Projektni zadatak iz Statistike

**Ovisnost ukupnog broja ostvarenih bodova na   
kolokvijima iz Difrafa i Intrafa**

Karmen Sović, Teo Šestak, Tihana Štifanić, Marina Švoger, Iva Tutiš

Prirodoslovno-matematički fakultet u Zagrebu  
Matematički odsjek  
akademska godina 2018./2019.  
  
profesor Siniša Slijepčević  
asistentica Ivana Geček Tuđen

# 1. Sažetak

Cilj ovog projekta je promatrati ovisnost ukupnog broja ostvarenih bodova na kolokvijima iz kolegija Diferencijalni račun funkcija više varijabli i Integralni račun funkcija više varijabli, kolegija druge godine preddiplomskog studija Matematika na Matematičkom odsjeku PMF-a u Zagrebu, u akademskoj godini 2017./2018. Difraf se sluša u zimskom semestru te je kolegij prethodnik Intrafu koji se sluša u ljetnom semestru. Studenti koji slušaju navedene kolegije podijeljeni su po abecedi u ukupno dvije grupe na predavanjima i dvije grupe na vježbama. Svaka grupa imala je profesora i asistenta koji su im predavali oba predmeta.

Ovim projektom željeli smo ispitati koreliranost i nezavisnost ukupnog broja ostvarenih bodova na dva kolokvija na jednom kolegiju te ukupnog broja ostvarenih bodova na dva kolokvija na drugom kolegiju kao i dolaze li podaci iz normalne distribucije. Za koreliranost smo koristili test koreliranosti dviju varijabli pomoću Pearsonovog koeficijenta korelacije, za nezavisnost χ2-test o nezavisnosti dviju varijabli, a za provjeru dolaze li podaci iz normalne distribucije Shapiro-Wilkov test. Također, u obzir smo uzimali samo bodove ostvarene na dva redovna kolokvija no ne i ostale elemente ocjenjivanja koji tvore završnu ocjenu pojedinog kolegija.

# 2. Prikupljanje podataka

Na web stranici kolegija (<https://web.math.pmf.unizg.hr/nastava/difraf/dif/> i <https://web.math.pmf.unizg.hr/nastava/difraf/int/>) pronašli smo rezultate kolokvija iz Difrafa i Intrafa studenata koji su te predmete slušali prošle akademske godine. Difraf su tada slušala 144 studenta, a Intraf 118 studenata, međutim, mi smo kao populaciju odlučili uzeti samo one studente koji su prošle godine slušali oba predmeta, a takvih je bilo 109. Zatim smo odabrali slučajan uzorak duljine 50. To smo učinili tako da smo tih 109 studenata i njihove rezultate poslagali u tablicu s numeriranim retcima te pomoću stranice za generiranje slučajnih brojeva (<https://www.random.org/integer-sets/>) generirali 50 slučajnih brojeva u rasponu od 1 do 109 koji se ne ponavljaju. Podatke koji se nalaze u tablici u retku čiji redni broj se nalazi u skupu brojeva koji su generirani smo zapisali u novu tablicu   
(*tablica 2.1.*) i time dobili uzorak.

|  |  |  |
| --- | --- | --- |
| Redni broj osobe | Kolokviji Difraf | Kolokviji Intraf |
| 1 | 72 | 97 |
| 2 | 106 | 57 |
| 3 | 93 | 97 |
| 4 | 179 | 171 |
| 5 | 81 | 73 |
| 6 | 108 | 131 |
| 7 | 135 | 129 |
| 8 | 109 | 116 |
| 9 | 80 | 73 |
| 10 | 124 | 96 |
| 11 | 100 | 104 |
| 12 | 113 | 103 |
| 13 | 98 | 104 |
| 14 | 89 | 87 |
| 15 | 165 | 92 |
| 16 | 88 | 89 |
| 17 | 67 | 50 |
| 18 | 110 | 84 |
| 19 | 127 | 114 |
| 20 | 114 | 94 |
| 21 | 146 | 134 |
| 22 | 131 | 129 |
| 23 | 101 | 44 |
| 24 | 85 | 82 |
| 25 | 165 | 185 |
| 26 | 138 | 122 |
| 27 | 96 | 98 |
| 28 | 97 | 78 |
| 29 | 103 | 92 |
| 30 | 92 | 103 |
| 31 | 127 | 105 |
| 32 | 90 | 51 |
| 33 | 115 | 101 |
| 34 | 137 | 156 |
| 35 | 118 | 57 |
| 36 | 225 | 202 |
| 37 | 198 | 194 |
| 38 | 161 | 125 |
| 39 | 108 | 74 |
| 40 | 132 | 149 |
| 41 | 96 | 84 |
| 42 | 129 | 110 |
| 43 | 82 | 93 |
| 44 | 146 | 116 |
| 45 | 102 | 80 |
| 46 | 131 | 125 |
| 47 | 128 | 95 |
| 48 | 88 | 65 |
| 49 | 131 | 93 |
| 50 | 110 | 98 |

*Tablica 2.1. Uzorak*

# 3. Opisna statistika

Opisnu statistiku provodimo za svaku varijablu posebno, dok na kraju promatramo zajednička dvodimenzionalna statistička obilježja.

## 3.1. Broj bodova na kolokviju iz kolegija Diferencijalni račun funkcija više varijabli

X = broj bodova na kolokviju iz Difrafa

Uzorak: (n = 50)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 72 | 106 | | 93 | | 179 | | 81 | | 108 | | 135 | | | 109 | | 80 | | 124 |
| 100 | 113 | | 98 | | 89 | | 165 | | 88 | | 67 | | | 110 | | 127 | | 114 |
| 146 | 131 | | 101 | | 85 | | 165 | | 138 | | 96 | | | 97 | | 103 | | 92 |
| 127 | 90 | | 115 | | 137 | | 118 | | 225 | | 198 | | | 161 | | 108 | | 132 |
| 96 | 129 | | 82 | | 146 | | 102 | | 131 | | 128 | | | 88 | | 131 | | 110 |
|  | |  | |  | |  | |  | |  | |  |  | |  | |  | |

Dio koda u R-u pomoću kojeg su u vektor učitani podaci:

x<-c(72, 106, 93, 179, 81, 108, 135, 109, 80, 124, 100, 113, 98, 89, 165, 88, 67, 110, 127, 114, 146, 131, 101, 85, 165, 138, 96, 97, 103, 92, 127, 90, 115, 137, 118, 225, 198, 161, 108, 132, 96, 129, 82, 146, 102, 131, 128, 88, 131, 110)

Na osnovu uzorka računamo aritmetičku sredinu, uzoračku varijancu i standardnu devijaciju uzorka koje iznose, redom:

Dio koda u R-u pomoću kojeg su izračunate aritmetička sredina i uzoračka varijanca:

mean(x)  
var(x)

Sortirani uzorak:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 67 | 72 | 80 | 81 | 82 | 85 | 88 | 88 | 89 | 90 |
| 92 | 93 | 96 | 96 | 97 | 98 | 100 | 101 | 102 | 103 |
| 106 | 108 | 108 | 109 | 110 | 110 | 113 | 114 | 115 | 118 |
| 124 | 127 | 127 | 128 | 129 | 131 | 131 | 131 | 132 | 135 |
| 137 | 138 | 146 | 146 | 161 | 165 | 165 | 179 | 198 | 225 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 67.00 | 95.25 | 110.00 | 131.25 | 225.00 |

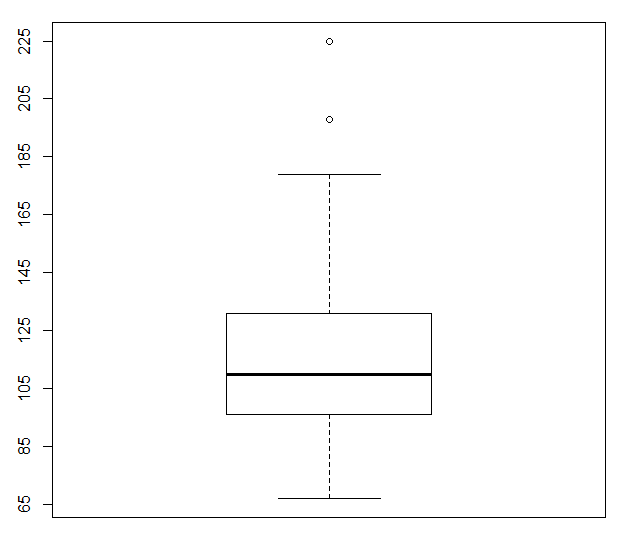
Kvartili:

Iz čega slijedi da je karakteristična petorka uzorka (67.00, 95.25, 110.00, 131.25, 225.00). Također, računamo interkvartil:

Dio koda u R-u pomoću kojeg je uzorak sortiran te su izračunati kvartili i interkvartil:

sort(x)  
quantile(x, type=6)  
IQR(x, type=6)

Iz ovoga slijedi da u uzorku ima 2 outliera iznad gornjeg „brka“, što se vidi na sljedećem dijagramu pravokutnika:



*Slika 3.1.1. Dijagram pravokutnika bodova na kolokvijima iz Difrafa*

Dio koda iz R-a pomoću kojeg je napravljen ovaj dijagram pravokutnika:

boxplot(x, yaxt='n')  
axis(2, at=seq(65, by=20, length=10))

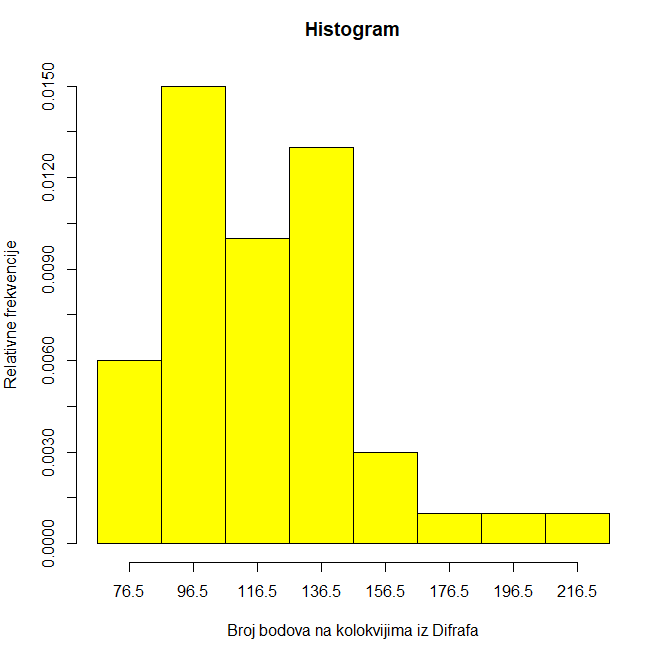
Pri čemu naredbom axis osiguravamo gušći prikaz brojeva na y-osi.

Prilikom crtanja histograma, podatke grupiramo u 8 razreda zbog boljeg prikaza podataka. Računamo potrebnu širinu razreda:

Potom izrađujemo tablicu relativnih frekvencija:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 |  | 6 | 0.12 | 0.006 | 76.5 |
| 2 |  | 15 | 0.3 | 0.015 | 96.5 |
| 3 |  | 10 | 0.2 | 0.01 | 116.5 |
| 4 |  | 13 | 0.26 | 0.013 | 136.5 |
| 5 |  | 3 | 0.06 | 0.003 | 156.5 |
| 6 |  | 1 | 0.02 | 0.001 | 176.5 |
| 7 |  | 1 | 0.02 | 0.001 | 196.5 |
| 8 |  | 1 | 0.02 | 0.001 | 216.5 |
| Σ |  | 50 | 1 |  |  |

*Tablica 3.1.2. Tablica relativnih frekvencija broja bodova na kolokvijima iz Difrafa*



*Slika 3.1.3. Histogram relativnih frekvencija broja bodova na kolokvijima iz Difrafa*

Dio koda iz R-a pomoću kojeg je napravljen ovaj histogram:

hist(x,probability=TRUE,breaks=seq(66.5, by=20, length=9),xlab="Broj bodova na kolokvijima iz Difrafa", ylab="Relativne frekvencije", main="Histogram",col="yellow", xaxt='n', yaxt='n')

axis(1, at=seq(76.5, by=20, length=8))  
axis(2, at=seq(0, by=0.0015, length=11))

Pri čemu dodatni parametri u hist služe za bolji prikaz histograma, kao i pravilno određivanje razreda, dok je dvjema naredbama axis osiguran željeni prikaz brojeva na x, odnosno y-osi.

## 3.2. Broj bodova na kolokviju iz kolegija Integrali funkcija više varijabli

Y = broj bodova na kolokviju iz Intrafa

Uzorak: (n = 50)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 97 | 57 | 97 | 171 | 73 | 131 | 129 | 116 | 73 | 96 |
| 104 | 103 | 104 | 87 | 92 | 89 | 50 | 84 | 114 | 94 |
| 134 | 129 | 44 | 82 | 185 | 122 | 98 | 78 | 92 | 103 |
| 105 | 51 | 101 | 156 | 57 | 202 | 194 | 125 | 74 | 149 |
| 84 | 110 | 93 | 116 | 80 | 125 | 95 | 65 | 93 | 98 |

Dio koda u R-u pomoću kojeg su u vektor učitani podaci:

y<-c(97, 57, 97, 171, 73, 131, 129, 116, 73, 96, 104, 103, 104, 87, 92, 89, 50, 84, 114, 94, 134, 129, 44, 82, 185, 122, 98, 78, 92, 103, 105, 51, 101, 156, 57, 202, 194, 125, 74, 149, 84, 110, 93, 116, 80, 125, 95, 65, 93, 98)

Na osnovu uzorka računamo aritmetičku sredinu, uzoračku varijancu i standardnu devijaciju uzorka koje iznose, redom:

Dio koda u R-u pomoću kojeg su izračunate aritmetička sredina i uzoračka varijanca:

mean(y)  
var(y)

Sortirani uzorak:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 44 | 50 | 51 | 57 | 57 | 65 | 73 | 73 | 74 | 78 |
| 80 | 82 | 84 | 84 | 87 | 89 | 92 | 92 | 93 | 93 |
| 94 | 95 | 96 | 97 | 97 | 98 | 98 | 101 | 103 | 103 |
| 104 | 104 | 105 | 110 | 114 | 116 | 116 | 122 | 125 | 125 |
| 129 | 129 | 131 | 134 | 149 | 156 | 171 | 185 | 194 | 202 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 44.00 | 83.50 | 97.50 | 122.75 | 202.00 |

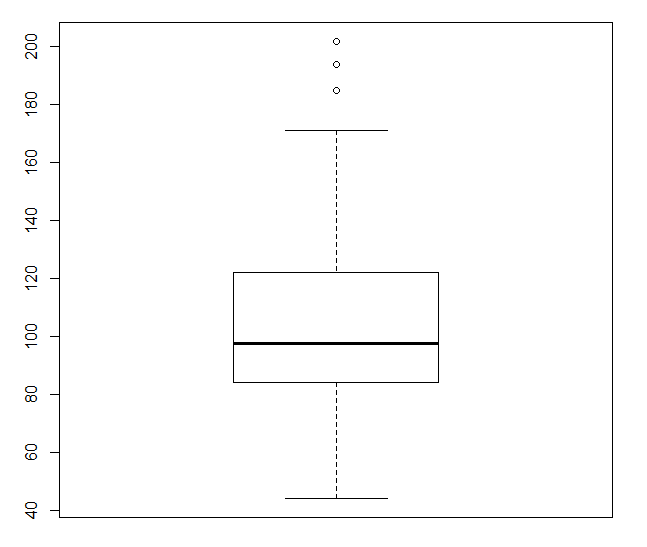
Kvartili:

Iz čega slijedi da je karakteristična petorka uzorka (44.00, 83.50, 97.50, 122.75, 202.00). Također, računamo interkvartil:

Dio koda u R-u pomoću kojeg je uzorak sortiran te su izračunati kvartili i interkvartil:

sort(y)  
quantile(y, type=6)  
IQR(y, type=6)

Iz ovoga slijedi da u uzorku ima 3 outliera iznad gornjeg „brka“, što se vidi na sljedećem dijagramu pravokutnika:



*Slika 3.2.1. Dijagram pravokutnika bodova na kolokvijima iz Intrafa*

Dio koda iz R-a pomoću kojeg je napravljen ovaj dijagram pravokutnika:

boxplot(y, yaxt='n')  
axis(2, at=seq(40, by=20, length=10))

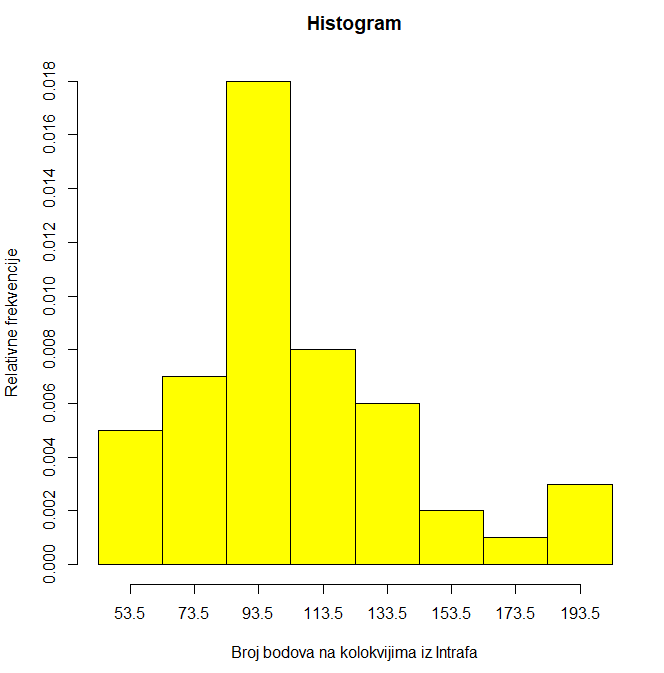
Pri čemu naredbom axis osiguravamo gušći prikaz brojeva na y-osi.

Prilikom crtanja histograma, podatke grupiramo u 8 razreda zbog boljeg prikaza podataka. Računamo potrebnu širinu razreda:

Potom izrađujemo tablicu relativnih frekvencija:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 |  | 5 | 0.1 | 0.005 | 53.5 |
| 2 |  | 7 | 0.14 | 0.007 | 73.5 |
| 3 |  | 18 | 0.36 | 0.018 | 93.5 |
| 4 |  | 8 | 0.16 | 0.008 | 113.5 |
| 5 |  | 6 | 0.12 | 0.006 | 133.5 |
| 6 |  | 2 | 0.04 | 0.002 | 153.5 |
| 7 |  | 1 | 0.02 | 0.001 | 173.5 |
| 8 |  | 3 | 0.06 | 0.003 | 193.5 |
| Σ |  | 50 | 1 |  |  |

*Tablica 3.2.2. Tablica relativnih frekvencija broja bodova na kolokvijima iz Intrafa*



*Slika 3.2.3. Histogram relativnih frekvencija broja bodova na kolokvijima iz Intrafa*

Dio koda iz R-a pomoću kojeg je napravljen ovaj histogram:

hist(y,probability=TRUE,breaks=seq(43.5, by=20, length=9),xlab="Broj bodova na kolokvijima iz Intrafa", ylab="Relativne frekvencije", main="Histogram",col="yellow", xaxt='n', yaxt='n')

axis(1, at=seq(53.5, by=20, length=8))  
axis(2, at=seq(0, by=0.002, length=10))

pri čemu dodatni parametri u hist služe za bolji prikaz histograma, kao i pravilno određivanje razreda, dok je dvjema naredbama axis osiguran željeni prikaz brojeva na x, odnosno y-osi.

## 3.3. Zajednička statistička obilježja

Spareni podaci, u obliku (X, Y), pri čemu su, kao i gore, X broj bodova ostvaren na kolokvijima iz Difrafa, a Y broj bodova ostvaren na kolokvijima iz Intrafa:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (72, 97) | (106, 57) | (93, 97) | (179, 171) | (81, 73) |
| (108, 131) | (135, 129) | (109, 116) | (80, 73) | (124, 96) |
| (100, 104) | (113, 103) | (98, 104) | (89, 87) | (165, 92) |
| (88, 89) | (67, 50) | (110, 84) | (127, 114) | (114, 94) |
| (146, 134) | (131, 129) | (101, 44) | (85, 82) | (165, 185) |
| (138, 122) | (96, 98) | (97, 78) | (103, 92) | (92, 103) |
| (127, 105) | (90, 51) | (115, 101) | (137, 156) | (118, 57) |
| (225, 202) | (198, 194) | (161, 125) | (108, 74) | (132, 149) |
| (96, 84) | (129, 110) | (82, 93) | (146, 116) | (102, 80) |
| (131, 125) | (128, 95) | (88, 65) | (131, 93) | (110, 98) |

Zbog naravi podataka (veliki raspon vrijednosti koje se postižu), prilikom izrade kontingencijske tablice podatke grupiramo u 5 kategorija prema ostvarenom broju bodova, i to u one koje su postigli manje 100 bodova, između 100 i 120, između 121 i 150, između 151 i 180 te više od 180 bodova.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X/Y | <100 | 100-120 | 121-150 | 151-180 | >180 | Σ |
| <100 | 14 | 2 | 0 | 0 | 0 | 16 |
| 100-120 | 9 | 4 | 1 | 0 | 0 | 14 |
| 121-150 | 3 | 4 | 6 | 1 | 0 | 14 |
| 151-180 | 1 | 0 | 1 | 1 | 1 | 4 |
| >180 | 0 | 0 | 0 | 0 | 2 | 2 |
| Σ | 27 | 10 | 8 | 2 | 3 | 50 |

Tablica 3.3.1. Kontingencijska tablica

# 4. Statistički testovi

## 4.1. Testovi koreliranosti dviju varijabli

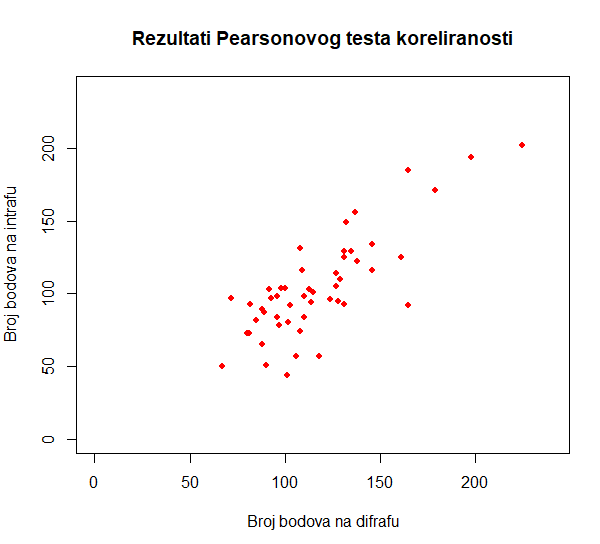
**4.1.1 Pearsonov test koreliranosti**

Za ovaj test nam je potrebna pretpostavka o normalnoj razdiobi varijabli *X* i *Y*.

Neka je oznaka za Pearsonov koeficijent korelacije varijabli *X* i *Y*.

Želimo testirati nultu hipotezu  
 to jest da varijable X i Y nisu korelirane, u odnosu na jednu od alternativa  
to jest da su varijable negativno (pozitivno) korelirane.

Stavimo li u graf dosadašnje podatke, dobivamo

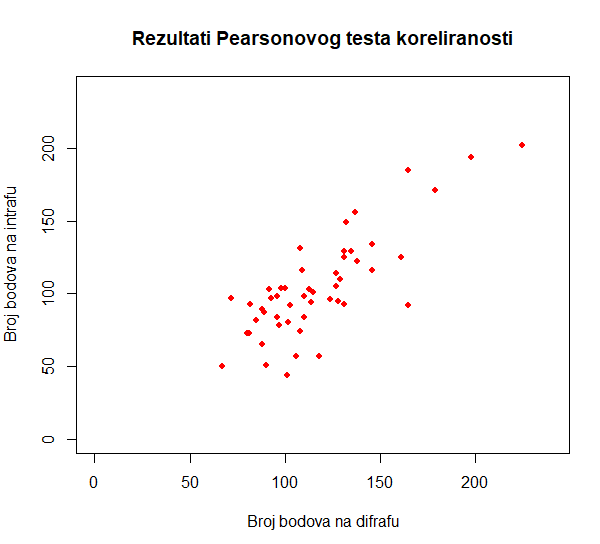


*Graf 4.1.1. Rezultati testa koreliranosti pomoću Pearsonovog koeficijenta korelacije*

gdje je kod u R-u kojim smo napravili graf podataka

plot(x,y, pch=20,col="red",main="Rezultati Pearsonovog testa koreliranosti",xlab="Broj bodova na difrafu",ylab="Broj bodova na intrafu", xlim=c(0,240), ylim=c(0, 240), cex=1.3)

Iz grafa je očito da oko podataka možemo „povući“ elipsu kojoj je šira os na pravcu s pozitivnim koeficijentom smjera, kao dolje:



*Graf 4.1.2. Rezultati testa koreliranosti pomoću Pearsonovog koeficijenta korelacije s pripadnom elipsom*

Naša testna t-statistika je  
gdje je .  
Također, definira se:

Provedemo li test, dobivamo podatke:

* Izračunata realizacija t-statistike je
* Broj stupnjeva slobode je
* Kritično područje za je
* p-vrijednost je
* 95% pouzdani interval je
* Pearsonov koeficijent korelacije je

gdje je kod u R-u kojim se izvršava test:

cor.test(x,y, method = "pearson")

## Vidimo da t upada u kritično područje, to jest da je pa odbacujemo hipotezu *H0* u korist *H1* na razini značajnosti . Iz vrijednosti Pearsonovog koeficijenta možemo zaključiti da su varijable X i Y pozitivno korelirane (uz pretpostavku normalne razdiobe varijabli) jer je

***4.1.2. Spearmanov test koreliranosti***

Međutim, pogledamo li histograme u odjeljku 3.1 i 3.2, jasno vidimo da razdiobe naših varijabli X i Y po razredima odstupaju od pretpostavljene normalne razdiobe. Proučimo sada koeficijent korelacije kakav bi smo dobili *Spearmanovim testom koreliranosti,* koji ne zahtjeva normalnu distribuciju varijabli.

Neka je oznaka za Spearmanov koeficijent korelacije varijabli *X* i *Y*.

Želimo testirati nultu hipotezu  
 to jest da varijable X i Y nisu korelirane, u odnosu na jednu od alternativa  
to jest da su varijable negativno (pozitivno) korelirane.

Test funkcionira tako da se vrijednosti varijable X „rangiraju“ po veličini od najvećeg do najmanjeg, gdje se višestrukim vrijednostima istog iznosa daje isti rang koji odgovara prosjeku rangova koje bi trebale biti dodjeljene tim vrijednostima (na primjer ako su 3. i 4. vrijednost istog iznosa, obje dobivaju rang 3.5). Sasvim analogno rangiramo vrijednosti varijable Y.

Zatim, gledajući uređene parove podataka i njihove pripadne rangove , računa se:

Te se koeficijent korelacije dobiva računom iz statistike:

Gdje je Spearmanov koeficijent korelacije, a broj parova podataka.

Kod u R-u kojim se izvršava test:

cor.test(x,y, method = "spearman")

Provedemo li test, dobivamo rezultat:

* p-vrijednost je:
* Spearmanov test koreliranosti je:
* Odbacuje se hipoteza *H0* u korist *H1*

## Iz rezultata i male p-vrijednosti odbacujemo hipotezu H0 u korist H1 , a iz vrijednosti Spearmanovog koeficijenta zaključujemo da su varijable X i Y pozitivno korelirane jer je

## 

## 4.2. χ2-test nezavisnosti dviju varijabli

Testiramo nezavisnost varijabli X i Y, pri čemu su brojevi bodova podjeljeni u razrede od manje od 100 bodova, između 100 i 120, između 121 i 150, između 151 i 180 te više od 180 bodova, sa opaženim frekvencijama kao u tablici 3.3.1.

Hipoteze su sljedeće:

*H0*: X i Y su nezavisne

*H1*: X i Y nisu nezavisne

Očekivane frekvencije uz hipotezu H0 dane su tablicom:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X/Y | <100 | 100-120 | 121-150 | 151-180 | >180 | Σ |
| <100 | 8.64 | 3.20 | 2.56 | 0.64 | 0.96 | 16 |
| 100-120 | 7.56 | 2.8 | 2.24 | 0.56 | 0.84 | 14 |
| 121-150 | 7.56 | 2.8 | 2.24 | 0.56 | 0.84 | 14 |
| 151-180 | 2.16 | 0.8 | 0.64 | 0.16 | 0.24 | 4 |
| >180 | 1.08 | 0.4 | 0.32 | 0.08 | 0.12 | 2 |
| Σ | 27 | 10 | 8 | 2 | 3 | 50 |

Tablica 4.2.1. Očekivane frekvencije

Testna statistika je:

,

gdje su i marginalne distribucije redom X i Y, podijeljene s duljinom uzorka, a distribucije vektora (X, Y).

Nadalje, kritično područje je , za dani α.

Računom dobivamo da je i p-vrijednost .

Dio koda u R-u kojim provodimo test:

tablica<-matrix(c(14,2,0,0,0,9,4,1,0,0,3,4,6,1,0,1,0,1,1,1,0,0,0,0,2),nrow=5,byrow=T)

chisq.test(tablica)

Dakle, za odbacujemo hipotezu *H0* u korist *H1*.

## 4.3. Shapiro –Wilk test

Shapiro –Wilkov test je inačica Kolmogorov-Smirnovljeva testa. Njime provjeravamo dolaze li podaci iz normalne distribucije.

Testiramo hipoteze:

*H0*: varijabla ima normalnu distribuciju  
*H1*: varijabla nema normalnu distribuciju

Testna statistika je

gdje je statistički niz , aritmetička sredina uzorka dok su parametri određeni izrazom

gdje je , a vrijednosti su očekivane vrijednosti statističkog niza nezavisne i jednako distribuirane slučajne varijable uzorkovane iz osnovnog skupa normalne distribucije, a V je matrica kovarijance tih statističkih nizova.

Nulta hipoteza se odbija ukoliko je vrijednost test veličine W premalena, odnosno ukoliko je p –vrijednost manja od razine vjerodostojnosti α.

U R-u se računa naredbom:

shapiro.test(x)

gdje je x numerički vektor, s vrijednostima između 3 i 5000, s tim da neke vrijednosti mogu nedostajati. Dobivamo:

Difraf

* W=0.9231
* p-vrijednost=0.003059

Stoga za α > 0.003059 odbacujemo *H0* u korist *H1*, to jest podaci nisu normalno distibuirani.

Intraf

* W=0.93351
* p-vrijednost=0.007524

Stoga za α > 0.007524 odbacujemo *H0* u korist *H1*, to jest podaci nisu normalno distibuirani.

# 5. Zaključak

Uz pretpostavku normalne distribuiranosti varijabli X i Y i toga što je p-vrijednost jako mala iz *Pearsonovog testa koreliranosti dviju varijabli* vidimo da za svaki „razuman“ odbacujemo hipotezu da varijable nisu korelirane. Bez potrebe za pretpostavkom normalne distribuiranosti varijabli X i Y, *Spearmanov test koreliranosti dviju varijabli* nas opet (zbog vrlo male p-vrijednosti) navodi na odbacivanje hipoteze nekoreliranosti. Oba testa nam indiciraju da su X i Y pozitivno korelirane.

Slično, iz *χ2-testa* za nezavisnost vidimo da za svaki „razuman“ varijable X i Y nisu nezavisne. Na primjer, za =0.01 možemo reći da brojevi bodova ostvareni na kolokviju iz Difrafa nisu nezavisni s brojevima bodova ostvarenim na kolokviju iz Intrafa te da su ti podaci pozitivno korelirani što smo na neki način i očekivali.

Shapiro-Wilkov test nam daje zaključak da za razinu značajnost α > 0.003059 za Difraf odnosno za α > 0.007524 za Intraf podaci nisu normalno distribuirani.

# 6. Literatura

<https://web.math.pmf.unizg.hr/nastava/difraf/dif/>  
<https://web.math.pmf.unizg.hr/nastava/difraf/int/>  
<https://www.random.org/integer-sets/>  
<https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test>  
Predavanja iz Statistike profesora M. Huzaka, Statistički testovi 3, slide 1-9  
Vježbe iz Statistike, χ2-test i Kolmogorov-Smirnovljev test, χ2-test o nezavisnosti, str. 137-139