

## ZADAĆA 2.

Rok za predaju (uploadati na Merlin): 14.4.2022.

1. Rješite primjer sa slide-a 18, predavanje 24.3. (generiranje dobi žena u RH, popis 2001).
2. U trgovini (gdje radi samo 1 prodavač), 25% kupaca kupuje tako da prilikom kupnje ne traži informacije o proizvodima. Vrijeme kupnje za tu skupinu kupaca slijedi eksponencijalnu distribuciju sa sredinom 2.5 (min). Preostalih 75% kupaca, prije kupnje traži informacije o proizvodima od prodavača, pa vrijeme kupnje za tu skupinu slijedi Erlangovu distribuciju sa sredinom 4 (min). Generirajte (generatorima slučajnih brojeva) vremena potrebna za kupovinu za 100 kupaca, te procijenite vrijednosti prva 4 momenta (mean, std, skewness, kurtosis) vremena kupovine (svih kupaca), te ukupno vrijeme (sum) kupovine svih kupaca. Koristite generatore slučajnih brojeva za Bernoullijevu, eksponencijalnu i Erlangovu distribuciju. Koristite vrijednost 7736 za početni SEED (call streaminit(seed);)
3. U SAS datoteci RAZRED su upisani podaci za visinu (Height) u inčima i težinu (Weight) u funtama, za 18 učenika i učenica.

Podaci se unose u SAS datoteku RAZRED izvođenjem slijedećeg data stepa:

```
data RAZRED;
  input Name $ Height Weight @@;
  datalines;
Alfred 81.0 130.5   Alice 56.5 84.0   Barbara 65.3 98.0
Carol 62.8 102.5   Henry 61.5 101.5   James 76.3 170.0
Jane 71.8 94.5    Janet 62.5 112.5   Jeffrey 61.5 85.0
John 59.0 99.5    Joyce 65.3 150.5   Judy 64.3 90.0
Louise 56.3 77.0   Mary 66.5 115.0   Philip 72.5 152.0
Robert 64.8 158.0  Ronald 67.0 137.0  Thomas 57.5 85.0
;
run;
```

- a) Pretvorite funte u kilograme, a inče u cm i nazovite nove varijable VISINA i TEZINA.
- b) Procijenite vrijednosti prva 4 momenta (mean, std, skewness, kurtosis) za VISINA i TEZINA na danim podacima.
- c) Generirajte 18 slučajnih brojeva koji slijede nenormalnu distribuciju sa vrijednostima prva 4 momenta procijenjenima u a) za VISINA. Procijenite vrijednosti prva 4 momenta za generirane brojeve i ispišite ih. (Koristite vrijednost seed-a: 6678)
- d) Generirajte 18 slučajnih brojeva koji slijede nenormalnu distribuciju sa vrijednostima prva 4 momenta procijenjenima u a) za TEZINA. Procijenite vrijednosti prva 4 momenta za generirane brojeve i ispišite ih. (Koristite vrijednost seed-a: 2211)
- e) Ponovite b)-c) još 4 puta, sa različitim početnim vrijednostima (tj vrijednostima seed-a: 33456, 4544, 4236, 56643). Izračunajte prosječne vrijednosti 4 momenta procijenjenih za VISINA I TEZINA (u 5 ponavljanja/iteracija)
- f) Kako bi trebalo promijeniti program (data step za generiranje slučajnih brojeva) ako bismo željeli ponoviti b)-c) 100 puta? (Samo navesti kod, bez izvođenja.)

UPUTA:

Za ispis i spremanje vrijednosti mean, std, skewness i kurtosis u izlaznu SAS datoteku IMEIZLAZNEDATOTEKE koristite proceduru MEANS na slijedeći način:

```
proc means data=IMEULAZNEDATOTEKE nway mean std skewness kurtosis;  
  var IMEULAZNEVARIJABLE;  
    output out=IMEIZLAZNEDATOTEKE mean=mean std=std skewness=skewness  
    kurtosis=kurtosis;  
run;
```

ili (ako se u data setu IMEULAZNEDATOTEKE nalaze podaci svih iteracija) :

```
proc means data=IMEULAZNEDATOTEKE nway mean std skewness kurtosis nway;  
  var IMEULAZNEVARIJABLE;  
    output out=IMEIZLAZNEDATOTEKE mean=mean std=std skewness=skewness  
    kurtosis=kurtosis;  
class iteracija;  
  
run;
```

4. Odredite (pomoću simulacije) prosječno vrijeme potrebno za kupovinu u trgovinama A i B, uz slijedeće pretpostavke:

- a. U trgovinama A i B radi po jedan prodavač. U trgovini A, 30% kupaca kupuje tako da prilikom kupnje ne traži informacije o proizvodima. Vrijeme kupnje za tu skupinu kupaca slijedi eksponencijalnu distribuciju sa sredinom 3 (min). Preostalih 70% kupaca se prije kupnje informira kod prodavača o proizvodima, pa vrijeme kupnje za tu skupinu slijedi Gamma distribuciju sa sredinom 5 (min). U trgovini B, 40% kupaca kupuje tako da prilikom kupnje ne traži informacije o proizvodima. Vrijeme kupnje za tu skupinu kupaca slijedi eksponencijalnu distribuciju sa sredinom 4 (min). Preostalih 60% kupaca se prije kupnje informira kod prodavača o proizvodima, pa vrijeme kupnje za tu skupinu slijedi Gamma distribuciju sa sredinom 7 (min). Generirajte (generatorima slučajnih brojeva) vremena potrebna za kupovinu za 300 kupaca u trgovini A, te u trgovini B.
- b. Uz pretpostavke navedene pod a) koliko je vremena (sati) potrebno da se usluži svih 300 kupaca u trgovini A, a koliko u trgovini B?
- c. Izračunajte prosječno vrijeme kupovine u trgovini A i u trgovini B. Koliko kupaca može maksimalno biti usluženo u jednom radnom danu od 12 sati u trgovini A, a koliko u trgovini B (uz pretpostavke navedene pod a).
- d. Pretpostavite sve isto kao pod a), osim distribucije vremena kupovine druge skupine kupaca (onih koji se prije kupnje informiraju kod prodavača o proizvodima). Vrijeme kupnje tih kupaca neka sada slijedi eksponencijalnu distribuciju sa sredinom 5 (min) u trgovini A, a eksponencijalnu distribuciju sa sredinom 7 (min) u trgovini B.

- e. Uz pretpostavke navedene pod d) Koliko je vremena (sati) potrebno da se usluži svih 300 kupaca u trgovini A, a koliko u trgovini B?
- f. Izračunajte prosječno vrijeme kupovine u trgovini A i u trgovini B. Koliko kupaca može maksimalno biti usluženo u jednom radnom danu od 12 sati u trgovini A, a koliko u trgovini B (uz pretpostavke navedene pod d).
- g. Ponovite a ), ali sada umjesto 300, generirajte 10000 vremena kupaca. Odredite srednju vrijednost, standardnu devijaciju, koeficijent asimetrije, koeficijent spljoštenosti i standardnu pogrešku za vrijeme potrebno za kupovinu u trgovinama A i B.
- h. Ponovite d ), ali sada umjesto 300, generirajte 10000 vremena kupaca. Odredite srednju vrijednost, standardnu devijaciju, koeficijent asimetrije, koeficijent spljoštenosti i standardnu pogrešku za vrijeme potrebno za kupovinu u trgovinama A i B.

**NAPOMENA:** Koristite vrijednost 47755 za početni SEED (call streaminit(seed);)

**UPUTA:**

Za procjenu statistika pod g) i h) koristite proceduru MEANS:

```
proc means data=imeulaznedatoteke mean sum std skewness kurtosis stderr;
var imevarijable;
class grupa ;
run;
```

5. Provedeno je istraživanje kojim se želi ispitati učinkovitost lijeka za smanjenje sistoličkog krvnog tlaka SKT). U studiji je sudjelovalo 14 ispitanika, kojima je tlak izmjeren prije (SKTprije) i nakon (SKTposlije) uzimanja lijeka.

Za učitavanje podataka u SAS datoteku TLAK koristite slijedeći SAS kod:

```
data TLAK;
    input SKTprije SKTposlije @@;
    datalines;
120 127    128 130    130 131    118 127
140 132    128 125    140 141    135 137
126 118    130 132    130 129    127 120
122 121    141 125
;
run;
```

- a) Testirajte nultu hipotezu da nije došlo do smanjenja sistoličkog krvnog tlaka. (Napišite nultu i alternativnu hipotezu, te primijenite proceduru TTEST sa naredbom PAIRED.) Može li se nulta hipoteza odbaciti?
- b) Generirajte 100 podataka (sa naredbama DO REP= 1 to 100; ... END; za petlju) za SKTprije i 100 za SKTposlije uz pretpostavku da slijede bivarijatnu normalnu distribuciju sa srednjim vrijednostima, standardnim devijacijama i korelacijskim koeficijentom (između SKTprije i SKTposlije) određenim iz podataka (iz datoteke TLAK). Koristite seed=5678. Usporedite vrijednosti srednjih vrijednosti, standardne devijacije i korelacijskog koeficijenta generiranih podataka sa vrijednostima iz

datoteke TLAK (tj. sa originalnim vrijednostima). ). (Za izračunavanje korelacijskog koeficijenta koristite **proc corr** u SAS-u ili **Analyze** → **Multivariate Methods** → **Multivariate** u Jmp-u)

- c) Generirajte 100 x 14 podataka (sa naredbama DO REP= 1 to 100; DO I=1 to 14; ... END; END; za dvostruku petlju), sa istim karakteristikama kao u b). Prikažite grafički distribuciju 100 t vrijednosti. (Primijenite proceduru TTEST sa naredbom PAIRED I sa naredbom BY rep, gdje je rep = 1 ...100.) Izračunajte osnovne deskriptivne statistike I centile za tValue varijablu iz izlaznog data seta TTESTS.
- d) Odredite 5. I 95. Centil za tValue, te prosječnu vrijednost, te ih usporedite sa t-vrijednosti iz koraka a). Da li je distribucija t-vrijednosti "slična" t-distribuciji sa 13 st. slobode?

#### UPUTA:

Za a) I c) Koristite ODS OUTPUT naredbu za ispis t vrijednosti u izlazni dataset TTESTS.

```
ods output ttests=ttests;
proc ttest data=tlak;
paired SKTprije* SKTposlije;
*by rep; *za c) ukloniti *;
RUN;
```

Prije izvođenja c) (a nakon testiranja programa sa 2-3 ponavljanja (&NREP)), primijenite macro %ODSOFF po uputama u Simulating data with SAS Ch6 Simulation Studies.pdf, str. 98-99.

Nakon izvođenja c) primijenite macro %ODSON.

#### UPUTA za b)

```
DO REP=1 ro 100;
  Ovdje upisati naredbe za generiranje SKTprije i SKTposlije
  OUTPUT;
END;
```

#### UPUTA za c)

```
DO REP=1 ro 100;
DO I=1 to 14;
  Ovdje upisati naredbe za generiranje SKTprije i SKTposlije
  OUTPUT;
END;
END;
```