

Pregled procedura u Statistici

PTF - Primijenjena i inženjerska matematika, ak. god. 2023./2024.

1 Pouzdani intervali

1.1 Pouzdani interval za očekivanje

1.1.1 z -interval

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz $\mathcal{N}(\mu, \sigma^2)$, σ^2 poznato

Računa se: pomoću formula

1.1.2 t -interval

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz $\mathcal{N}(\mu, \sigma^2)$, σ^2 nepoznato ili X_1, \dots, X_n slučajan uzorak, n velik, $\mu = EX_1$, $\sigma^2 = \text{Var}(X_1) < \infty$

Računa se:

- pomoću formula
- na osnovu stupca podataka Statistica: Basic Statistics → Descriptive statistics označiti varijablu, zatim [Advanced] → ☒ Conf. limits for means, upisati razinu pouzdanosti → Summary

1.2 Pouzdani interval za vjerojatnost događaja (proporciju)

Pretpostavke: X_1, \dots, X_n velik slučajan uzorak iz Bernoullijeve distribucije

Računa se:

- pomoću formula
- Statistica: Power Analysis → Interval Estimation → One proportion, Z, Chi-Square Test → OK unijeti relativnu frekvenciju, veličinu uzorka i pouzdanost, zatim Compute, u ispisu očitati Pi (Crude)

2 Testovi temeljeni na jednom uzorku (varijabli)

2.1 Testiranje hipoteza o očekivanju

Nul-hipoteza: $H_0 : \mu = \mu_0$

Uzorak: X_1, \dots, X_n

2.1.1 z -test

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz $\mathcal{N}(\mu, \sigma^2)$, σ^2 poznato

Provodi se: pomoću formula

2.1.2 t -test

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz $\mathcal{N}(\mu, \sigma^2)$, σ^2 nepoznato ili X_1, \dots, X_n slučajan uzorak, n velik, $\mu = EX_1$, $\sigma^2 = \text{Var}(X_1) < \infty$

Provodi se:

- pomoću formula
- ako su dostupni samo sumarni podaci – Statistica: Basic Statistics → Difference tests → Difference between two means → Single mean 1 vs population mean 2
- ako je dostupan originalni uzorak – Statistica: Basic Statistics → t-test, single sample (p -vrijednost je za dvostranu alternativnu hipotezu)

2.2 Testiranje hipoteza o vjerojatnosti događaja (proporciji)

Nul-hipoteza: $H_0 : p = p_0$

Uzorak: X_1, \dots, X_n

2.2.1 z -test

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz Bernoullijeve distribucije s parametrom p , n velik

Provodi se:

- pomoću formula
- koristeći χ^2 test (p -vrijednost je za dvostranu alternativnu hipotezu)

2.3 Testiranje hipoteza o distribuciji

Nul-hipoteza: H_0 : distribucija iz koje dolazi uzorak jednaka je teorijskoj

Uzorak: X_1, \dots, X_n

2.3.1 χ^2 -test

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz neke diskretne distribucije, teorijska distribucija diskretna

Provodi se: Statistica: Nonparametrics → Observed versus expected X2

2.3.2 Testovi normalnosti

Pretpostavke: X_1, \dots, X_n slučajan uzorak iz neprekidne distribucije, teorijska distribucija normalna

Provodi se: Statistica: Basic Statistics → Descriptive statistics [Normality] → Shapiro-Wilk W test → Histograms

3 Testovi temeljeni na dva uzorka (dvije varijable)

3.1 Usporedba očekivanja - nezavisni uzorci

Nul-hipoteza: $H_0 : \mu_1 = \mu_2$
Uzorci: $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ i $X_1^{(2)}, \dots, X_{n_2}^{(2)}$

3.1.1 z -test

Pretpostavke: $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ slučajan uzorak iz $\mathcal{N}(\mu_1, \sigma_1^2)$, $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ slučajan uzorak iz $\mathcal{N}(\mu_2, \sigma_2^2)$, σ_1^2 i σ_2^2 poznati

Provodi se: pomoću formula

3.1.2 Welchov t -test

Pretpostavke: $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ slučajan uzorak iz $\mathcal{N}(\mu_1, \sigma_1^2)$, $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ slučajan uzorak iz $\mathcal{N}(\mu_2, \sigma_2^2)$, σ_1^2 i σ_2^2 nepoznati, ne nužno jednake (za velike uzorke možemo koristiti i bez pretpostavke normalnosti)

Provodi se:

- ako su uzorci organizirani u dva stupca – Statistica: Basic Statistics → t-test, independent, by variables [Options] → t-test with separate variance estimates (p -vrijednost je za dvostranu alternativnu hipotezu)
- ako su uzorci organizirani u jedan stupac uz grupirajuću varijablu – Statistica: Basic Statistics → t-test, independent, by groups [Options] → t-test with separate variance estimates (p -vrijednost je za dvostranu alternativnu hipotezu)

3.2 Usporedba proporcija - nezavisni uzorci

Nul-hipoteza: $H_0 : p_1 = p_2$
Uzorci: $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ i $X_1^{(2)}, \dots, X_{n_2}^{(2)}$

3.2.1 z-test

Pretpostavke: $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ slučajan uzorak iz Bernoullijeve distribucije s parametrom p_1 , $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ slučajan uzorak iz Bernoullijeve distribucije s parametrom p_2

Provodi se:

- pomoću formula
- Statistica: Basic Statistics → Difference tests → Difference between two proportions

3.3 Usporedba očekivanja - vezani uzorci

Nul-hipoteza: $H_0 : \mu_1 = \mu_2$

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n

3.3.1 t-test

Pretpostavke: razlike $D_i = X_i - Y_i, i = 1, \dots, n$ su normalno distribuirane ili je uzorak velik – test se svodi na test na jednom uzorku

Provodi se:

- pomoću formula
- Statistica: Basic Statistics → t-test, dependent samples (p -vrijednost je za dvostranu alternativnu hipotezu)

3.4 Empirijska distribucija slučajnog vektora

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n iz diskretnih distribucija

Uputa: Statistica: Basic Statistics → Tables and banners → Specify tables (select variables) označiti varijable → OK [Options] → Percentages of total count; za procjene uvjetnih distribucija označiti Percentages of row counts i Percentages of column counts

3.5 Testiranje hipoteze o nezavisnosti

Nul-hipoteza: H_0 : slučajne varijable iz kojih dolaze uzorci su nezavisne

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n

3.5.1 χ^2 -test

Pretpostavke: uzorci su iz diskretnih distribucija

Provodi se: Statistica: Basic Statistics → Tables and banners → Specify tables (select variables) označiti varijable → OK [Options] → Pearson & M-L Chi-square → [Advanced] → Detailed two-way tables, u ispisu očitati p -vrijednost za Pearson Chi-square

3.6 Jednostavna linearna regresija

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n iz neprekidnih distribucija

3.6.1 Dijagram raspršenosti

Uputa: Statistica: Graphs → Scatterplot → Variables označiti varijable → OK

3.6.2 Procjena regresijskog pravca

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → Summary: Regression results parametri su u ispisu označeni kao stupac b – intercept (slobodni član) i koeficijent smjera

3.6.3 Predikcija

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → [Residuals/assumptions/prediction] → Predict dependent variable unijeti vrijednost nezavisne varijable → OK, u ispisu očitati vrijednost Predicted

3.6.4 Analiza reziduala - homogenost varijanci grešaka

Graf predviđenih vrijednosti u odnosu na reziduale

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → [Residuals/assumptions/prediction] → Perform residuals analysis → [Scatterplots] → Predicted vs. residuals

3.6.5 Analiza reziduala - nezavisnost grešaka

Graf vrijednosti nezavisne varijable u odnosu na reziduale

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → [Residuals/assumptions/prediction] → Perform residuals analysis → [Residuals] → Residuals vs. independent var. označiti nezavisnu varijablu → OK

Dijagram raspršenosti susjednih grešaka

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → [Residuals/assumptions/prediction] → Perform residuals analysis → Summary: Residuals & predicted stupac Residual kopirati i dodati u tablicu podataka, zatim isto to napraviti ali kopirati od drugog retka tablice i zadnji redak obrisati; na Statistica: Graphs → Scatterplot nacrtati dijagram raspršenosti za dva nova stupca

3.6.6 Analiza reziduala - normalnost grešaka

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → [Residuals/assumptions/prediction] → Perform residuals analysis → Summary: Residuals & predicted stupac Residual kopirati i dodati u tablicu podataka te na njemu provesti Shapiro-Wilk test

3.6.7 Koeficijent determinacije

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → Summary: Regression results u tablici Summary Statistics očitati Multiple R2 vrijednost

3.6.8 Zaključivanje o koeficijentu smjera

Uputa: Statistica: Multiple Regression → Variables označiti varijable → OK → Summary: Regression results u retku za nezavisnu varijablu očitati p -vrijednost

3.7 Testiranje hipoteze o koreliranosti

Nul-hipoteza: H_0 : slučajne varijable iz kojih dolaze uzorci su korelirane ($\rho = 0$)

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n

3.7.1 Test o koeficijentu korelacije

Pretpostavke: uzorci su iz neprekidnih distribucija

Provodi se:

- Statistica: Basic Statistics → Correlation matrices [Options] → Display detailed table of results

3.8 Testiranje hipoteze o monotonj vezi

Nul-hipoteza: H_0 : postoji monotona veza između slučajnih varijabli iz kojih dolaze uzorci ($\rho_S = 0$)

Uzorci: X_1, \dots, X_n i Y_1, \dots, Y_n

3.8.1 Test o Spearmanovom koeficijentu korelacije ranga

Provodi se:

- Statistica: Nonparametrics → Correlations → Compute: Detailed report

4 Testovi temeljeni na više od dva uzorka (varijable)

4.1 Usporedba očekivanja

Nul-hipoteza: H_0 : $\mu_1 = \mu_2 = \dots = \mu_k$

Uzorci: $Y_{11}, \dots, Y_{1n_1}, \dots, Y_{k1}, \dots, Y_{kn_k}$

4.1.1 Jednofaktorska analiza varijance (ANOVA)

Pretpostavke: $Y_{11}, \dots, Y_{1n_1}, \dots, Y_{k1}, \dots, Y_{kn_k}$ su redom slučajni uzorci iz $\mathcal{N}(\mu_1, \sigma^2), \dots, \mathcal{N}(\mu_k, \sigma^2)$

Provodi se:

- Statistica: Basic Statistics → Breakdown & one-way ANOVA, odabrati zavisnu varijablu i faktor, → Analysis of Variance i očitati p -vrijednost
- homogenost varijanci može se testirati Leveneovim testom (nul-hipoteza je da su varijance homogene) na [ANOVA & tests] → Levene's tests
- ukoliko je pretpostavka homogenosti varijanci narušena, može se koristiti Welchova varijanta testa na [ANOVA & tests] → Perform Welch's F-Test → Analysis of Variance