalaze u datoteci `zad50r.dat` (Devore, Jay L., Probability and Statistics for Engineering and the Sciences, 1982., Brooks/Cole Publishing Company, Monterey, California, str. 470).

- (a) Prikažite podatke  $(x, y)$  u Kartezijevom koordinatnom sustavu. Provedite prilagodbu linearnog modela  $y = \theta_0 + \theta_1 x$  podacima i dobiveni pravac prikažite na istom grafu zajedno s empirijskim podacima. Izračunajte statistiku  $R^2$  te sprovedite test osnovne hipoteze da je linearan model adekvatan.
- (b) Adekvatno transformirajte podatke iz (a) kako sugerira predloženi model te ponovite (a) dio zadatka na ovako transformiranim podacima.
- (c) Nacrtajte graf reziduala, graf standardiziranih reziduala (za transformirane podatke) te provjerite da li (standardizirani) reziduali dolaze iz jedinične normalne distribucije, i to upotrebom dva kriterija: grafičkog, koji se sastoji od grafa normalnih vjerojatnosti, te Kolmogorov-Smirnovljevog testa.
- (d) Procijenite 95% pouzdane intervale za parametre  $\theta_0$  i  $\theta_1$  prilagođenog linearnog modela transformiranim podacima.
- (e) Uz pretpostavku da je linearan model dobar za transformirane podatke, napišite kako glasi model za originalne podatke (iz (a)). Nacrtajte pripadnu regresijsku funkciju zajedno s originalnim podacima. Također, prikažite točke  $(y, \hat{y})$  zajedno s pravcem  $y = x$  u

Kartezijevom koordinatnom sustavu ( $\hat{y}$  je procjena od  $y$  na osnovu modela za originalne podatke) te odgovorite na pitanje je li taj model dobar za originalne podatke.

- (f) Na istom grafu s podacima (posebno za originalne i transformirane podatke) prikažite gornje i donje krivulje koje definiraju 95% pouzdane intervale, prvo za srednju vrijednost od  $Y$  (uz dano  $x$ ), te zatim i za  $Y$  (uz dano  $x$ ) te ih prikažite zajedno s regresijskom funkcijom u Kartezijevom koordinatnom sustavu .

### ZADATAK C

Test termičke izdržljivosti je proveden u svrhu istraživanja odnosa između temperature i vijeka trajanja vodiča (žice) prevučenog poliesterom ("Thermal Endurance of Polyester Enameled Wires Using Twisted Wire Specimens", IEEE Trans. Insulation, 1965., str. 38-44). Izvršena su mjerenja kod tri različite temperature ( $x$  =temperatura,  $y$  =vijek trajanja). Podaci o mjerenjima se nalaze u datoteci **zad52r.dat** (Devore, Jay L., Probability and Statistics for Engineering and the Sci-ences, 1982., Brooks/Cole Publishing Company, Monterey, California, str. 475).

- (a) Prikažite podatke  $(x, y)$  u Kartezijevom koordinatnom sustavu. Sugerira li graf linearnu vezu među tim podacima?
- (b) Transformirajte podatke iz (a) tako da uzmete  $x_0 = 1/x$  i  $y_0 = \ln(y)$ . Prikažite podatke  $(x_0, y_0)$  u Kartezijevom koordinatnom sustavu. Provedite prilagodbu linearnog modela  $y_0 = \theta_0 + \theta_1 x_0$  transformiranim podacima i dobiveni pravac prikažite na istom grafu zajedno s transformiranim podacima. Izračunajte statistiku  $R^2$  te sprovedite test o adekvatnosti ovog linearnog modela za transformirane podatke.
- (c) Nacrtajte graf reziduala, graf standardiziranih reziduala (za transformirane podatke) te provjerite da li (standardizirani) reziduali dolaze iz jedinične normalne distribucije, i to upotrebom dva kriterija: grafičkog, koji se sastoji od grafa normalnih vjerojatnosti, te Kolmogorov-Smirnovljevog testa.
- (d) Procijenite 95% pouzdane intervale za parametre  $\theta_0$  i  $\theta_1$  prilagođenog linearnog modela transformiranim podacima.
- (e) Uz pretpostavku da je linearan model dobar za transformirane podatke, napišite kako glasi model za originalne podatke (iz (a)). Nacrtajte pripadnu regresijsku funkciju zajedno s originalnim podacima. Također, prikažite točke  $(y, \hat{y})$  zajedno s pravcem  $y = x$  u Kartezijevom koordinatnom sustavu ( $\hat{y}$  je procjena od  $y$  na osnovu modela za originalne podatke) te odgovorite na pitanje je li taj model dobar za originalne podatke.

- (f) Na istom grafu s podacima (posebno za originalne i transformirane podatke) prikažite gornje i donje krivulje koje definiraju 95% pouzdane intervale, prvo za srednju vrijednost od  $Y$  (uz dano  $x$ ), te zatim i za  $Y$  (uz dano  $x$ ) te ih prikažite zajedno s regresijskom funkcijom u Kartezijevom koordinatnom sustavu .

**Napomena:** Ukoliko imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se mailom na [stjepan.sebek@fer.hr](mailto:stjepan.sebek@fer.hr).