

Prilagodba modela podacima i linearna regresija

ZADATAK A

Na prijelazu iz 19. u 20. stoljeće među vodećim statističarima toga doba vodila se živa rasprava o problemu vjerojatnosnog modeliranja. Posebno, postoji dobro dokumentirana korespondencija između W.F.R. Weldon (1860.-1906.), F. Galtona (1822.-1911.), F.Y. Edgewortha (1845.-1926.) i K. Pearsona (1857.-1936.) o problemu prilagodbe binomnog modela opaženim frekvencijama u pokusu bacanja igračih kocaka (E.S. Pearson (1965) Some incidents in the early history of biometry and statistics 1890-94, Biometrika, 52, 3-18). Na osnovi podataka o (I.) 7006 bacanja, odnosno (II.) 26306 bacanja 12 kocaka, u dolje navedenoj tablici navedene su frekvencije broja opaženih petica i šestica.

Broj 5-ica ili 6-ica	(I.) Frekvencija	(II.) Frekvencija
0	45	185
1	327	1149
2	886	3265
3	1475	5475
4	1571	6114
5	1404	5194
6	787	3067
7	367	1331
8	112	403
9	29	105
10	2	14
11	1	4
12	0	0

- Nacrtajte histograme podataka.
- Za svaki od uzoraka (I.) i (II.) posebno testirajte dolaze li iz binomne distribucije i to, prvo, iz $B(12, 1/3)$, te, drugo, iz $B(12, p)$, gdje je p nepoznati parametar.
- Grafički usporedite histograme podataka iz a) sa histogramima, prvo, $B(12, 1/3)$ binomne razdiobe, te, drugo, podacima odgovarajuće prilagođene binomne $B(12, p)$ binomne razdiobe.
- Odredite 95% pouzdani interval za nepoznati parametar p iz b) dijela zadatka, za oba uzorka (I.) i (II.).
- Za oba uzorka (I.) i (II.) posebno sprovedite test osnovne hipoteze $H_0 : p = \frac{1}{3}$ u odnosu na alternativu $H_1 : p \neq \frac{1}{3}$.

ZADATAK B

Pedesetih je godina prošlog stoljeća škotski fizičar James Forbes proučavao vezu između vrelišta vode i atmosferskog tlaka u svrhu razvoja jednostavne metode određivanja nadmorske visine planine. Podatci dobiveni mjerenjima, koje je tih godina izvršio na raznim lokacijama u Alpama i Škotskoj, nalaze se u datoteci *forbes.dat*. Kao potporu svojim podacima, naveo je i rezultate mjerenja koje je na Himalajima sproveo dr. Joseph Hooker u datoteci *hooker.dat*. U prvom stupcu datoteka je vrelište, a u drugom atmosferski tlak. (Forbes, J.D. (1857) Further experiments and remarks on the measurement of heights by the boiling point of the water, Transaction of the Royal Society of Edinburgh, 21, 135-143).

- a) Prikažite podatke. Kakvu vezu među podacima sugerira graf?
- b) Provedite prilagodbu linearnog modela $y = \theta_0 + \theta_1 x$, potom kvadratičnog modela $y = \theta_0 + \theta_1 x + \theta_2 x^2$ podacima i dobiveni pravac te parabolu grafički prikažite zajedno s empirijskim podacima.
- c) Nacrtajte graf reziduala, graf standardiziranih reziduala te provjerite dolaze li (standardizirani) reziduali iz jedinične normalne distribucije (upotrebom dva kriterija: normalnog vjerojatnosnog grafa, te Kolmogorov-Smirnovljevog testa).
- d) Izračunajte statistiku R^2 . Je li prilagodba linearnog, odnosno kvadratičnog modela podacima zadovoljavajuća? Koji model biste radije odabrali? Argumentirajte.
- e) Sprovedite test prihvatljivosti reduciranog (linearnog) u odnosu na prošireni (kvadratični) model. Procijenite 95% pouzdane intervale za parametre onog modela kojeg biste odabrali nakon provedenog testa.
- f) Za model koji ste odabrali u e) dijelu zadatka, na grafu s empirijskim podacima prikažite donje i gornje krivulje koje definiraju 95% pouzdane intervale, prvo, za srednju vrijednost od Y (uz dano x), te, drugo, za Y (uz dano x).

Napomena: Ukoliko imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na mail adresu vanessa.keranovic@fer.hr.