

6. feladatsor

Statisztika

2011. december 7.

- Egy $n = 10$ szervert tartalmazó kiszolgáló minden szervere minden pillanatban $0 < p < 1$ valószínűséggel foglalt, a foglaltságok szerverenként függetlenek. Tehát a foglaltak száma $Binom(n, p)$ eloszlásúnak tekinthető. A p -t szeretnénk meghatározni, ehhez 10 mérést végeztünk: 2,3,2,5,4,6,3,1,0,1.
A minta alapján határozzuk meg p legvalószínűbb értékét, azaz adjuk meg a p maximumlikelihood becslését általában $Binom(n, p)$ (fix n esetén), majd a konkrét példában.
- Adjuk meg az (a) exponenciális, (b) Poisson, (c) geometriai eloszlásból vett minták maximum likelihood becslését.
- Tegyük fel, hogy egy kérdőívvel a megkérdezettek jövedelmi viszonyait akarják felderíteni. A korábbi tapasztalatok szerint a magas jövedelműek 0,2 valószínűséggel alacsony jövedelműnek vallják magukat. Az alacsony jövedelműek csupán 0,1 valószínűséggel állítják, hogy ők a magas jövedelműek. Adjunk maximum likelihood becslést a tényleges θ arányra az alapján, hogy a beérkezett kérdőívek közül x szólt magas, $n - x$ pedig alacsony jövedelemről.
- A családok jövedelmét egy olyan skálán mérjük, ahol $X = 1$ a létminimumnak felel meg. Feltételezzük, hogy a jövedelem eloszlása az $f(x) = \frac{\theta}{x^{\theta+1}}$ ($x \geq 1$) sűrűségfüggvénnyel adható meg. (Ez az úgynevezett Pareto-eloszlás). Adjunk maximum likelihood becslést a θ -ra, ha 10 véletlenszerűen választott család jövedelme alapján: 1,53, 2,76, 19,65, 4,16, 7,31, 1,21, 254,2, 5,45, 1,12, 1,63.
- Egy alkatrészt élettartama exponenciális eloszlású θ/t várható értékkel ha t hőmérsékleten működtetjük. Tegyük fel, hogy az n megfigyelést a különböző t_1, \dots, t_n hőmérsékleten végeztük és x_1, \dots, x_n élettartamokat figyeltünk meg. Adjunk maximum likelihood becslést θ -ra.
- Egy város energiafogyasztása normális eloszlású ismeretlen μ várható értékkel és a korábbi tapasztalatok alapján ismert σ szórással. n napon keresztül végeztünk méréseket x_1, \dots, x_n eredménnyel, majd az $(n + 1)$ -dik naptól m napon keresztül át csak a város egyik kerületéből érkeztek adatok, ahol a fogyasztás várható értéke az egész város fogyasztásának a fele: y_1, \dots, y_m a kapott adatsor. Tételezzük fel, hogy a szórással itt is σ . Adjunk maximum likelihood becslést μ -re.
- Egy cukorgyárban a cukrot 1000 gramm névértékű zacskókba csomagolják. A gyártási technológiából eredően az egy zacskóba kerülő cukor szórása 50 gramm, a várható értéke azonban ismeretlen, jelölje m gramm. Megvizsgálunk 25 zacskót, és azt tapasztaljuk, hogy a bennük lévő cukor mennyisége átlagosan 986 gramm. Elfogadjuk-e 95%-os szinten az $m = 1000$ hipotézist az $m \neq 1000$ hipotézis ellenében? Mi lenne a helyzet, ha a szórással csak 20 gramm lenne?
- Egy oldat sótartalmát mérjük laborban. 5 mérést végzünk, melyek során a sótartalomra rendre a következő értékek adódtak (gramm/liter): 7.7, 8.1, 7.7, 7.5, 7.0. Az oldatról előzetesen azt állították, sótartalma 7 g/l. Elfogadjuk-e ezt 95%-os szinten azon hipotézis ellenében, hogy a sótartalom nem 7 g/l?
- Megvizsgálták, hogy 10 ember mekkora távot tudott 5 perc alatt lefutni. Ezután mindenki 3 napot diétázott, és így is megmérték a futásteljesítményt. Azt szeretnénk kideríteni, hogy a diéta befolyásolta-e a futásteljesítményt. Bizonyítható-e 95%-os szinten, hogy a diéta javított a teljesítményen?

Diéta előtt	1520	1830	1620	1740	1970	2130	1910	2000	1980	1900
Diéta után	1630	1810	1700	1800	1930	2100	1960	2160	2040	1970

- Egy gyárban azt vizsgálják, milyen módon lehetne növelni a munkások termelékenységét. Kétféle módot tesztelnek: (A) fizetésemelés, (B) munkakörülmények javítása. Két külön csoporton tesztelnek. Az alábbi két táblázat tartalmazza a termelékenység változását.

(A)						
munkás	1	2	3	4	5	6
változás	1.1	0.2	-0.1	2.2	1.3	1.3

(B)										
munkás	1	2	3	4	5	6	7	8	9	10
változás	1.3	2.5	1.2	0.8	0.3	1.9	3.2	2.4	2.2	3.2

- (a) Fogadjuk el vagy utasítjuk el 95%-os szinten azt a nullhipotézist, hogy a fizetésemelés nem változtatja a termelékenységet.
- (b) Fogadjuk el vagy utasítjuk el 95%-os szinten azt a nullhipotézist, hogy a munkakörülmények javítása nem változtatja a termelékenységet.
- (c) Fogadjuk el vagy utasítjuk el 95%-os szinten azt a nullhipotézist, hogy a munkakörülmények javítása nem növeli jobban a termelékenységet, mint a fizetésemelés.

11. Kétféle tápot tesztelnek, az egyiket 6 csirkén, a másikat pedig 8 (az előzőektől különböző) csirkén. A tesztelésnél azt vizsgálják, mekkora a testsúlynövekedés a tápszer nélküli állapothoz képest. Az eredmény:

A típusú táp						
csirke	1	2	3	4	5	6
növekedés	2.1	1.2	1.4	2.2	0.4	1.7

B típusú táp								
csirke	1	2	3	4	5	6	7	8
növekedés	1.7	2.2	1.1	1.8	2.5	0.9	1.6	1.7

Döntsük el kétmintás t-próba segítségével, hogy a két táp hatása egyformának tekinthető-e 95%-os szinten.

12. Amikor az embereket megkérdezik, hogy mekkora a tömegük, gyakran mondanak a valóságosnál kisebb értékeket. Szeretnénk eldönteni az alábbi adathalmazról, hogy igazi mérésből származik, vagy az emberek megkérdezéséből nyerték. Azt a tényt fogjuk használni, hogy mérés esetén az utolsó számjegyek eloszlásának egyenletesnek kell lennie a $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ halmazon. Döntsünk 0,95 szinten arról a hipotézisről, hogy mérésből származnak az adatok.

utolsó számjegy	0	1	2	3	4	5	6	7	8	9
mérések száma	35	4	4	3	4	24	2	4	8	2

- 1.) Határozzuk meg H_0 és H_1 hipotéziseket.
- 2.) Határozzuk meg a próbastatisztika értékét. Mekkora a szabadságfoka a próbastatisztikának?
- 3.) Határozzuk meg a 0,95 szignifikanciaszinthez tartozó kritikus értéket a χ^2 eloszlás táblázatából.
- 4.) Hasonlítsuk össze a kritikus értéket és a próbastatisztika értékét, majd döntsünk arról, hogy elvetjük-e H_0 -t, vagy nem.
13. Egy tóban háromféle hal él: amúr, makréla és ponty. Ottó bácsi, az öreg horgász azt súgja nekünk, hogy a tóban kétszer annyi a ponty, mint a makréla vagy az amúr. Kifogtunk 60 halat; döntsük el ez alapján 95%-os szinten, hallgathatunk-e Ottó bácsira.

amúr	makréla	ponty
11	14	35

14. Azt szeretnénk megtudni, hogy a bukósikak színe és a baleseti sérülések típusa között van-e összefüggés. Az utóbbi néhány év adatai alapján a következő táblázatot kaphatjuk:

	fekete	fehér	sárga/narancs
Kontroll (nem sérült)	491	377	31
Balesetes (sérült vagy meghalt)	213	112	8

$\varepsilon = 0,05$ elsőfajú hibavalószínűséggel döntsünk arról a hipotézisről, hogy a csoport (kontroll vagy balesetes) független a bukósikak színétől.

15. Egy csavar lehet hibás méret illetve szilárdság alapján is. Megvizsgáltunk 460 darab csavart.

	jó méret	rossz méret
jó szilárdság	416	16
rossz szilárdság	23	5

Döntsük el 95%-os szignifikanciaszinten, hogy az, hogy egy csavarnak megfelelő-e a szilárdsága illetve a mérete, független-e.