Treći kolokvij iz Statistike - 2008:05:29:09:00

Glavno Pravilo. Ako Dragana i Nenad u pomoćnim materijalima nađu bilo kakve dokaze, formulacije teorema, vlasnika takve opreme će odmah udaljiti i ocijeniti skor na trećem kolokviju ocjenom $s^+ = \min\{0, B_\leq\}$, gdje je $B_\leq$ broj dotad postignutih bodova. Mobiteli zabranjeni. Sve, osim formula, pisaćeg pribora i kalkulatora odložiti na prve klupe.

1. Teorem Pearsona (5%). Dokaz teorema za $r = 2$ (5%). Teorem Fishera, koji proširuje Pearsonov rezultat (5%).

2. Lema Neyman–Pearson (5%). Dokaz LNP (15%).

3. Markan peca. Komaraca k'o u priči, a velike k'o rode. Svako malo on se lupi po ramenu, leđima. Naravno, ne gledajući. Uglavnom razmazuje sopstvenu krvcu i proklinje se što se nije rodio k'o Tristan ilite Ahil. Jaran mu Pilip puši sjedeći na panju iza njega i bilježi broj ubijenih beštija po jednoj intervenciji. Usput se smije Markanu, jer balkanski humor se bazira na nesreći i muci drugoga. No, po Pilipovoj bilježnici Markanov osvetnički pohod je rezultirao sljedećim uzrokom:

† komarci	0	1	2	3	4
ovoliko puta	47	25	15	7	6

.

Ispitati pokoravaju li se † komarci eksponencijalnoj razdiobi $\mathcal{E}(\lambda)$ s pragom značajnosti 0.01 i 0.05? (20%)

4. Po sljedećoj tablici su zasijane površine u po 10 sēla regionā A i B pšenicom:

A	75	163	326	442	254	125	559	254	101	359
B	52	149	289	381	278	111	634	278	128	355

.

S $\alpha = 0.01$ testirati WMW testom hipotezu da je razdioba zasijanih površina jednaka, protiv alternativne. (15%)

5. Obilježje populacije $\xi \sim \mathcal{N}(m, 4)$. U uzorku opsega $n = 25$ je $\overline{X}_{25} = 14.7$. Pri $\alpha = 0.01$ testirati hipotezu

$$\mathsf{H}_0(m = 16) \longleftrightarrow \mathsf{H}_1(m < 16)\,.$$

(15%)

6. Uzorak opsega $n = 25$ dolazi iz populacije s obilježjem $\xi \sim \mathcal{N}(0, \sigma^2)$. Testirati hipotezu

$$\mathsf{H}_0\big(\sigma = 1\big) \longleftrightarrow \mathsf{H}_1\big(\sigma > 1\big)$$

nalazeći kritičnu oblast pomoću leme Neyman–Pearson s $\alpha = 0.05$. (15%)

7. (Nagradni zadatak = 20%). Novčić se baca 100 puta i registrira dobiveni rezultat.

1. Primjenom χ^2–testa testirati hipotezu da se grb i pismo podjednako često pojavljuju, ako smo registrirali 45 puta grb. Nivo značajnost neka je $\alpha = 0.01$.

2. Koji je minimalni broj pojavljivanja grba pri kojem uzorak ne proturiječi hipotezi u prethodnoj otčki zadatka?

SRETNO!

Tibor Poganj

RJEŠENJA

1. Bilježnica.

2. Bilježnica.

3. Budući da je parametar λ nepoznat, ocjenjujemo ga ML metodom $\widehat{\lambda} = \frac{1}{\overline{X}_n}$. Iz uzorka je $\overline{X}_{100} = (0 \cdot 3288 + 1 \cdot 25 + 2 \cdot 15 + 3 \cdot 7 + 4 \cdot 6)/100 = 1$, prema tome $\widehat{\lambda} = 1$. Teorijska razdioba je $\mathcal{E}(1)$, tj. gustoća i funkcija razdiobe su:

$$f(x) = e^{-x}\,\chi_{(x,\infty)}(x), \qquad F(x) = \left(1 - e^{-x}\right)\chi_{(x,\infty)}(x)\,.$$

Formiramo χ^2–tabelu:

S_k	$(-\infty, 1/2)$	$[1/2, 3/2)$	$[3/2, 5/2)$	$[5/2, 7/2)$	$[7/2, \infty)$
n_k	47	25	15	7	6
p_k	0.39	0.38	0.14	0.05	0.04
np_k	39	38	14	5	4
$(n_k - np_k)^2$	64	169	1	4	4
$\frac{(n_k-np_k)^2}{np_k}$	1.64	4.45	0.07	0.80	1.00

iz koje nalazimo vrijednost statistike $\chi^2_4 = 7.96$, dok je $\chi^2_{5-1-1;0.01} = \chi^2_{3;0.01} = 11.345$, $\chi^2_{3;0.05} = 7.815$. Kritična oblast je $\mathsf{C}_0 = [\chi^2_{3;\alpha}, \infty)$, pa krvavi matematički model *Markan–Pilip* na nivou 0.01 prihvaćamo, dok na grubljem 0.05 odbacujemo.

4. Formiramo WMW tablicu preko varijacijskih nizova

$A_{(k)}$	75	101	125	163	254	254	326	359	442	559
rang$(A_{(k)})$	3	3	5	8	9	10	14	16	18	19
$B_{(k)}$	52	111	128	149	278	278	289	355	381	634
rang$(B_{(k)})$	1	4	6	7	11	12	13	15	17	20

odakle jasno izlazi $T_A = 104$. Test–statistika $U_0 = T_A - 10 \cdot 11/2 = 104 - 55 = 49$, i kritična oblast je $\mathsf{C}_0 = \{X\colon |U| \ge c_\alpha\}$. Kako je $U \sim \mathcal{N}(50, 175)$, koristći CGT imamo vrijednost c_α. Naime,

$$\mathsf{P}\big\{|U| \ge c_{0.01}\big\} = \mathsf{P}\Big\{|U^*| \ge \frac{c_{0.01} - 50}{13.23}\Big\} = 1 - 2\Phi\Big(\frac{c_{0.01} - 50}{13.23}\Big) = 0.01\,,$$

odakle je $c_{0.01} = 84.13$. Kako $U_0 = 49 \notin \mathsf{C}_0$, hipotezu prihvaćamo. (Ne obaziremo se na iste vrijednosti unutar iste grupe podataka, rangovi se ne mijenjaju njihovim permutiranjem).

5. Test–statistika je

$$T = \frac{\overline{X}_n - m}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)\,,$$

dok je kritična oblast $\mathsf{C}_0 = (-\infty, z_{0.01}]$. Nalazimo granicu kritične oblasti:

$$\mathsf{P}\{T \le z_{0.01}\} = \Phi(z_{0.01}) - \Phi(-\infty) = 0.5 + \Phi(z_{0.01}) = 0.01$$

odakle je $z_{0.01} = -2.25$. Iz uzorka dobivamo

$$T_0 = \frac{14.7 - 16}{2} \cdot 5 = -3.25 \in \mathsf{C}_0\,.$$

Hipotezu odbacujemo.

6. Kada napišemo alternativnu hipotezu u obliku $\mathsf{H}_1(\sigma^2 = \sigma_1^2 > 1)$ imajući na umu da je $\sigma_0 = 1$, imamo već razrađenu proceduru za dobivanje kritične oblasti (bilježnica ili Pauše!):

$$\frac{L_{\sigma_0^2}(X)}{L_{\sigma_1^2}(X)} = \sigma_1^{25} \exp\Big\{ -(1 - \sigma_1^{-2}) \sum_{j=1}^{25} x_j^2 \Big\} \le C\,.$$

Dakle, kritična oblast je

$$\mathsf{C}_0 = \{X\colon \sigma_0^{-2} \sum_{j=1}^{25} X_j^2 \ge c_\alpha\} = \{X\colon \sum_{j=1}^{25} X_j^2 \ge c_\alpha\} = \{X\colon \chi_{25}^2 \ge c_\alpha\}\,.$$

gdje iz tablica za χ^2–razdiobu nalazimo da je $c_{0.05} = \chi^2_{25;0.05} = 37.652$. Zaključujemo:

$$\mathsf{C}_0 = [37.652, \infty)\,.$$

7. 1. Opseg uzroka je $n = 100$, broj stupnjeva slobode 1. Kako je broj grbova 45, pisama 55, vrijedi:

$$\chi_1^2 = \frac{(45-50)^2}{50} + \frac{(55-50)^2}{50} = 1\,.$$

Nadalje, zbog $\chi^2_{1;0.01} = 6.635$ nema razloga odbaciti hipotezu.

2. Tražimo najmanji takav prirodan broj N za koji je

$$\frac{(N-50)^2}{50} + \frac{(100-N-50)^2}{50} < 6.635\,.$$

Najmanje cjelobrojno rješenje ove nejednakosti je $N = 38$.