Az ismétlő órához nem kapcsolódnak gyakorlatok

Gyakorló feladatok megoldásai (2, ismétlés):

1.
    a Független mintás, kétmintás
    b Összetartozó mintás, páros
    c Egymintás
    d Független mintás, kétmintás

2. Khi-négyzet próba vagy a Kolmogorov-Smirnov próba

    Khi-négyzet próba
    a=read.table("S:/kata/fizetes.txt", sep=",", header=T)
    a=as.matrix(a)
    N=36, pl. length(a)
    2*36^(2/5)
    k≈8
    library(nortest)
    pearson.test(a, 8,adjust=T)
    p=0.55, azaz a változó normális eloszlású

    Kolmogorov-Smirnov próba
    ks.test(a, pnorm, mean(a), sd(a))
    p=0.995, azaz a változó normális eloszlású

Gyakorló feladatok megoldásai (3):

1. Paraméteres, vagy nemparaméteres: ennek eldöntéséhez normalitás-vizsgálatot kell végeznünk, azaz khi-négyzet próbát vagy a Kolmogorov-Smirnov próbát.

    Khi-négyzet próba
    a=read.table("S:/kata/IQ.txt", sep=",", header=F)
    a=as.matrix(a)
    N=132, pl. length(a)
    2*132^(2/5)
    k≈14
    library(nortest)
    pearson.test(a, 14,adjust=T)
    p<0.001, azaz a változó eloszlása különbözik a normális eloszlástól

    Kolmogorov-Smirnov próba
    ks.test(a, pnorm, mean(a), sd(a))
    p=0.01, azaz a változó nem normális eloszlású

    Az eredményeknek megfelelően csak nemparaméteres eljárást használhatunk annak eldöntéséhez, hogy a középérték lehet-e a populáció szintjén 100. A binomiális próbát vagy a Wald-Wolfowitz próbát használtahjuk. Kivitelezzük r-ben a binomiális próbát (ez egyszerűbb)!

    Binomiális próba a medián vizsgálatára
    library(car)

b=recode(a, “0:100=’0’; else=’1’”)
binom.test(sum(b), length(b), p=0.5)
p=0.46, azaz a minta nem mond ellent a populációban feltételezett 100-as IQ középértéknek.

Gyakorló feladatok megoldásai (4):

1. Kétmintás, független mintás t-próbának megfelelő nemparaméteres eljárások: a Kolmogorov-Smirnov próba, a Wald-Wolfowitz próba vagy a független mintás Wilcoxon, azaz a Mann-Whitney.

   Kivitelezzük a Kolmogorov-Smirnov próbát!
   a=read.table("S:\\kata\\suti.txt", sep=",", header=T)
   a=as.matrix(a)
   ks.test(a[,1], a[,2])
   p=0.4, azaz nem szignifikáns, a cserkészfiúk és cserkészlányok süti-eladásának mediánja nem tér el szignifikáns mértékben, a populáció szintjén nincs jelentős különbség a fiúk és a lányok teljesítménye közt a sütik eladása terén.

2. Kétmintás, független mintás t-próbának megfelelő nemparaméteres eljárások: a Kolmogorov-Smirnov próba, a Wald-Wolfowitz próba vagy a független mintás Wilcoxon, azaz a Mann-Whitney.

   Kivitelezzük a Mann-Whitney próbát (a független mintás Wilcoxon)!
   a=read.table("S:\\kata\\Shakespeare.txt", sep=",", header=F)
   a=as.matrix(a)
   wilcox.test(a[,1],a[,2], correct=F)
   p=0.053, azaz nem szignifikáns az eredmény, nincs szignifikáns eltérés a kérdő névmások mediánja közt az eredeti és az újonnan talált Shakespeare műben, azaz ezen karakterisztika alapján akár Shakespeare mű is lehet.

Gyakorló feladatok megoldásai (5):

1. Páros, összetartozó mintás t-próbának megfelelő nemparaméteres eljárás: a páros Wilcoxon.
   a=read.table("S:\\kata\\ertelmetlen.txt", sep=",", header=T)
   a=as.matrix(a)
   wilcox.test(a[,1],a[,2], paired=T, correct=F)
   p=0.019, szignifikáns, azaz a mediánok a populáció szintjén eltérnek. A negyedik és az ötödik próba során a vizsgálati személyek eltérő módon teljesítettek.

2. Páros, összetartozó mintás t-próbának megfelelő nemparaméteres eljárás: a páros Wilcoxon.
   a=read.table("S:\\kata\\reklam.txt", sep=",", header=F)
   a=as.matrix(a)
   wilcox.test(a[,1],a[,2], paired=T, correct=F)
   p=0.063, azaz nem szignifikáns, a mediánok a populáció szintjén a minta alapján lehetnek azonosak, a reklám nem volt hatásos, kb. annyian tértek be a bevásárlóközpontba előtte, mint utána.

Gyakorló feladatok megoldásai (6):

1. A független mintás variancia analízisnek megfelelő, Kruskal-Wallis próbát használhatjuk.
   a=read.table("S:\\kata\\Picasso.txt", sep=",", header=T)
   a=as.matrix(a)
   kruskal.test(a[,1]~a[,2])
   p<0.001, azaz szignifikáns az eredmény, van legalább két olyan csoport melyek közt jelentős eltérés van a tetszés tekintetében.
   Ez nem része a feladatnak, de páronkénti vizsgálattal (független mintás Wilcoxon) eldönthető lenne, hogy mely csoportok közt van különbség.
   wilcox.test(a[,1][a[,2]==1],a[,1][a[,2]==2], correct=F)
   wilcox.test(a[,1][a[,2]==1],a[,1][a[,2]==3], correct=F)
   wilcox.test(a[,1][a[,2]==2],a[,1][a[,2]==3], correct=F)
   Minden pár esetében p<0.001, azaz minden végzetségi osztály tetszés(közép)értéke szignifikánsan eltér egymástól.

2. A független mintás variancia analízisnek megfelelő, Kruskal-Wallis próbát használhatjuk.
   a=read.table("S:\\kata\\segit.txt", sep=",", header=T)
   a=as.matrix(a)
   kruskal.test(a[,1]~a[,2])
   p=0.004, azaz szignifikáns az eredmény, a különböző korú gyerekek csoportjai közt van legalább kettő, melyek mediánjuk alapján eltérő mértékben altruisták.
   Ez nem része a feladatnak, de páronkénti vizsgálattal (független mintás Wilcoxon) eldönthető lenne, hogy mely csoportok közt van különbség.
   wilcox.test(a[,1][a[,2]==1],a[,1][a[,2]==2], correct=F)
   p=0.009, azaz az egyes és a kettes csoport közt jelentős az eltérés.
   wilcox.test(a[,1][a[,2]==1],a[,1][a[,2]==3], correct=F)
   p=0.002, azaz az egyes és a hármas csoport közt jelentős az eltérés.
   wilcox.test(a[,1][a[,2]==2],a[,1][a[,2]==3], correct=F)
   p=0.36, azaz az egyes és a hármas csoport közt nem jelentős az eltérés.

3. A összetartozó mintás variancia analízisnek megfelelő, Friedman próbát használhatjuk.
   a=read.table("S:\\kata\\irogep.txt", sep=",", header=T)
   a=as.matrix(a)
   friedman.test(a)
   p=0.007, azaz szignifikáns az eredmény, van legalább két olyan írógéptípus, melyek eltérő teljesítményhez vezetnek.
   Ez nem része a feladatnak, de páronkénti vizsgálattal (páros Wilcoxon) eldönthető lenne, hogy mely csoportok közt van különbség.
   A kivitelezés módja más, lásd az adatbázis külalakját, egy típus egy oszlop!
   A boxplot(a[,1], a[,2], a[,3], a[,4], a[,5]) parancs is segíthet a jelentős különbségek felfedezésésben, amit aztán a páronkénti vizsgálattal igazolhatunk, számszerűsíthetünk.
   wilcox.test(a[,3],a[,4],paired=T, correct=F)
   p=0.031, azaz a C és D típusú írógépeken mutatott teljesítmény jelentősen különbözik egymástól.

Gyakorló feladatok megoldásai (8):

1.

$$\pi_p = 0.15, \omega_p = \frac{\pi_p}{\pi_{np}} = \frac{0.15}{0.85} = 0.18:1, \pi_{np} = 1 - \pi_p = 0.85, \omega_{np} = \frac{1}{\omega_p} = \frac{1}{0.18} = 5.56:1$$

2.

$$\omega_b = 0.03, \pi_b = \frac{\omega_b}{\omega_b + 1} = \frac{0.03}{1.03} = 0.029, \pi_{nb} = 1 - \pi_b = 1 - 0.029 = 0.971$$

3.Illeszkedésvizsgálat binomiális próbával. R parancs: binom.test(23,100,p=0.15). p=0.03, azaz szignifikáns az eltérés 0.05-ös szignifikanciaszinten. Így a konzervatív hipotézist elvetjük, az adataink ellentmondanak annak az állításnak, hogy általában a buszsofőrök 15%-a alkalmatlan a stressz-tűrést vizsgáló teszt alapján.

4.Illeszkedésvizsgálat binomiális próbával. R parancs: binom.test(81,200,p=0.34). p=0.06, azaz nem szignifikáns az eltérés 0.05-ös szignifikanciaszinten. Így, az adataink alapján a debreceni populáció a magyar populációnak megfelelő internetezési gyakoriságot mutat.

5.Igen, mert kísérletről van szó. Prospektív a vizsgálat.

$$\pi_{őrizet,előlőkésző} = \frac{19}{61} = 0.31, \pi_{őrizet,kontroll} = \frac{32}{62} = 0.51, \pi_{őrizet,kontroll} - \pi_{őrizet,előlőkészí} = 0.51 - 0.31$$

$$= 0.2, \omega_{őrizet,kontroll} = \frac{32}{30} = 1.07:1, \phi_{őrizet,kontr/előlőkészí} = (32/30)/(19/42) \approx 2.36$$

Gyakorló feladatok megoldásai (9):

1. N>20, n>5, minden cellánál, ezért χ^2 -próbát lehet használni.
   Függetlenségvizsgálatról van szó. R parancs pl.:
   a=matrix(c(114,157,158,255), ncol=2, byrow=T)
   chisq.test(a, correct=F)
   p=0.32, tehát 0.05-ös szignifikancia szinten nem mutatkozik szignifikáns különbség mutatkozik a két csoportot tekintetében a szálláshelyre vonatkozóan. Azaz nem függ a nemtől a szálláshelyválasztás. A szabadságfok (2-1)(2-1)=1.

2.N>20, n>5, minden cellánál, ezért χ^2 -próbát lehet használni.
   a=matrix(c(23,15,10,6,12,30), ncol=3, byrow=T)
   chisq.test(a, correct=F)
   p=0.0000391
   Függ a szülő nemétől a kommunikáció módja. Függetlenségvizsgálatról van szó. A szabadságfok (3-1)(2-1)=2.

3. Függetlenségvizsgálat esetében azt vizsgáljuk, hogy a függő változó eloszlására más-e a független változó különböző értékei esetén. Kontingencia táblázatban gondolkodva ez azt jelenti, hogy megvizsgáljuk, hogy az oszlopokban lévő gyakoriságok függnek-e a sorok gyakoriságaitól. (Csak multinomiális vagy Poisson mintavétel esetén tehető ez meg.)

4. Homogenitásvizsgálat esetében azt vizsgáljuk, hogy a két változó eloszlására függ-e egymástól. Kontingencia táblázatban gondolkodva ez azt jelenti, hogy az oszlopokban és sorokban lévő gyakoriságok függnek egymástól. (Minden típusú mintavétel esetén megtehető).

5. A változó eloszlása megfelel a feltételezett eloszlásnak.

6. A két változó eloszlása függ egymástól.

7. Retrospektív vizsgálat esetén a mintavétel a függő változó különböző értékei mentén történik.

8. Lehet.

9. $$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{43}{0.43 + 89}} = 0.069$$

   Gyenge, majdnem elhanyagolható kapcsolat.

10. $$V = \sqrt{\frac{\chi^2}{N(k-1)}} = \sqrt{\frac{10.4}{42(3-1)}} \approx 0.35$$

    Biztos, de gyenge kapcsolat

<u>Gyakorló feladatok megoldásai (10):</u>

1. Van olyan cella, amely kevesebb megfigyelést tartalmaz, mint öt, így a χ^2 -próbát nem lehet használni, a Fisher-teszt kivitelezhető. R parancs:
   a=matrix(c(3,2,1,4), ncol=2, byrow=T)
   fisher.test(a, a="greater")
   p=0.26, tehát 0.05-ös szignifikancia szinten nem mutatkozik szignifikáns különbség az abortuszt elfogadók és elutasítók körében az eutanáziára vonatkozó attitűd tekintetében. Homogenitásvizsgálatról van szó.

2. Közepes egybehangzóság

3. Erős kapcsolatot

4. Gyenge kapcsolatot

5. Kiváló egybehangzóság

6. Gyenge egybehangzóság

7. Bevisszük az adatokat adat nevű mátrixba R-ben.
   Pl. így adat=matrix(c(45,5,6,10,70,3,7,5,56), ncol=3, byrow=T)
   library(vcd)
   Kappa(adat)
   Eredmény: (unweighted value): kb. Kappa=0.74

Jó egybehangzóságot mutat a két teszt.

8. Először ki kell számolnunk a marginálisokat!

Ruha /Bögre szín	Piros	Sárga	Kék	Zöld	Összes
Piros	15	10	8	5	5
Sárga	14	12	15	3	3
Kék	12	5	8	11	11
Zöld	11	9	7	6	6
Összes	52	36	38	25	151

Ha nem vennénk figyelembe a független változót (ruha színe), akkor mindig piros bögrét küldenénk. Ekkor 52 esetben döntenénk optimálisan. Ha figyelembe vennénk a vásárolt ruha színét, akkor piros bögrét küldenénk, ha a személy leggyakrabban piros, kék vagy zöld színű ruhát vásárol; és kék bögrét küldenénk, ha leggyakrabban sárga színű ruhát vásárol. Ekkor 15+12+11+15 esetben döntenénk optimálisan. A lambda értéke ebből következően:

$$\lambda = \frac{\sum_i O_{im} - O_{+m}}{N - O_{+m}} = \frac{15+15+12+11-52}{151-52} = \frac{1}{99} \approx 0.01$$

Ez alapján gyenge, majdnem elhanyagolható kapcsolatról van szó.

Gyakorló feladatok megoldásai (11):

1. Konkordáns 5: AG, BE, BG, DF, DG
   Diszkordáns 5: AC, BC, BD, CE, EF
   x-ben kapcsolt 6: AD, AE, CF, CG, DE, FG
   y-ban kapcsolt 5:AB, AF, BF, CD, EG

2. Minden ordinális skálatípusú adatra fejlesztett asszociációs mutató értéke nulla lesz, mert P-Q=5-5=0

3. N kiszámolható a táblázatból, értéke 277
   A két változó értékeinek száma közül a kisebbik m=3
   Az eredmények, ha el nem számoltam:

$$\Gamma = \frac{P-Q}{P+Q} = \frac{4814-14733}{4814+14733} = \frac{-9919}{19547} \approx -0.51$$

$$D_{x|y} = \frac{P-Q}{P+Q+T_x} = \frac{4814-14733}{4814+14733+13112} = \frac{-9919}{32659} \approx -0.30$$

$$D_{y|x} = \frac{P-Q}{P+Q+T_y} = \frac{4814-14733}{4814+14733+9152} = \frac{-9919}{28699} \approx -0.35$$

$$D_{sym} = \frac{P-Q}{P+Q+\frac{T_y+T_x}{2}} = \frac{4814-14733}{4814+14733+\frac{13112+9152}{2}} = \frac{-9919}{30679} \approx -0.32$$

$$\tau = \frac{2(P-Q)}{N(N-1)} = \frac{2(4814-14733)}{277(277-1)} = \frac{-19838}{76452} \approx -0.26$$

$$\tau_b = \sqrt{D_{x|y} D_{y|x}} = \sqrt{-.03 * -0.35} \approx -0.33$$

$$\tau_c = \frac{2k(P-Q)}{N^2(k-1)} = \frac{6(4814-14733)}{277^2 * 2} = \frac{-59514}{153458} \approx -0.39$$

4. A rangok:
   2db 1-es van, rangjuk 1.5
   3db 2-es van, rangjuk 4
   2db 3-as van, rangjuk 6.5
   2db 4-es van, rangjuk 8.5
   5db 5-ös van, rangjuk 12
   1db 6-os van, rangja 15
   3db 7-es van, rangjuk 17

   Így az értékek rangokká konvertálva:
   8.5,12,4,17,1.5,12,6.5,17,15,12,12,8.5,4,1.5,17,6.5,4,12

5. Spearman féle rangkorreláció.
   a=c(5,1,3,7,8,6,4,2)
   b=c(4,7,6,1,3,2,5,8)
   cor.test(a,b, method="spearman")
   r=-09
   Tehát, erős fordított arányosság van a kopaszodás mértéke és a vonzóság megítélése között. Minél kopaszabb valaki, annál kevésbé vonzónak ítélik meg.

6. Spearman féle rangkorreláció.
   a=c(2,4,1,5,3,6,8,10,9,7)
   b=c(6,8,7,5,5,4,2,0,0,0)
   cor.test(a,b, method="spearman")
   r=-093
   Tehát, erős fordított arányosság van a tanártól való távolság és a teljesítmény között. Minél távolabb ül valaki, annál rosszabb a teljesítménye. Mivel ez megfigyelés., nem feltételezhetjük, hogy ha közelebb ülne valaki, rögtön javulna a teljesítménye. Lehet az ok egy látens mögöttes változó, mint pl. a motiváció mértéke.