Rešenja uradio L. Tenjović, a naknadno izmenila Nevena Mijatović Zadatak 4.

Zadatak 1.

Na tri pitanja ponuđeni su sledeći odgovori:

1. pitanje: DA NE

2. pitanje: DA ? NE

3. pitanje: 1 2 3 4 5

- Na koliko različitih načina je moguće odgovoriti na ova tri pitanja?

Odgovor:

U programu SPSS za sva računanja na ovim vežbama koristimo komandu COMPUTE.

Otvoriti bilo koji fajl sa podacima (konkretni podaci nisu važni, fajl se otvara da bi mogla da se koristi komanda COMPUTE). Umesto otvaranja postojećeg fajla u novom fajlu se mora uneti barem jedan podatak u matricu podataka kako bi mogla da se koristi komanda COMPUTE.

Prema osnovnom pravilu kombinatorike $m_1 * m_2 * m_3 = 2 * 3 * 5 = 30$.

Zadatak 2.

Fajl isti kao za zadatak 1.

- Ako se iz populacije koja ima 20 članova slučajno bira (uzorkovanjem **bez vraćanja**) 5 jedinica u uzorak koliko različitih uzoraka veličine 5 je moguće izvući?

Odgovor:

a) *Ukoliko uzorak definišemo kao neuređen skup (što je uobičajeno u empirijskim istraživanjima) a pošto je reč o uzorkovanju bez vraćanja, tražimo broj kombinacija bez ponavljanja sa po 5 elemenata od ukupno 20 elemenata:*

n = 20
r=5

$$nCr = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{20*19*18*17*16}{5!} = \frac{1860480}{120} = 15504$$

b) *Ukoliko, pak, uzorak definišemo kao uređen skup (što je neuobičajeno u empirijskim istraživanjima) a pošto je reč o* uzorkovanju bez vraćanja *tražimo broj* varijacija *bez ponavljanja sa po 5 elemenata od ukupno 20 elemenata*:

nPr = n! / (n-r)! = n (n-1)…(n - r + 1) = 20!/(20-5)! = 20*19*18*17*16 = 1860480.

Zadatak 3.

Fajl isti kao za zadatak 1.

a) Kolika je verovatnoća da se u 3 nezavisna bacanja ispravnog novčića sva tri puta dobije "pismo"? b) Kolike su šanse tog događaja?

Odgovor:

a) Sva tri puta "pismo" u 3 nezavisna bacanja predstavlja "povoljan" ishod i takav ishod može biti samo jedan. To je ishod PPP ("pismo" u prvom, "pismo" u drugom i "pismo" u trecem bacanju).

Ukupan broj mogućih ishoda možemo dobiti ispisivanjem mogućih ishoda:

S = {GGG, GPP, GGP, GPG, PGG, PPG, PGP, PPP}

Ukupan broj mogucih ishoda možemo dobiti i na osnovu Osnovnog pravila kombinatorike:

*U svakom bacanju su dva moguća ishoda, a ima 3 nezavisna bacanja. Prema tome n = 2 * 2 * 2 = 8.*

Možemo primeniti i obrazac za broj varijacija sa ponavljanjem od 2 elementa od ukupno 3 elementa:

$$nPr = n^r = {}_3P_2 = 2^3 = 8.$$

Budući da je ukupan broj ishoda jednak 8, a da je samo 1 ishod "povoljan", verovatnoću događaja PPP u 3 bacanja novčića računamo po obrascu koji verovatnoću definiše na klasičan način:

$$P(A) = \frac{n_A}{n} = \frac{1}{8} = 0.125$$

Alternativno: 3 nezavisna bacanja su u pitanju, verovatnoća pisma u svakom je 1/2. "Pismo" u prvom, "pismo" u drugom i "pismo" u trećem bacanju su nezavisni događaji, pa je verovatnoća njihovog zajedničkog javljanja jednaka proizvodu pojedinačnih verovatnoća ova tri događaja: (1/2) (1/2)* (1/2) = 1/8 = 0.125.*

b) Šanse događaja PPP računamo po obrascu:

$$O(A) = \frac{P(A)}{1 - P(A)} = \frac{0.125}{1 - 0.125} = 0.143$$

Uočimo da su šanse događaja PPP, budući da je verovatnoća tog događaja veoma niska, veoma bliske verovatnoći tog događaja.

Zadatak 4 .

Fajl isti kao za zadatak 1.

U uzorku PISA istraživanja u Srbiji sa 4879 đaka, nalazi se 2902 đaka koji nemaju visoko obrazovane roditelje i 1977 đaka kojima je bar jedan roditelj visoko obrazovan. Među đacima koji nemaju visoko obrazovane roditelje je 1625 postiglo funkcionalnu pismenost, a među đacima koji imaju bar jednog visoko obrazovanog roditelja je njih 1387 postiglo funkcionalnu pismenost.

a) Kolika je verovatnoća da na slučaj odabran đak iz ovog uzorka bude funkcionalno pismen?
b) Kolika je verovatnoća da na slučaj odabran đak iz uzorka bude funkcionalno pismen ako znamo da je reč o đaku sa bar jednim visoko obrazovanim roditeljem?
c) Kolika je verovatnoća da na slučaj odabran đak iz uzorka bude funkcionalno nepismen ako znamo da je reč o đaku bez visoko obrazovanih roditelja?
d) Da li su (kada imamo na umu ovih 4879 đaka) "imati bar jednog visoko obrazovanog roditelja" i "biti funkcionalno pismen" statistički nezavisni događaji?
e) Kolike su šanse postizanja funkcionalne pismenosti u ovom uzorku?
f) Kolike su šanse postizanja funkcionalne pismenosti među đacima kojima je bar jedan roditelj visoko obrazovan?
g) Kolike su šanse funkcionalne nepismenosti među đacima bez visoko obrazovanih roditelja?

<u>Odgovor:</u>

Napravićemo sledeću tabelu:

	Funcionalno nepismeni	Funkcionalno pismeni	Ukupno
Bez visoko obrazovanih roditelja	1525	1387	2902
Sa bar jednim visoko obrazovanim roditeljem	590	1625	1977
Ukupno	*1867*	*3012*	4879

a) Verovatnoća da slučajno odabrani đak bude funkcionalno pismen računa se kao:

P(Funkcionalno pismen) = Ukupan broj funkcionalno pismenih / Ukupan broj đaka = 3012 / 4879 ≈ 0.6177

Dakle, verovatnoća je 61.77%.

b) Verovatnoća da đak bude funkcionalno pismen ako ima bar jednog visoko obrazovanog roditelja je uslovna verovatnoća P(Funkcionalno pismen|Visoko obrazovani roditelj) I računa se kao:

P(Funkcionalno pismen|Visoko obrazovani roditelj) =

Broj funkcionalno pismenih đaka sa visoko obrazovanim roditeljima / Ukupan broj đaka sa visoko obrazovanim roditeljima=1387 / 1977 ≈ 0.7014

Verovatnoća je 70.14%.

c) Verovatnoća da đak bude funkcionalno nepismen ako nema visoko obrazovane roditelje računa se kao uslovna verovatnoća P(Funkcionalno nepismen|Bez visoko obrazovanih roditelja):

P(Funkcionalno nepismen|Bez visoko obrazovanih roditelja)=

Broj funkcionalno nepismenih đaka bez visoko obrazovanih roditelja / Ukupan broj đaka bez visoko obrazovanih roditelja = 1277/2902 ≈0.4399

Verovatnoća je 43.99%.

d) Da bismo izračunali statističku nezavisnost događaja “Visoko obrazovani roditelj” i “Funkcionalno pismen đak” možemo posmatrati više brojeva iz tabele učestalosti, tj. doći do istog zaključka korišćenjem više postupaka.

Prvi postupak bi bilo poređenje totalne verovatnoće događaja funkcionalne pismenosti (61.77%) i uslovne verovatnoće funkcionalne pismenosti đaka sa bar jednim visoko obrazovanim roditeljem (70.14%). Pošto se verovatnoće razlikuju, “Bar jedan visoko obrazovani roditelj” i “Funkcionalno pismen đak” nisu statistički nezavisni događaji.

Alternativno: Oslonićemo se i na sledeći postupak. Proverićemo prvo na osnovu verovatnoće ova dva događaja događaj šta je statistički očekivano da se nađe u preseku. Dakle posmatramo prvo samo pojedinačne raspodele za svaku varijablu posebno (u gornjoj tabeli to su kolone “ukupno”).

Verovatnoća događaja “Imati bar jednog visoko obrazovanog roditelja” je 1977/4879 ≈ 0.4025. Verovatnoća događaja “biti funkcionalno pismen” je 3012 / 4879 ≈ 0.6177. Proizvod pojedinačnih verovatnoća, odnosno presek ta dva računa se kao 0. 4025 *0.6177 = 0.2503. Na osnovu raspodele pojedinačnih varijabli, statistički očekivano je da u njihovom preseku u kategoriji funkcionalno pismenih đaka sa visoko obrazovanim roditeljima bude ¼ đaka.

Međutim, kada podatke za dve varijable zaista ukrstimo (“cross-tabs”), vidimo da je u kategoriji funkcionalno pismenih đaka sa visoko obrazovanim roditeljima 1625/4879 đaka, što je 0.3331, odnosno 1/3, trećina đaka. Budući da očekivana verovatnoća od 0.25 nije ista kao podacima dobijena verovatnoća od 0.33, u pitanju je sistematsko zakrivljenje, odnosno statistički su zavisni događaji. Tehnički izrečeno: verovatnoća stvarnog zajedničkog dešavanja ova dva događaja (0.3331) nije jednaka proizvodu pojedinačnih verovatnoća ta dva događaja (0.2503*).*

e) Šanse za postizanje funkcionalne pismenosti

Šanse se računaju kao:

O(Funkcionalno pismen) = P(Funkcionalno pismen) / P(Funkcionalno nepismen) = 0.6177 / (1 - 0.6177) ≈ 1.613

Šanse su 1.613.

e) Šanse za funkcionalnu pismenost među đacima sa bar jednim visoko obrazovanim roditeljem. Primetimo da je ovo takozvana uslovna šansa.

O(Funkcionalno pismen | Visoko obrazovani roditelj) = P(Funkcionalno pismen | Visoko obrazovani roditelj) / (1 - P(Funkcionalno pismen | Visoko obrazovani roditelj)) = 0.7014 / (1 - 0.7014) ≈ 2.348

Šanse su 2.348

f) Šanse za funkcionalnu nepismenost među đacima bez visoko obrazovanih roditelja. Primetimo da je i ovo takozvana uslovna šansa.

O(Funkcionalno nepismen | Bez visoko obrazovanih roditelja) = P(Funkcionalno nepismen | Bez visoko obrazovanih roditelja) / (1 - P(Funkcionalno nepismen | Bez visoko obrazovanih roditelja)) = 0.4399 / (1 - 0.4399) ≈ 0.785

Šanse su 0.785.

Zadatak 5.

Fajl isti kao za zadatak 1.

Test pomoću detektora laži primenjuje se na populaciji na kojoj ogromna većina ispitanika nema razloga da laže u vezi sa određenim pitanjem, tako da je P(Istina)=0.99, P(Laž)=0.01, pri čemu je P(Istina) verovatnoća da osoba govori istinu a P(Laž) verovatnoća da osoba laže. Na osnovu ispitivanja pouzdanosti detektora laži zna se da je:

P(+| Laž) = 0.88 (Verovatnoća da se testom otkrije da osoba laže).

(Ova verovatnoća se kod dijagnostičkih testova uobičajeno zove senzitivnošću testa).

P(-| Laž) = 0.12

P(-| Istina) = 0.86 (Verovatnoća da se testom potvrdi da osoba govori istinu)

(Ova verovatnoća se kod dijagnostičkih testova uobičajeno zove specifičnošću testa).

P(+| Istina) = 0.14

(Pozitivan ishod testa, kada detektor signalizira da osoba laže, označimo oznakom +, a negativan ishod testa, kada detektor ne signalizira da osoba laže, oznakom -).

Recimo da je za neku osobu u takvom ispitivanju test pozitivan, tj. da detektor laži signalizira da osoba laže. Izračunati (korišćenjem Bajesovog pravila) kolika je verovatnoća da je ishod testa pogrešan, tj. da osoba u stvari govori istinu? Drugim rečima kolika je verovatnoća P(Istina | +), tj. kolika je verovatnoća da osoba govori istinu ako znamo da je test pozitivan?

Odgovor:

Primenom oblika Laplas-Bajesove teoreme/pravila (za situaciju kada se skup mogućih ishoda S može predstaviti unijom dva međusobno isključiva događaja, B i B^c):

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(B)\mathrm{P}(A|B)}{\mathrm{P}(B)\mathrm{P}(A|B) + \mathrm{P}(B^c)\mathrm{P}(A|B^c)}$$

na ovu situaciju dobijamo:

$$\mathrm{P}(I|+) = \frac{\mathrm{P(I)P}(+|I)}{\mathrm{P}(I)\mathrm{P}(+|I) + \mathrm{P}(L)\mathrm{P}(+|L)} = \frac{0.99 * 0.14}{0.99 * 0.14 + 0.01 * 0.88} = 0.94$$

Dakle, u trijažnom ispitivanju detektorom laži u ovakvoj populaciji, tj. u populaciji u kojoj su uglavnom osobe koje su "nevine", tj. nemaju razloga da lažu, postojalo bi 94% pogrešnih alarma, tj. detektor laži bi pretežno pogrešno alarmirao da osobe koje govore istinu lažu.

Zadatak 6.

Fajl isti kao za zadatak 1.

U jednom časopisu objavljen je tekst u kojem stoji da *istraživanja seksualnih partnera osoba koje su zaražene virusom HIV pokazuju da je rizik inficiranja neinficiranog partnera u jednom seksualnom (vaginalnom) odnosu bez zaštite 1/500.* Novinar, zatim rezonuje ovako: *statistički. 500 seksualnih odnosa sa inficiranim partnerom daju 100% verovatnoću inficiranja ("statistički, ne nužno u realnosti", kaže novinar).*

- Izračunati kolika je verovatnoća inficiranja virusom HIV pod pretpostavkom da su prenošenja virusa u prvom, drugom, … 500-tom seksualnom (vaginalnom) odnosu bez zaštite **uzajamno statistički nezavisni** događaji.

Odgovor: Najjednostavnije je da ovu verovatnoću izračunamo tako što sračunamo, pod istim pretpostavkama, verovatnoću komplementarnog događaja ("neinficiranje virusom HIV u 500 vaginalnih odnosa bez zaštite"):

P(neinficiranje virusom HIV u jednom vaginalnom odnosu bez zaštite) = (1-1/500) = 0.998;

Budući da su prenošenja virusa u prvom, drugom, … 500-tom vaginalnom odnosu bez zaštite ***uzajamno statistički nezavisni*** *događaji, primenili bismo uopštenje teoreme o zajedničkoj verovatnoći dva statistički nezavisna događaja:*

Ako su $A_1, A_2, \ldots, A_k$ *uzajamno statistički nezavisni događaji tada je njihova zajednička verovatnoća (verovatnoća njihovog zajedničkog dešavanja), u oznaci* $P(A_1 \cap A_2 \cap \ldots \cap A_k)$*, jednaka proizvodu pojedinačnih verovatnoća tih događaja:*

$$\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \prod_{i=1}^{k} A_i$$

Prema tome:

P(neinficiranje virusom HIV u 500 vaginalnih odnosa bez zaštite) = $0.998^{500} = 0.3675$.

Odatle, na osnovu teoreme o komplementu:

P(inficiranje virusom HIV u 500 vaginalnih odnosa bez zaštite) = 1 – 0.3675 = 0.63.

Naravoučenije: iako verovatnoća nije jednaka 1 treba pri seksualnom odnosu koristiti zaštitu, tj. kondom!