

REŠENJA ZADATAKA

*** Na dokumentu reldi2.tsv ispitajte sledeće hipoteze:

(1) Prosečna dužina reči razlikuje se u hrvatskom, srpskom (latinično pismo) i srpskom (ćirilično pismo), bilo da se iskaže (a) numerički ili (b) kategorički

a)

- Provera normalnosti distribucije

```
shapiro.test(reldi2$duzina.reci[reldi2$korpus=="hrwac"]) # i tako za svaki korpus posebno
```

ili

```
tapply(reldi2$duzina.reci, reldi2$korpus, shapiro.test) # jednopotezno izračunavanje za sva tri korpusa
```

- Test

```
kruskal.test(reldi2$duzina.reci~reldi2$korpus)
```

- Deskripcija

```
tapply(reldi2$duzina.reci, reldi2$korpus, median)
```

- Grafički prikaz

```
boxplot(reldi2$duzina.reci~reldi2$korpus)
```

b)

- Tabela unakrsne klasifikacije

```
duzina.reci.2.po.korpusima=table(reldi2$korpus, reldi2$duzina.reci.2)
```

- Test

```
chisq.test(duzina.reci.2.po.korpusima)
```

- Grafički prikaz

```
mosaicplot(duzina.reci.2.po.korpusima)
```

(2) Infinitivi su češći u hrvatskom nego u srpskom jeziku

- Promijenili .tsv file (find-replace \tsrwac.(lat|cyr)\t - \tsrwac\t)

```
reldi2.5 = read.delim("reldi2-samo_srwac.tsv")  
summary(reldi2.5)
```

- Plotamo jer to volimo

```
boxplot(reldi2.5$infinitivi~reldi2.5$korpus)
```

ili

```
boxplot(reldi2$infinitivi[reldi2$korpus=="hrvac"],reldi2$infinitivi[reldi2$korpus=="srwac.cyr"|reldi2$korpus=="srwac.lat"]) # ako korpusi na srpskom nisu objedinjeni u izvornoj tabeli, potrebno ih je objединити u formuli; isto važi i za funkcije ispod
```

- Provjeravamo je li distribucija normalna (nije)

```
tapply(reldi2.5$infinitivi, reldi2.5$korpus, shapiro.test)
```

- Testiramo infinite i statistički su značajni

```
wilcox.test(reldi2.5$infinitivi~reldi2.5$korpus)
```

ili

```
wilcox.test(reldi2$infinitivi[reldi2$korpus=="hrvac"],reldi2$infinitivi[reldi2$korpus=="srwac.cyr"|reldi2$korpus=="srwac.lat"])
```

- Veznik da je češći u srpskom nego u hrvatskom jeziku

```
shapiro.test(reldi2$da[reldi2$korpus=="hrvac"])
```

```
shapiro.test(reldi2$da[reldi2$korpus=="srwac.cyr"|reldi2$korpus=="srwac.lat"])
```

```
kruskal.test(reldi2.5$da~reldi2.5$korpus)
```

```
boxplot(reldi2.5$da~reldi2.5$korpus)
```

```
tapply(reldi2.5$da, reldi2.5$korpus, median)
```

*** Na dokumentu cirlat.csv ispitajte sledeće hipoteze:

(1) Reči koje se različito čitaju i različite su reči na ćirilici i latinici (primer: KOB) obrađuju se sporije od reči koje su iste na oba pisma (primer: JAMA)

```
shapiro.test(cirlat$RT3)
```

```
shapiro.test(cirlat$RT5)
```

```
wilcox.test(cirlat$RT3, cirlat$RT5, paired=TRUE)
```

```
boxplot (cirlat$RT3, cirlat$RT5)
```

(2) Za reči koje se mogu pročitati i na ćirilici, ali su reči samo na latinici (primer: TAJAC) postoji povezanost između procenta tačnosti odgovora i vremena reakcije

```
shapiro.test(cirlat$Tacno4)
```

```
shapiro.test(cirlat$RT4)
```

```
cor.test(cirlat$Tacno4, cirlat$RT4, method="spearman")
```

```
plot(cirlat$Tacno4,cirlat$RT4)
```