Rešenja uradio L. Tenjović

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:*

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

b) *Ukoliko, pak, uzorak definišemo kao* <u>*uređen*</u> *skup (što inače nije uobičajeno u empirijskim istraživanjima) a pošto je reč o* uzorkovanju <u>bez vraćanja</u> *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?

<u>Odgovor:</u>

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 populaciji od 1 000 odraslih osoba, ima 400 muškaraca i 600 žena. Među muškarcima je njih 20 obolelo od depresije, a među ženama je njih 60 obolelo od depresije.

a) Kolika je verovatnoća da na slučaj odabrana osoba iz ove populacije boluje od depresije?
b) Kolika je verovatnoća da na slučaj odabrana osoba iz ove populacije ima depresiju ako znamo da je reč o osobi ženskog pola?
c) Kolika je verovatnoća da na slučaj odabrana osoba iz ove populacije <u>nema</u> depresiju ako znamo da je reč o osobi muškog pola?
d) Da li su (kada imamo na umu ovu populaciju) "biti žena" i "imati depresiju" statistički nezavisni događaji?
e) Kolike su šanse oboljevanja od depresije u ovoj populaciji?

f) Kolike su šanse oboljevanja od depresije među ženama u ovoj populaciji?
g) Kolike su šanse neoboljevanja od depresije među muškarcima u ovoj populaciji?

<u>Odgovor:</u>

Napravićemo sledeću tabelu:

	Ima depresiju	Nema depresiju	Ukupno
Muškarci	20	380	400
Žene	60	540	600
Ukupno	80	920	1000

a) *Verovatnoću da na slučaj izabrana osoba iz ove populacije ima depresiju određujemo kao relativnu frekvenciju ukupnog broja obolelih (zbir kolone "Ima depresiju) i veličine populacije:*

Broj obolelih/n = 80/1000 = 0.08.

b) *Verovatnoću da na slučaj izabrana osoba iz ove populacije ima depresiju ako znamo da je reč o osobi ženskog pola određujemo kao relativnu frekvenciju broja obolelih žena i ukupnog broja žena ($n_ž$ u imeniocu):*

Broj obolelih žena/ $n_ž$ = 60/600 = 0.1.

Uočimo da je to isto kao da smo računali uslovnu verovatnoću po obrascu preko verovatnoća:

P(depresija|žene) = P(imati depresiju $\cap$ biti žena)/P(biti žena) = (60/1000)/(600/1000) = 60/600 = 0.1

c) *Verovatnoću da na slučaj odabrana osoba iz ove populacije <u>nema</u> depresiju ako znamo da je reč o osobi muškog pola određujemo kao relativnu frekvenciju muškaraca koji nisu oboleli od depresije i ukupnog broja muškaraca:*

Broj muškaraca koji nisu oboleli od depresije/ n_m = 380/400 = 0.95.

d) *Kada imamo na umu ovu populaciju, verovatnoća događaja "biti žena" je 600/1000 = 0.6 a verovatnoća događaja "imati depresiju" je 80/1000 = 0.08. Verovatnoća zajedničkog dešavanja ova dva događaja, tj. verovatnoća događaja ("biti žena" $\cap$ "imati depresiju") je 60/1000 = 0.06. Budući da 0.6*0.08 = 0.048, te da 0.048 $\neq$ 0.06, ne možemo reći da su "biti žena" i "imati depresiju" statistički nezavisni događaji jer verovatnoća zajedničkog dešavanja ova dva događaja (0.06) nije jednaka proizvodu pojedinačnih verovatnoća ta dva događaja (0.048).*

e) *O(oboljevanja od depresije) = P(oboljevanja od depresije)/P(neoboljevanja od depresije) = 0.08/0.92 = 0.09.*

Ako šanse računamo preko frekvencija tada je O(oboljevanja od depresije) = (80/1000)/(920/1000) = 80/920 = 0.09.

Uočimo razliku između verovatnoće (broj obolelih/ukupan broj ljudi) i šansi (broj obolelih/broj neobolelih).

f) *O(oboljevanja od depresije$\mid$ žene) = P(oboljevanja od depresije$\mid$ žene)/P(neoboljevanja od depresije$\mid$ žene) = 0.1/0.9 = 0.11.*

Uočimo da je reč o uslovnim šansama te za njihovo računanje preko verovatnoća koristimo odgovarajuće uslovne verovatnoće.

Ako šanse računamo preko frekvencija tada je O(oboljevanja od depresije$\mid$ žene) = (60/600)/(540/600) = 60/540 = 0.11.

Uočimo razliku uslovne verovatnoće (broj obolelih žena/ukupan broj žena) i uslovnih šansi (broj obolelih žena/broj neobolelih žena).

g) O(neoboljevanja od depresije| muškarci) = P(neoboljevanja od depresije| muškarci)/P(oboljevanja od depresije| muškarci) = 0.95/0.05 = 19.

Uočimo da je reč o uslovnim šansama te za njihovo računanje preko verovatnoća koristimo odgovarajuće uslovne verovatnoće.

Ako šanse računamo preko frekvencija tada je O(neoboljevanja od depresije| muškarci) = (380/400)/(20/400) = 380/20 = 19.

Uočimo razliku uslovne verovatnoće (broj muškaraca koji nisu oboleli/ukupan broj muškaraca) i uslovnih šansi (broj neobolelih muškaraca/broj obolelih muškaraca).

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?

<u>Odgovor:</u>

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!