

KOLOKVIJ 2 IZ RAČUNARSKE STATISTIKE

PMF-Matematika, Školska godina 2013/2014

Predavač: Doc.dr.sc. Vesna Lužar-Stiffler

Datum: _____

Ime, Prezime: _____

Potpis: _____

NAPOMENA: Rješenja kolokvija (sas programe) poslati na email adresu: **lubura.snjezana@gmail.com**

1. Monte Carlo metodom procijenite slijedeći određeni integral:

$$I = \int_0^1 \frac{2 \ln(1+x)}{x} dx$$

Koristite slijedeće Importance Sampling (IS) funkcije (tj. funkcije distribucija):

- a) Uniformnu na intervalu 0,1 (U(0,1))
- b) Beta na intervalu (0,1), uz vrijednosti parametara a=1, b=2
- c) Beta na intervalu (0,1), uz vrijednosti parametara a=1, b=1.5
- d) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.01
- e) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.001
- f) Podrežanu („truncated“) normalnu distribuciju sa vrijednostima parametara $\mu=0$, $\sigma=1.5$ (N(0,1.5²)) , podrežanu („truncated“) na intervalu (0,1).

Za svaku importance sampling funkciju generirajte 10000 replikacija/ponavljjanja.

Koristite SEED= 489123 (za sve importance sampling funkcije).

U svakom ponavljanju izračunajte kumulativni prosjek i kumulativnu standardnu pogrešku.

Odredite procjene standardne pogreške za metode a)-f) nakon 10000 replikacija, te efikasnosti metoda b)-f) u odnosu na a).

UPUTA ZA f: Kod generiranja po podrežanoj normalnoj distribuciji zadržavaju se samo one generirane vrijednosti koje su u intervalu (0,1). Funkciju gustoće vjerojatnosti takve „truncated“ normalne distribucije na (0,1) intervalu potrebno je prilagoditi kako bi zadovoljavala svojstva funkcije gustoće vjerojatnosti. Koristite funkcije PROBNORM i PDF.

UPUTA ZA IZRAČUNAVANJE PODINTEGRALNE FUNKCIJE: Za ln (prirodni logaritam) koristite SAS funkciju LOG

Upišite rezultate zaokružene na 4 decimalna mjesta:

	Standardna pogreška:	Efikasnost (u odnosu na a):
a)		-----
b)		
c)		
d)		
e)		
f)		

2. U data setu LIJEKOVI su zapisani podaci mjerenja efikasnosti dvaju lijekova (A i B). Eksperiment je proveden tako da su pacijenti na slučajan način podijeljeni u dvije grupe (A i B). Grupa A (liječnik = "A") je dobivala lijek A, a grupa B lijek B, te je za svakog pacijenta procijenjena efikasnost.

```
data lijekovi;
input lijek $ efikasnost @@;
datalines;
A 4.4 A 4.8 A 3.8 A 4.2 A 5.1 A 4.6 A 4.0 A 2.1
B 5.2 B 4.7 B 5.5 B 4.9 B 5.0 B 4.9
;
```

- a) Testirajte hipotezu $H_0: \mu_A = \mu_B$ (da je EFIKASNOST jednaka bez obzira na LIJEK (A ili B)) nasuprot $H_1: \mu_A \neq \mu_B$ (da je EFIKASNOST lijeka A različita od EFIKASNOSTI lijeka B na razini statističke značajnosti $\alpha = 0.05$ primjenom t testa za 2 nezavisna uzorka.)

UPUTA: Koristite proceduru TTEST. Prvo testirajte hipotezu o jednakosti varijanci na razini statističke značajnosti $\alpha = 0.05$, te u zavisnosti od ishoda, primijenite ili „pooled“ ili „Satterthwaite“ test za hipotezu $H_0: \mu_A = \mu_B$.

Može li se hipoteza H_0 odbaciti na razini statističke značajnosti $\alpha = 0.05$?

Upišite rezultate zaokružene na 4 decimalna mjesta:

TEST	Vrijednost test statistike (F, t)	p-vrijednost	Nulta hipoteza se odbacuje (da/ne)
Jednakosti varijanci			
Jednakosti sredina			

- b) Testirajte hipotezu $H_0: \mu_A = \mu_B$ nasuprot $H_1: \mu_A \neq \mu_B$ na razini statističke značajnosti $\alpha = 0.05$ primjenom bootstrap testa.

UPUTA: Koristite metodu B bootstrap uzorkovanja za testiranje hipoteza. Primijenite slijedeće vrijednosti makro varijabli SEED:

```
%let seed=486227; *za grupu A (liječnik="A");
%let seed= 947123; *za grupu B (liječnik="B");
```

Za svaku grupu izvedite po 1000 bootstrap ponavljanja (replikacija).

Može li se hipoteza H_0 odbaciti na razini statističke značajnosti $\alpha = 0.05$?

Upišite rezultate zaokružene na 4 decimalna mjesta:

Welch t statistika na uzorku:	
Prosječna vrijednost bootstrap vrijednosti Welch t statistike:	
5%-tni percentil bootstrap vrijednosti Welch t statistike:	
95%-tni percentil bootstrap vrijednosti Welch t statistike:	
Bootstrap p-vrijednost za test hipoteze $H_0: \mu_A = \mu_B$ nasuprot $H_1: \mu_A \neq \mu_B$:	
Nulta hipoteza se odbacuje (da/ne)	

3. Koristite podatke iz zadatka 2.

a) Parametarskom bootstrap metodom odredite bootstrap srednju vrijednost, standardnu pogrešku i 90% bootstrap interval pouzdanosti (po percentilnoj metodi) za **t vrijednost** testa za testiranje razlike između prosječnih vrijednosti varijable EFIKASNOST za lijek A i prosječnih vrijednosti varijable EFIKASNOST za lijek B.

$$(t = (\text{meanA} - \text{meanB}) / \sqrt{\text{stdeA}^2 + \text{stdeB}^2}),$$

gdje su meanA i stdeA prosječna vrijednost i standardna pogreška za grupu A,
a meanB i stdeB prosječna vrijednost i standardna pogreška za grupu B.

Upišite rezultate zaokružene na 4 decimalna mjesta:

Bootstrap srednja vrijednost t statistike	
Bootstrap standardna pogreška t vrijednosti	
90% bootstrap interval pouzdanosti (po percentilnoj metodi) za t statistiku	

b) Parametarskom bootstrap metodom odredite bootstrap srednju vrijednost, standardnu pogrešku i 90% bootstrap interval pouzdanosti za **apsolutnu vrijednost razlike između medijana** varijable EFIKASNOST za lijek A i medijana varijable EFIKASNOST za lijek B.

U a) i b):

Primijenite slijedeće vrijednosti makro varijabli SEED:

%let seed= 123456; *za grupu A (lijek="A");

%let seed= 987654; *za grupu B (lijek="B");

Pretpostavite da varijabla EFIKASNOST slijedi normalnu distribuciju.

Za svaku grupu izvedite 1000 bootstrap ponavljanja (replikacija).

Upišite rezultate zaokružene na 4 decimalna mjesta:

Bootstrap srednja vrijednost apsolutne vrijednosti razlike medijana	
Bootstrap standardna pogreška apsolutne vrijednosti razlike medijana	
90% bootstrap interval pouzdanosti (po percentilnoj metodi) za apsolutnu vrijednost razlike medijana	

UPUTA za a):

Koristite proceduru MEANS :

```
proc means data=imeulaznogdataseta nway noprint;  
  var imevarijable;  
  class rep imegrupnevarijable;  
  output out=imeizlaznogdataseta mean=mean stderr=stderr;  
run;
```

NAPOMENA: Nakon izvođenja procedure MEANS će u izlaznom datasetu („imeizlaznogdataseta“) prosječna vrijednost će biti spremljena kao varijabla sa imenom MEAN, a standardna pogreška kao varijabla sa imenom STDERR.

UPUTA za b):

Koristite proceduru MEANS :

```
proc means data=imeulaznogdataseta nway noprint;  
  var imevarijable;  
  class rep imegrupnevarijable;  
  output out=imeizlaznogdataseta median=medijan;  
run;
```

NAPOMENA: Nakon izvođenja procedure MEANS će u izlaznom datasetu („imeizlaznogdataseta“) medijan biti spremljen kao varijabla sa imenom MEDIJAN.