

## Vježbe 9. i 10. Testiranje hipoteza - dvije populacije

### Usporedba srednjih vrijednosti dviju populacija

#### Zadatak 1: Komarci

Komarci pronalaze svoje žrtve dijelom pomoću mirisa pa ima smisla pitati se utječe li ono što jedemo i pijemo na našu privlačnost za komarce. Studija provedena u Zapadnoj Africi (Lefèvre i sur., 2010.), promatrajući vrste komaraca koji prenose malariju, istraživala je utječe li konzumacija lokalnog piva na privlačnost za komarce.

Istraživači su otvorili posudu s 50 komaraca pored svakog od 25 sudionika koji nisu konzumirali alkohol i izmjerili omjer komaraca koji su napustili posudu i poletjeli prema sudionicima. Isti postupak ponovili su 15 minuta nakon što su isti sudionici popili litru piva te su izmjerili "promjenu u aktivaciji" (vrijednost nakon minus vrijednost prije).

Ovaj postupak također je proveden na 18 sudionika koji su dobili vodu umjesto piva. Promjena u aktivaciji komaraca zabilježena je za obje skupine: onu koja je pila pivo i onu koja je pila vodu. Podaci se nalaze u bazi "**komarci.csv**".

1. Nacrtajte usporedni boxplot podataka za obe populacije.
2. Provjerite normalnost podataka pomoću histograma i normalnog vjerojatnosnog grafa.
3. Uz pretpostavku normalnosti, testirajte hipotezu o jednakosti varijanci.
4. Provedite odgovarajući test usporedbe.

#### Zadatak 2: Banke

Promatramo broj ljudi koji dolaze u banku C i banku D. Nasumice smo odabrali dane u godini i zabilježili broj osoba koji su taj dan posjetili promatranu banku. Podatci se nalaze u bazi "**banke.csv**".

1. Nacrtajte usporedni boxplot podataka za obe populacije.
2. Provjerite normalnost podataka pomoću histograma i normalnog vjerojatnosnog grafa.
3. Testirajte hipotezu o jednakosti varijanci.
4. Provedite odgovarajući test usporedbe.

#### Zadatak 3: Laboratorij (DZ)

U nekom laboratoriju mjerimo vrijeme da laboratorijski uređaj završi s analizom na uzorku. Analiza se vrši na dva laboratorijska uređaja koji dolaze iz firme F, odnosno G. Svaki je uzorak podijeljen na dva manja, tako da oba uređaja mogu analizirati isti uzorak. Podatci se nalaze u bazi "**labos.csv**". Možete li na razini značajnosti od 5% tvrditi da je jedan od uređaja brži od drugog?

## Usporedba srednjih vrijednosti dviju populacija - sparene vrijednosti

### Zadatak 4: Mortalitet

U datoteci **"mortalitet.csv"** dani su podatci o mortalitetu u državama u vrijeme niskog poreza i nakon prijelaza na visoku stopu poreza na nekretnine. Odredite je li došlo do promjene mortaliteta na razini značajnosti 5%.

### Zadatak 5: Poliklinika (DZ)

Poliklinika Erdut uvela je novi način podjele termina za preglede - putem aplikacije. Poliklinika misli da će na taj način privući više klijenata. Imaju podatke o posjetama prije uvođenja aplikacije, i nakon uvođenja. Pokrenite kod iz dokumenta **"kako\_vjezbati\_testiranje.R"** s već postavljenim brojevima. Potom rezultirajuću bazu **"mojabaza.csv"** neka bude baza koja sadrži informacije o broju klijenata prije (varijabla) i poslije (varijabla2) uvođenja aplikacije. Postavite hipoteze i na razini značajnosti od 5% provjerite možete li tvrditi da je aplikacija povećala posjećenost.

## Testiranje hipoteza o varijanci

### Zadatak 6: Ribe

Zapanjujuća raznolikost ciklidnih riba u Viktorijinom jezeru održava se zahvaljujući preferencijama ženki prema mužjacima vlastite vrste. Kako bi se razumjelo kako su te vrste uopće nastale, važno je znati genetsku osnovu te preferencije kod ženki. Istraživači su križali dvije vrste ciklida, *Pundamilia pundamilia* i *P. nyererei*, te su uzgojili njihovo potomstvo (hibridi prve generacije, "F1") do odrasle dobi. Izmjerali su stupanj preferencije ženki F1 riba prema mužjacima vrste *P. pundamilia* u usporedbi s mužjacima vrste *P. nyererei* (Haesler i Seehausen, 2005). Zatim su križali F1 hibride međusobno kako bi proizveli drugu generaciju hibrida ("F2"), koje su također uzgojili do odrasle dobi i izmjerili isti indeks preferencije ženki. Ako mali broj gena igra važnu ulogu u određivanju te preferencije, tada će varijanca indeksa preferencije biti različita između ove dvije generacije (najveća će biti kod F2 hibrida). Istraživači su izmjerili preferencije kod 20 F1 jedinki i 33 F2 jedinke. Podaci se nalaze u datoteci **"ribe.csv"**.

1. Nacrtajte usporedni boxplot za obe populacije. Interpretirajte podatke.
2. Uz pretpostavku normalnosti podataka, provedite testiranje jednakosti varijanci.

## Normalnost podataka

### Zadatak 7: Normalnost

Za populacije iz prethodnih zadataka, provjerite normalnost podataka pomoću histograma te pomoću normalnog vjerojatnosnog grafa, izračunajte zakošenost i spljoštenost. Potrebne naredbe se nalaze u paketu **"e1071"**.