GRAĐEVINSKI FAKULTET
Odsek za hidrotehniku
i vodno ekološko inženjerstvo

STOHASTIČKA HIDROLOGIJA

Godišnji zadatak

Vežba 1: Testiranje homogenosti i slučajnosti niza godišnjih maksimuma proticaja

Za zadati niz godišnjih maksimuma sprovesti sledeće testove:

a) testovi homogenosti:

- z-test
- F-test
- t-test
- Men-Vitni test
- Kolmogorov-Smirnov test

b) testovi slučajnosti:

- test kvadrata uzastopnih razlika
- Vold-Volfovic test
- Bartletov test za koeficijent autokorelacije

Za sve testove usvojiti prag značajnosti od 5%.

Vežba 2: Određivanje velikih voda metodom godišnjih maksimuma

Za zadati niz godišnjih maksimuma uraditi sledeće:

- odrediti statistike originalnog niza (X = Q), odgovarajućeg logaritmovanog niza (Y = log X);
- proveriti postojanje izuzetaka pomoću testa Grubsa i Beka;
- odrediti empirijsku raspodelu proračunom kompromisne verovatnoće;
- odrediti kvantile velikih voda pomoću sledećih teorijskih raspodela: log-normalna, Gumbelova, Pirson 3 i log-Pirson 3;
- prikazati teorijske raspodele i empirijske tačke na dijagramu normalne verovatnoće;
- testirati saglasnost teorijskih raspodela sa empirijskom raspodelom pomoću testova Kolmogorova i Kramer-Mizesa; na osnovu testova usvojiti najpovoljniju teorijsku raspodelu;
- odrediti interval poverenja od 95% za odabranu teorijsku raspodelu i prikazati ga grafički na dijagramu normalne verovatnoće

Vežba 3: Metoda pikova za proračun velikih voda

Za niz pikova iznad zadatog praga uraditi sledeće:

a) Obrada broja javljanja pikova:

- formiranje niza
- proračun statistika
- proračun empirijskih frekvencija
- izbor teorijske raspodele
- proračun teorijskih frekvencija
- testiranje slaganja teorijskih i empirijskih frekvencija pomoću hi-kvadrat testa

– grafički prikaz teorijskih i empirijskih frekvencija

b) Obrada visine pikova:

- proračun statistika
- proračun empirijske raspodele
- proračun teorijskih raspodela (eksponencijalna i Vejbulova)
- testiranje slaganja teorijskih i empirijskih raspodela pomoću testa Kramer-Mizesa
- grafički prikaz teorijskih raspodela i empirijskih tačaka na papiru Gumbelove verovatnoće

c) Formiranje raspodele godišnjih maksimuma:

- formiranje niza godišnjih maksimuma
- proračun empirijske raspodele
- proračun teorijskih raspodela
- testiranje slaganja teorijskih i empirijskih raspodela pomoću testa Kramer-Mizesa
- usvajanje najbolje kombinacije raspodela broja javljanja i visine pikova
- grafički prikaz teorijskih raspodela i empirijskih tačaka na papiru Gumbelove verovatnoće

Vežba 4: Linearna regresija (korelacija)

Za zadate nizove godišnjih padavina na dve kišomerne stanice (nizovi X i Y) uraditi sledeće:

- odrediti statistike nizova na dve stanice (srednje vrednosti, standardne devijacije, koeficijent korelacije);
- odrediti koeficijente linearne regresije;
- odrediti standardnu grešku ocene $S_{Y|X}$ (varijansu reziduala);
- odrediti pojaseve oko regresione linije od $\pm S_{Y|X}$ i $\pm 2S_{Y|X}$;
- odrediti interval poverenja regresione linije od 90%;
- grafički predstaviti model, empirijske tačke, pojaseve i interval poverenja;
- koristeći se dobijenim modelom, popuniti nedostajuće podatke u nizu Y.

Nastavnici:

doc. dr Zoran Radić
doc. dr Jasna Plavšić

PREGLED TEORIJE UZ VEŽBE

VEŽBA 1

Testiranje homogenosti i slučajnosti statističkih nizova

Pod homogenošću hidrološkog niza podrazumeva se da on potiče iz jedne populacije veličine koja se razmatra. Tipičan primer nehomogenog niza je niz godišnjih maksimuma proticaja među kojima su oni nastali usled jakih letnjih pljuskova i oni nastali usled prolećnog otapanja snega.

Slučajnost niza znači da elementi niza moraju biti međusobno nezavisni.

Testiranje homogenosti i slučajnosti niza obavlja se odgovarajućim statističkim testovima. U opštem slučaju procedura testiranja se sastoji od sledećih koraka:

- Formulisanje nulte hipoteze H_0 da nema razlike između dva niza podataka.
- Izbor statističkog testa.
- Izbor praga značajnosti α.
- Utvrđivanje regiona prihvatanja H_0 za tu vrstu testa i za usvojeno α.
- Proračun vrednosti statistike na kojoj se zasniva test. Ako je sračunata vrednost izvan regiona prihvatanja, odbaciti H_0.

Postoje dve vrste statističkih testova:

- parametarski i
- neparametarski.

Parametarski testovi se zasnivaju na testiranju neke karakteristike niza (srednja vrednost, varijansa) i uključuju određene pretpostavke o nizu:

a) podaci su međusobno nezavisni;
b) podaci potiču iz normalno raspoređenih populacija;
c) te populacije moraju imati istu varijansu.

Za neparametarske testove, koji ne koriste statistike niza, obično je dovoljna prva pretpostavka. Najčešće korišćeni parametarski testovi su z-test i t-test za testiranje srednje vrednosti i F-test za testiranje varijanse. Među neparametarskim testovima to su testovi Men-Vitni (Mann-Whitney) i Kolmogorov-Smirnov.

1 Testiranje homogenosti statističkih nizova

1.1 Parametarski testovi

1.1.1 Testiranje jednakosti dve srednje vrednosti sa poznatim varijansama: z-test

Ovaj test testira jednakost srednjih vrednosti dva uzorka. Koristi se se za veliki uzorak. Pretpostavlja se da dva uzorka potiču iz dve normalno raspoređene populacije, X_1: N(μ_1, σ_1) i X_2: N(μ_2, σ_2). Uzorak X_1, dužine N_1, ima srednju vrednost $\overline{x}_1$ i standardnu devijaciju S_1, dok uzorak X_2, dužine N_2 ima srednju vrednost $\overline{x}_2$ i standardnu devijaciju S_2. Nulta i alternativna hipoteza glase:

$$H_0: \quad \mu_1 = \mu_2$$
$$H_a: \quad \mu_1 \neq \mu_2$$

Ako se posmatra razlika srednjih vrednosti dva uzorka, ona takođe predstavlja normalno raspoređenu slučajnu promenljivu:

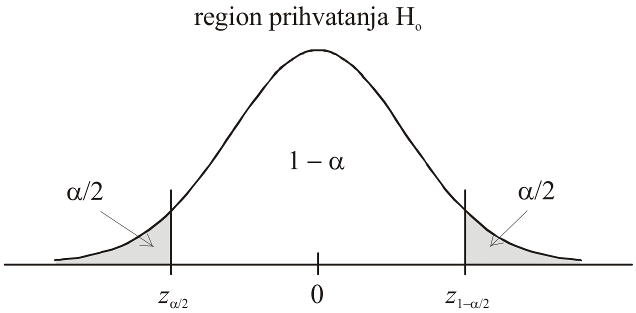
$$\overline{X}_1 - \overline{X}_2 : N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 / N_1 + \sigma_2^2 / N_2})$$

Pod uslovom da važi hipoteza H_0 ($\mu_1 - \mu_2 = 0$), standardizovana normalna promenljiva za razliku srednjih vrednosti uzoraka glasi:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_1^2 / N_1 + S_2^2 / N_2}}$$

Promenljiva z ima raspodelu N(0,1). Ovaj test je dvostrani test, jer vrednost z može biti i pozitivna i negativna. Za zadati prag značajnosti α, region prihvatanja hipoteze H_0 je (vidi sliku):

$$z_{\alpha/2} < z < z_{1-\alpha/2}$$



1.1.2 Testiranje jednakosti dve srednje vrednosti sa nepoznatim ali jednakim varijansama: t-test

Ovaj test je sličan z-testu, osim što se primenjuje za uzorke manjeg obima. Dodatna pretpostavka je da su varijanse dva uzorka jednake, tako da ovu hipotezu treba proveriti pre sprovođenja testa (pomoću Fišerovog F-testa za varijanse). Ocena jedinstvene varijanse σ za dva uzorka glasi:

$$S^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}$$

Razlika srednjih vrednosti dva uzorka $\bar{X}_1 - \bar{X}_2$ kao slučajna promenljiva ima srednju vrednost $\mu_1 - \mu_2$ (koja je, prema H_0, jednaka 0) i varijansu $\sigma\sqrt{1/N_1 + 1/N_2}$. Standardizovana promenljiva

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{1/N_1 + 1/N_2}}$$

prati Studentovu raspodelu sa $N_1 + N_2 - 2$ stepeni slobode (videti dodatak A). I ovde se radi o dvostranom testu, pa je region prihvatanja hipoteze H_0 definisan sa

$$t_{\alpha/2} < t < t_{1-\alpha/2}$$

1.1.3 Testiranje jednakosti varijansi dva uzorka: F-test

Nulta hipoteza ovog testa je da su varijanse dva niza jednake. Posmatra se statistika

$$F = \frac{S_1^2}{S_2^2}$$

u kojoj je S_1^2 veća od dve varijanse ($S_1^2 > S_2^2$). Ova statistika prati Fišerovu (Fisher) F-raspodelu sa $\nu_1 = N_1 - 1$ i $\nu_2 = N_2 - 1$ stepeni slobode (videti dodatak A). Ovaj test je jednostrani test, pa je region prihvatanja nulte hipoteze:

$$F < F_{1-\alpha,\nu_1,\nu_2}$$

1.2 Neparametarski testovi

Dva testa o kojima će ovde biti reči koriste se za ispitivanje istorodnosti raspodela dva poduzorka (dobijena podelom originalnog uzorka).

1.2.1 Men-Vitni test (Mann i Whitney)

Nulta hipoteza ovog testa je da je uzorak homogen. Dva poduzorka dužina N_1 i N_2 se spajaju i uređuju u rastući niz, pri čemu se označava pripadnost prvom odnosno drugom poduzorku. Označava se redosled (rang) članova zajedničkog niza. Sabiranjem rangova članova prvog poduzorka dobija se zbir R_1, a sabiranjem rangova drugog poduzorka dobija se zbir R_2. Zatim se računaju statistike

$$U_1 = N_1 N_2 + \frac{N_1(N_1+1)}{2} - R_1$$

$$U_2 = N_1 N_2 + \frac{N_2(N_2+1)}{2} - R_2$$

Između ovih statistika postoji veza:

$$U_1 + U_2 = N_1 N_2$$

Između U_1 i U_2 bira se manja vrednost:

$$U = \min\{U_1, U_2\}$$

Za N_1, $N_2 > 8$, statistika U približno prati normalnu raspodelu sa srednjom vrednošću i varijansom:

$$\mu_U = \frac{N_1 N_2}{2}, \quad \sigma_U^2 = \frac{N_1 N_2 (N_1 + N_2 + 1)}{12}$$

a statistika

$$u = \frac{U - \mu_U}{\sigma_U}$$

prati standardnu normalnu raspodelu N(0,1). Region prihvatanja nulte hipoteze za prag značajnosti α je:

$$u_{\alpha/2} < u < u_{1-\alpha/2}$$

1.2.2 Kolmogorov-Smirnov test za dva uzorka

Dva poduzorka se poređaju u rastući niz, a zatim se sračunaju njihove kumulativne relativne frekvencije $F_1(x_i)$ i $F_2(x_j)$. Na grafiku x-$F(x)$ posmatra se najveća razlika između ordinata kumulativnih relativnih frekvencija dva poduzorka:

$$D_{\max} = \max\left|F_1(x_i) - F_2(x_j)\right|$$

Za zadati prag značajnosti α, region prihvatanja nulte hipoteze je:

$$D_{\max} < D_0(\alpha, n)$$

gde je:

$$n = \frac{N_1 N_2}{N_1 + N_2}$$

Kritične vrednosti D_0 za ovaj test date su u tabeli u prilogu.

2 Testiranje slučajnosti statističkih nizova

U svim testovima slučajnostu nulta i alternativna hipoteza glase:

H_0: niz je slučajan (članovi niza su međusobno nezavisni)

H_a: niz nije slučajan (članovi niza su međusobno zavisni)

2.1 Test kvadrata uzastopnih razlika

Nulta hipoteza ovog testa je da se uzorak sastoji od međusobno nezavisnih podataka osmatranja iz normalno raspoređene populacije. Varijansa populacije σ^2 ocenjuje se na dva načina. Prva ocena je kvadrat nepristrasne ocene standardne devijacije:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Druga ocena koristi kvadrate razlika uzastopnih članova neuređenog uzorka:

$$\frac{d^2}{2} = \frac{1}{2(N-1)}\sum_{i=1}^{N-1}(x_{i+1} - x_i)^2$$

Ako se odnos ove dve ocene označi sa δ:

$$\delta = \frac{d^2/2}{S^2}$$

i pod pretpostavkom da važi nulta hipoteza, očekivanje i varijansa promenljive δ su:

$$\mu_\delta = 1, \quad \sigma_\delta^2 = \frac{N-2}{N^2-1}$$

Statistika

$$u = \frac{\delta - \mu_\delta}{\sigma_\delta} = \frac{\delta - 1}{\sqrt{\dfrac{N-2}{N^2-1}}}$$

prati standardizovanu normalnu raspodelu N(0,1). Region prihvatanja nulte hipoteze za prag značajnosti α je tada:

$$u_{\alpha/2} < u < u_{1-\alpha/2}$$

2.2 Vold-Volfovic test (Wald i Wolfowitz)

U ovom testu nulta hipoteza je da je uzorak sastavljen od nezavisnih članova i da je stacionaran. Za uzorak obima N, WW-test posmatra statistiku R:

$$R = \sum_{i=1}^{N-1} x_i x_{i+1}$$

U slučaju da su članovi niza nezavisni, statistika R prati normalnu raspodelu sa srednjom vrednošću i varijansom:

$$\overline{R} = \frac{s_1^2 - s_2}{N-1}$$

$$\sigma_R^2 = \frac{s_2^2 - s_4}{N-1} + \frac{s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s_2^2 - 2s_4}{(N-1)(N-2)} - \overline{R}^2$$

gde je $s_k = \sum_{i=1}^{N} x_i^k$. Veličina

$$u = \frac{R - \overline{R}}{\sigma_R}$$

prati standardnu normalnu raspodelu N(0,1). Region prihvatanja nulte hipoteze za prag značajnosti α je tada:

$$u_{\alpha/2} < u < u_{1-\alpha/2}$$

2.3 Bartletov test za koeficijent autokorelacije

Međusobna zavisnost članova nekog niza odražava se i na vrednosti koeficijenta autokorelacije. Koeficijent autokorelacije prvog reda je mera zavisnosti između uzastopnih članova niza. Analogno tome, koeficijent autokorelacije k-tog reda je mera zavisnosti između svakog k-tog člana niza. Red k se naziva i pomak.

Koeficijent autokorelacije reda k (k = 1, 2, ...) definiše se kao:

$$\rho_k = \frac{E[(X_i - \mu\ )(X_{i-k} - \mu_{X_{i-k}})]}{\sigma_{X_i} \sigma_{X_{i-k}}}$$

Ocena koeficijenta autokorelacije na osnovu uzorka je tada:

$$r_k = \frac{\frac{1}{N} \sum_{i=k+1}^{N} (x_i - \overline{x})(x_{i-k} - \overline{x})}{S^2}$$

U opštem slučaju vremenskih serija očekuje se da međuzavisnost članova niza opada sa pomakom k. Ukoliko je niz sastavljen od međusobno nezavisnih članova, svi koeficijenti autokorelacije jednaki su nuli. Kako se na osnovu uzorka ne mogu dobiti koeficijenti korelacije tačno jednaki nuli, već različiti od nule, potrebno je odgovarajućim testom utvrditi da li se oni značajno razlikuju od nule. Jedan od takvih testova je Bartletov test.

Nulta hipoteza Bartletovog testa glasi:

$$H_0: \quad \rho_k = 0, \quad k = 1,2,\ldots$$

Standardna devijacija k-tog koeficijenta korelacije glasi:

$$S_{r_k} = \sqrt{\frac{1}{N}\left(1 + 2\sum_{j=1}^{k} r_j^2\right)}, \quad k = 1,2,\ldots$$

Region prihvatanja nulte hipoteze za prag značajnosti α je:

$$z_{\alpha/2} \cdot S_{r_k} < r_k < z_{1-\alpha/2} \cdot S_{r_k}$$

gde su $z_{\alpha/2}$ i $z_{1-\alpha/2}$ standardizovane normalne promenljive sa funkcijom raspodele $\alpha/2$ odnosno $1 - \alpha/2$.

VEŽBA 2

Statistička analiza nizova godišnjih maksimuma

Metoda godišnjih maksimuma podrazumeva statističku analizu maksimalnih godišnjih vrednosti veličine koja se analizira (u ovoj vežbi proticaja). Veličina koja se analizira tretira se kao *slučajna promenljiva X* sa nekom funkcijom raspodele.

Funkcija raspodele definiše se kao verovatnoća da će slučajna promenljiva X biti manja ili jednaka od neke vrednosti x:

$$F(x) = P\{X \le x\}$$

Suprotna verovatnoća, tj. verovatnoća da će slučajna promenljiva imati vrednost veću od zadate vrednosti x naziva se *verovatnoća prevazilaženja*:

$$R(x) = P\{X > x\} = 1 - F(x)$$

Kada slučajna promenljiva X predstavlja godišnji maksimum, verovatnoća njene pojave iskazuje se i *povratnim periodom*, koji se izražava u godinama:

$$T(x) = \frac{1}{P\{X > x\}} = \frac{1}{R(x)} = \frac{1}{1 - F(x)}$$

Postupak statističke analize sastoji se od sledećih koraka:

1. Formiranje niza godišnjih maksimuma

Niz se formira od maksimalnih vrednosti proticaja za svaku godinu tokom perioda osmatranja. Pored osnovnog niza X, formira se i niz logaritama $Y = \log X$ (potreban za logaritmaske raspodele).

2. Proračun statistika osnovnog niza godišnjih maksimuma i logaritamskog niza

- srednja vrednost: $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$
- standardna devijacija: $$S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} x_i^2 - \frac{N}{N-1}\bar{x}^2}$$
- koeficijent varijacije: $$Cv_x = \frac{S_x}{\bar{x}}$$
- koeficijent asimetrije: $$Cs_x = \frac{N^2}{(N-1)(N-2)}\frac{1}{S_x^3}\sum_{i=1}^{N}(x_i - \bar{x})^3$$

Analogno se računaju: $\bar{y}$, S_y, C_{vy} i C_{sy}.

3. Provera postojanja izuzetaka u nizu

Za ovu proveru koristi se test Grubsa i Beka. U nizu postoji gornji izuzetak ako je

$$Y_{\max} > Y_g$$

odnosno donji izuzetak ako je

$$Y_{\min} < Y_d$$

Granice Yg i Yd određuju se kao:

$$Y_g = \bar{y} + K_{QN} S_y\,, \quad Y_d = \bar{y} - K_{QN} S_y$$

gde je:

$$K_{QN} = -3.62201 + 6.28446 N^{1/4} - 2.49835 N^{1/2} + 0.491436 N^{3/4} - 0.037911 N$$

4. Proračun empirijske raspodele

Empirijska raspodela podrazumeva dodeljivanje kompromisnih verovatnoća vrednostima iz niza. Pri tome se članovi niza uređuju po rastućem redosledu. Za kompromisne verovatnoće u ovoj vežbi koristi se Vejbulova formula:

$$F_e(x_i) = p_i = \frac{i}{N+1}$$

5. Proračun teorijskih raspodela

Proračun teorijskih raspodela podrazumeva da se odrede *kvantili*, odnosno vrednosti proticaja zadate verovatnoće pojave ili zadatog povratnog perioda. Da bi se to uradilo, neophodno je da se odrede parametri teorijskih raspodela. Proračun parametara teorijskih raspodela na osnovu statistika niza kao i proračun kvantila za različite teorijske raspodele prikazani su u dodatku A.

6. Testiranje saglasnosti teorijskih i empirijske raspodele

Provera slaganja teorijskih i empirijske raspodele može se sprovesti pomoću različitih testova, koji su prikazani u dodatku C.

Na osnovu rezultata testova usvaja se najpogodnija teorijska funkcija raspodele.

7. Interval poverenja teorijske funkcije raspodele

Kvantili određeni prema usvojenoj teorijskoj raspodeli predstavljaju *tačkastu ocenu* slučajne promenljive (ovde velikih voda) određene verovatnoće pojave. Pod tačkastom ocenom se podrazumeva da se daje jedna vrednost kao ocena. Pored ovakve ocene, može se odrediti i *intervalna ocena* slučajne promenljive, kojom se daje interval (opseg) vrednosti između kojih se tačan kvantil može naći. Takva ocena se naziva i *interval poverenja*. Pojam poverenja koji se vezuje za intervalne ocene kvantila predstavlja poverenje koje imamo u takvu ocenu. Tako se npr. govori od intervalu poverenja od 90% ili od 95%. Što je veće poverenje koje želimo da imamo u ocenu, granice u kojima se tačna veličina može naći biće šire.

Intervalu poverenja β odgovara *prag značajnosti* α:

$$\alpha = \frac{1-\beta}{2}$$

Na primer, za β = 90%, α = 5%.

Za veličinu slučajne promenljive zadatog povratnog perioda $X(T)$, gornja i donja granica intervala poverenja mogu se predstaviti izrazima:

– gornja granica:

$$X_g(T) = X(T) + |z_\alpha| S_{X(T)}$$

– donja granica:

$$X_d(T) = X(T) - |z_\alpha| S_{X(T)}$$

gde je $z_\alpha = z(\alpha) = -z(1-\alpha)$ standardizovana normalna promenljiva, a $S_{X(T)}$ standardna greška ocene $X(T)$. Proračun standardne greške ocene kvantila $S_{X(T)}$ prikazan je u dodatku D.

VEŽBA 3

Statistička analiza metodom pikova

Mana statističke analize godišnjih maksimuma je u tome što u takve nizove ulazi samo jedan događaj iz svake godine osmatranja, dok npr. druga ili treća najveća vrednost u toku godine mogu biti veće od maksimalnog događaja iz neke druge godine, a ipak ne ulaze u niz. Ovaj nedostatak se može prevazići formiranjem nizova *prekoračenja* ili *pikova*, u koje ulaze sve vrednosti iznad neke bazne vrednosti (odnosno ispod bazne vrednosti za nizove minimuma). Bazna vrednost se obično bira tako da u niz uđe bar jedan podatak iz svake godine. Vrednosti koje čine niz pikova moraju biti nezavisne; to znači da se ne mogu uzeti proticaji iz dva uzastopna dana, jer pripadaju istom meteorološkom događaju. Niz pikova se sastoji od različitog broja podataka za svaku godinu, zbog čega raspodela niza pikova nije direktno uporediva sa raspodelom odgovarajućeg niza godišnjih ekstrema.

Ulazne podatke za metodu pikova čini niz vrednosti strogo većih od izabranog praga ($x > x_B$). Ukupan broj godina osmatranja je N, a ukupan broj vrednosti iznad praga (pikova) je M. Statistička analiza sastoji se iz tri koraka:

(1) određivanje raspodele broja pikova u godini dana,
(2) određivanje raspodele samih pikova, i
(3) kombinacija prethodne dve raspodele u raspodelu godišnjih ekstrema.

1. Raspodela broja pikova u godini dana

Broj pikova u godini dana je diskretna slučajna promenljiva n koja može da uzme vrednosti $n = 0, 1, 2, ...$, pa shodno tome ima raspodelu verovatnoće:

$$P\{n = k\} = p_k, \quad k = 0,1,2,\dots$$

Ako se posmatra serija pikova za N godina, onda je $n_1, n_2, ..., n_N$ broj pikova u svakoj godini, a ukupan broj pikova tokom N godina je $M = n_1 + n_2 + ... + n_N$.

Srednja vrednost i standardna devijacija broja pikova su:

$$\overline{n} = \frac{1}{N}\sum_{i=1}^{N} n_i = \frac{M}{N}, \qquad S_n^2 = \frac{1}{N}\sum_{i=1}^{N}(n_i - \overline{n})^2$$

Ako se frekvencije vrednosti k ($k = 0, 1, 2, ...$) označe sa f_k, onda se statistike broja pikova mogu odrediti i na sledeći način:

$$\overline{n} = \frac{1}{N}\sum n_k f_k = \frac{M}{N}, \qquad S_n^2 = \frac{1}{N}\sum n_k^2 f_k - \overline{n}^2$$

Važna statistika diskretnih promenljivih je *indeks disperzije*:

$$I = \frac{S_n^2}{\overline{n}}$$

Diskretne teorijske raspodele koje se koriste za prilagođavanje broja javljanja pikova su najčešće Poasonova, binomna i negativna binomna raspodela (videti dodatak B). Kriterijum za izbor teorijske raspodele je vrednost indeksa disperzije:

$I < 1$: binomna raspodela
$I = 1$: Poasonova raspodela
$I > 1$: negativna binomna raspodela

U praktičnim proračunima Poasonova raspodela može se usvojiti ako je indeks disperzije približno jednak jedinici, tj. ako je $0.8 < I < 1.2$.

2. Visina pikova

Visina pikova se definiše kao kontinualna slučajna promenljiva $Z = X - x_B$. Raspodela visine pikova označava se sa $H(z)$:

$$H(z) = H(x - x_B) = P\{Z \le z\}$$

Za visinu pikova najčešće se koriste eksponencijalna i Vejbulova raspodela. Eksponencijalna raspodela ima jedan parametar (α), a Vejbulova dva (α i β). Eksponencijalna raspodela poseban slučaj Vejbulove raspodele za parametar β = 1. Parametri raspodela se određuju iz statistika uzorka:

- srednja vrednost: $\bar{z} = \frac{1}{M}\sum_{i=1}^{M} z_i$
- standardna devijacija: $S_z = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(z_i - \bar{z})^2}$
- koeficijent varijacije: $C_{vz} = \frac{S_z}{\bar{z}}$
- koeficijent asimetrije: $C_{sz} = \frac{M^2}{(M-1)(M-2)}\frac{1}{S_z^3}\sum_{i=1}^{M}(z_i - \bar{z})^3$

Za proračun teorijskih raspodela videti dodatak A. Empirijska raspodela se računa prema jednoj od poznatih formula za M članova niza pikova.

3. Godišnji maksimumi

Godišnji maksimumi predstavljaju slučajnu promenljivu X, koja se definiše kao najveća vrednost od svih pikova tokom godinu dana:

$$X = x_B + \max\{Z_j; j = 0,1,\ldots,n\}$$

Raspodela godišnjih maksimuma $F(x)$ je verovatnoća:

$$F(x) = P\{X \le x\}$$

i ona je definisana samo za vrednosti $x > x_B$. Opšti izraz za raspodelu $F(x)$ dobija se kombinovanjem raspodela broja pikova i visine pikova, uzimajući u obzir sve moguće kombinacije broja pikova u godini dana. Ako u godini ima $n = k$ pikova, onda su to pikovi $Z_1, Z_2, ..., Z_k$. Verovatnoća da je godišnji maksimum X, kao najveća vrednost od svih Z, manji ili jednak od x podrazumeva i da su svi ostali pikovi manji ili jednaki od $x - x_B$:

$$P\{X \le x \mid n = k\} = P\{Z_1 \le x - x_B, Z_2 \le x - x_B, \ldots, Z_k \le x - x_B\} =$$
$$= P\{Z_1 \le x - x_B\} \cdot P\{Z_2 \le x - x_B\} \cdot \ldots \cdot P\{Z_k \le x - x_B\} = [H(x - x_B)]^k$$

Ukoliko tokom godine nema nijednog pika ($n = 0$), to znači da su svi pikovi manji od bazne vrednosti pa i godišnji maksimum, pa je $P\{X \le x_B \mid n = 0\} = 1$. Kako je $x > x_B$, sledi i:

$$P\{X \le x \mid n = 0\} = 1$$

Kombinovanjem prethodne dve verovatnoće za sve moguće vrednosti broja pikova u godini dana, dobija se:

$$F(x) = P\{X \le x\} = P\{n = 0\} \cdot P\{X \le x \mid n = 0\} + P\{n = 1\} \cdot P\{X \le x \mid n = 1\} + P\{n = 2\} \cdot P\{X \le x \mid n = 2\} + \ldots =$$
$$= P\{n = 0\} \cdot 1 + P\{n = 1\} \cdot H(x - x_B) + P\{n = 2\} \cdot [H(x - x_B)]^2 + \ldots$$

odnosno

$$F(x) = \sum_{k=1}^{\infty} p_k [H(x - x_B)]^k$$

U zavisnosti od tipa raspodele broja pikova i tipa raspodele visine pikova, gornji izraz se svodi na:

- Poasonova raspodela broja pikova: $F(x) = \exp\{-\lambda(1-H)\}$
- Binomna raspodela broja pikova: $F(x) = [1 - p(1-H)]^a$
- Negativna binomna raspodela broja pikova: $F(x) = \left[1 + \left(\frac{1}{p} - 1\right)(1-H)\right]^{-b}$

gde su λ, p, a i b parametri odgovarajućih diskretnih raspodela, a u zavisnosti od raspodele visine pikova $1 - H$ je:

- Eksponencijalna raspodela visine pikova: $1 - H = \exp\{-(x - x_B)/\alpha\}$
- Vejbulova raspodela visine pikova: $1 - H = \exp\{-(x - x_B)^\beta / \alpha\}$

Inverzni izrazi, odnosno izrazi za kvantile zadate vrednosti funkcije raspodele F (odnosno povratnog perioda T) glase:

- Eksponencijalna raspodela visine pikova: $x = x_B + \alpha[-\ln(1-H)]$
- Vejbulova raspodela visine pikova: $x = x_B + \alpha[-\ln(1-H)]^{1/\beta}$

gde $1 - H$ zavisi od raspodele broja pikova:

- Poasonova raspodela broja pikova: $1 - H = -\frac{\ln F}{\lambda}$
- Binomna raspodela broja pikova: $1 - H = \frac{1 - F^{1/a}}{p}$
- Negativna binomna raspodela broja pikova: $1 - H = \frac{F^{-1/b} - 1}{1/p - 1}$

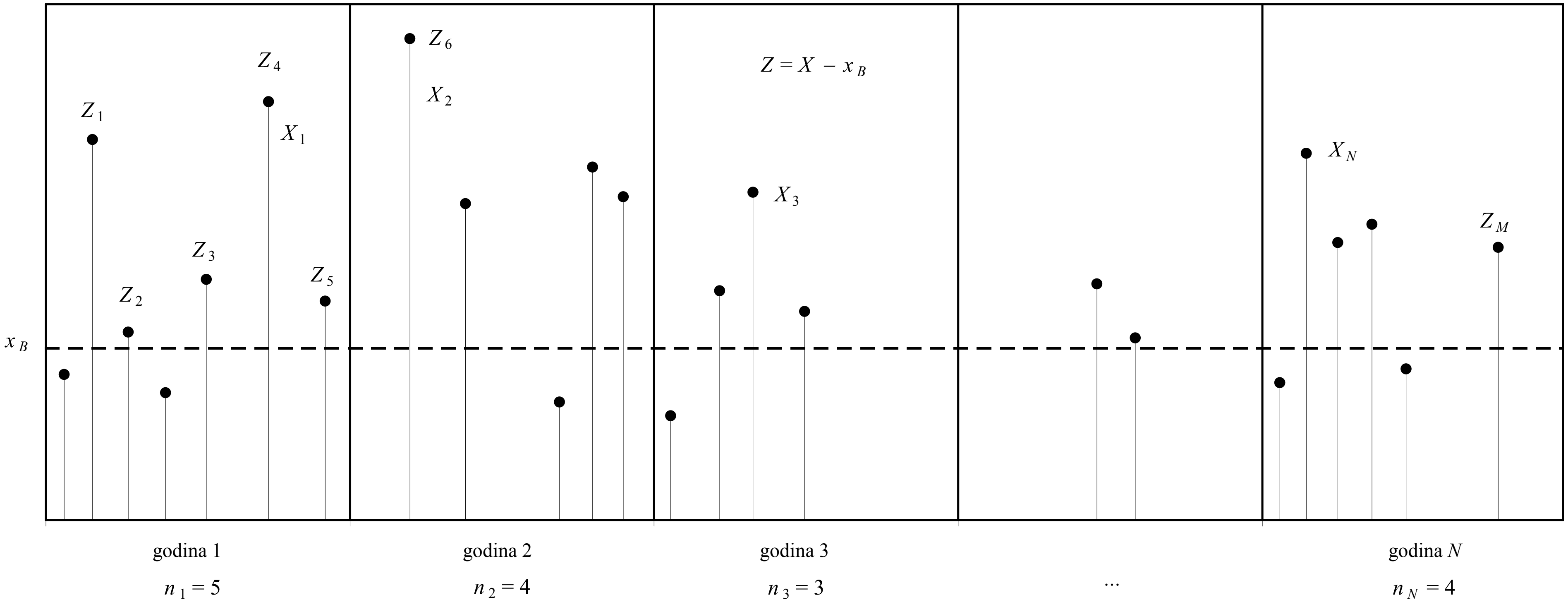
Funkcija raspodele godišnjih maksimuma $F(x)$ definisana je samo za vrednosti $x > x_B$, odnosno $z > 0$. Može se pokazati iz prethodnih izraza da se kvantili mogu sračunati samo za vrednosti funkcije raspodele:

- Poasonova raspodela broja pikova: $F > e^{-\lambda}$
- Binomna raspodela broja pikova: $F > (1-p)^a$
- Negativna binomna raspodela broja pikova: $F > p^b$

Empirijska raspodela godišnjih maksimuma određuje se proračunom kompromisnih verovatnoća za niz od N godina.

Metoda pikova

(1) broj pikova u godini dana n_i ($i = 1, 2, ..., N$)
(2) visina pikova Z_j ($j = 1, 2, ..., M$)
(3) godišnji maksimumi X_i ($i = 1, 2, ..., N$)



VEŽBA 4

Korelacija i regresija

Korelacija je veza između dve ili više promenljivih.

Regresija je postupak određivanja funkcionalne zavisnosti između dve ili više promenljivih, odnosno određivanja koeficijenata (parametara) u jednačinama modela za koji se pretpostavlja da važi između promenljivih. Najjednostavniji regresioni model je linearna regresija:

$$Y = aX + b$$

Koeficijenti a i b nazivaju se *regresionim koeficijentima*. Regresija može biti i nelinearna i višestruka (sa više od dve promenljive). Primeri regresije su:

– logaritamska regresija: $Y = a \ln X + b$
– eksponencijalna regresija: $Y = ae^{bX}$
– stepena regresija: $Y = aX^b$ Y = a Xb
– višestruka linearna regresija: $Y = a_1 X_1 + a_2 X_2 + \ldots + a_M X_M + b$

Nelinearna regresija se može svesti na linearnu odgovarajućim transformacijama. Tako se logaritamska regresija svodi na linearnu ako se stavi $X' = \ln X$, kod eksponencijalne ako se uvede smena $Y' = \ln Y$, a kod stepene ako se obe promenljive logaritmuju.

Koeficijent korelacije predstavlja meru povezanosti dve promenljive, i to njihove linearne povezanosti. On se definiše kao:

$$\rho = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Korelaciona matrica opisuje povezanost više od dve promenljive i predstavlja matricu koeficijenata korelacije između svake dve promeljive.

1 Regresioni model i odstupanja od modela

Ako je Y slučajna promenljiva, a $Y_m(X)$ regresioni model zavisnosti između X i Y, tada se slučajna promenljiva Y može prikazati u obliku

$$Y = Y_m(X) + \varepsilon$$

gde se Y naziva zavisnom promenljivom, a X nezavisnom promenljivom. Član ε predstavlja odstupanja između vrednosti slučajne promenljive Y i modela Y_m:

$$\varepsilon = Y - Y_m$$

Ovo odstupanje se naziva i greška ili rezidual. Odstupanje ε je slučajna promenljiva koja prati normalnu raspodelu sa srednjom vrednošću 0 i standardnom devijacijom σ_ε, odnosno varijansom σ_ε^2:

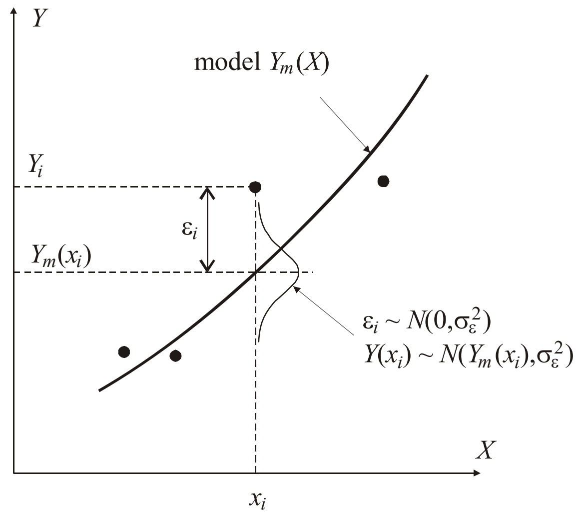
$$E[\varepsilon] = \mu_\varepsilon = 0\,, \qquad D[\varepsilon] = \sigma_\varepsilon^2$$

To znači da promenljiva Y za fiksiranu vrednost x ima srednju vrednost jednaku vrednosti modela i da odstupa od modela sa standardnom devijacijom σ_ε:

$$E[Y \mid X = x] = E[Y_m(x) + \varepsilon] = Y_m(x) + E[\varepsilon] = Y_m(x)$$

$$D[Y \mid X = x] = D[Y_m(x) + \varepsilon] = D[\varepsilon] = \sigma_\varepsilon^2$$

Drugim rečima, regresioni model je očekivano mesto svih eksperimentalnih tačaka, a pojedine eksperimentalne tačke odstupaju od modela sa standardnom devijacijom σ_ε.



Slika 1. Regresioni model i odstupanja eksperimentalnih tačaka.

2 Ocena parametara modela

Nepoznati parametri u modelu (regresioni koeficijenti) najčešće se određuju *metodom najmanjih kvadrata*. Suština ove metode je u tome da se minimizira suma kvadrata odstupanja osmotrenih vrednosti od modela:

$$SS = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} [Y_i - Y_m(x_i)]^2 \tag{1}$$

Ako su a_1, a_2, ..., a_K parametri modela, rešavanjem sistema jednačina

$$\frac{\partial SS}{\partial a_1} = 0, \quad \frac{\partial SS}{\partial a_2} = 0, \quad \ldots, \quad \frac{\partial SS}{\partial a_K} = 0$$

dolazi se do ocene parametara koji će dati najmanju sumu kvadrata grešaka.

2.1 Ocena parametara linearne regresije

U slučaju linearne regresije

$$Y_m = aX + b$$

jednačina (1) se svodi na:

$$SS = \sum_{i=1}^{N} [y_i - ax_i - b)]^2$$

pa se formira sledeći sistem od dve jednačine sa nepoznatim parametrima a i b:

$$\frac{\partial SS}{\partial a} = \sum_{i=1}^{N} 2(y_i - ax_i - b)(-x_i) = a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i y_i = 0 \tag{2}$$

$$\frac{\partial SS}{\partial b} = \sum_{i=1}^{N} 2(y_i - ax_i - b)(-1) = a\sum_{i=1}^{N} x_i + bN - \sum_{i=1}^{N} y_i = 0 \tag{3}$$

Iz jednačine (3) odmah sledi:

$$\hat{b} = \overline{y} - \hat{a}\overline{x}$$

dok se preuređivanjem jednačine (2) dobija:

$$\hat{a} = r\frac{S_y}{S_x}$$

gde je r ocena koeficijenta korelacije dobijena iz uzorka (videti odeljak o linearnoj regresiji). Sa $\hat{a}$ i $\hat{b}$ su označene ocene parametara na osnovu uzorka, dok su a i b parametri koji važe za populaciju.

3 Ocena grešaka odnosno pogodnosti modela

3.1 Standardna greška ocene

Nepristrasna ocena standardne devijacije odstupanja naziva se *standardna greška ocene* i u opštem slučaju jednaka je:

$$\hat{\sigma}_\varepsilon^2 = S_{Y|X}^2 = \frac{1}{N-M-1}\sum_{i=1}^{N}[Y_i - Y_m(x_i)]^2 = \frac{1}{N-M-1}\sum_{i=1}^{N}\varepsilon_i^2 \qquad (4)$$

gde je M broj nezavisnih promenljivih u regresionom modelu. Kod proste regresije tipa $Y = f(X)$, $M = 1$, pa sledi:

$$\hat{\sigma}_\varepsilon^2 = S_{Y|X}^2 = \frac{1}{N-2}\sum_{i=1}^{N}[Y_i - Y_m(x_i)]^2 = \frac{1}{N-2}\sum_{i=1}^{N}\varepsilon_i^2$$

3.2 Objašnjena i neobjašnjena varijansa

Ukupna varijacija slučajne promenljive Y definiše se kao suma kvadrata odstupanja Y od srednje vrednosti $\overline{Y}$. Može se pokazati da važi sledeća jednakost (videti sliku 2):

$$\sum(Y-\overline{Y})^2 = \sum(Y-Y_m)^2 + \sum(Y_m-\overline{Y})^2 \qquad (5)$$

Prvi izraz s desne strane predstavlja sumu kvadrata reziduala i naziva se *neobjašnjena varijansa*, a drugi izraz predstavlja sumu kvadrata odstupanja modela od srednje vrednosti ili *objašnjenu varijansu*, pa se kao mera pogodnosti modela može sračunati procenat objašnjene varijanse.

3.3 Korelacioni odnos

Odnos objašnjene i ukupne varijanse zove se korelacioni odnos (ili koeficijent determinacije):

$$R^2 = \frac{\sum(Y_m-\overline{Y})^2}{\sum(Y-\overline{Y})^2}$$

i praktično pokazuje procenat objašnjene varijanse pomoću modela.

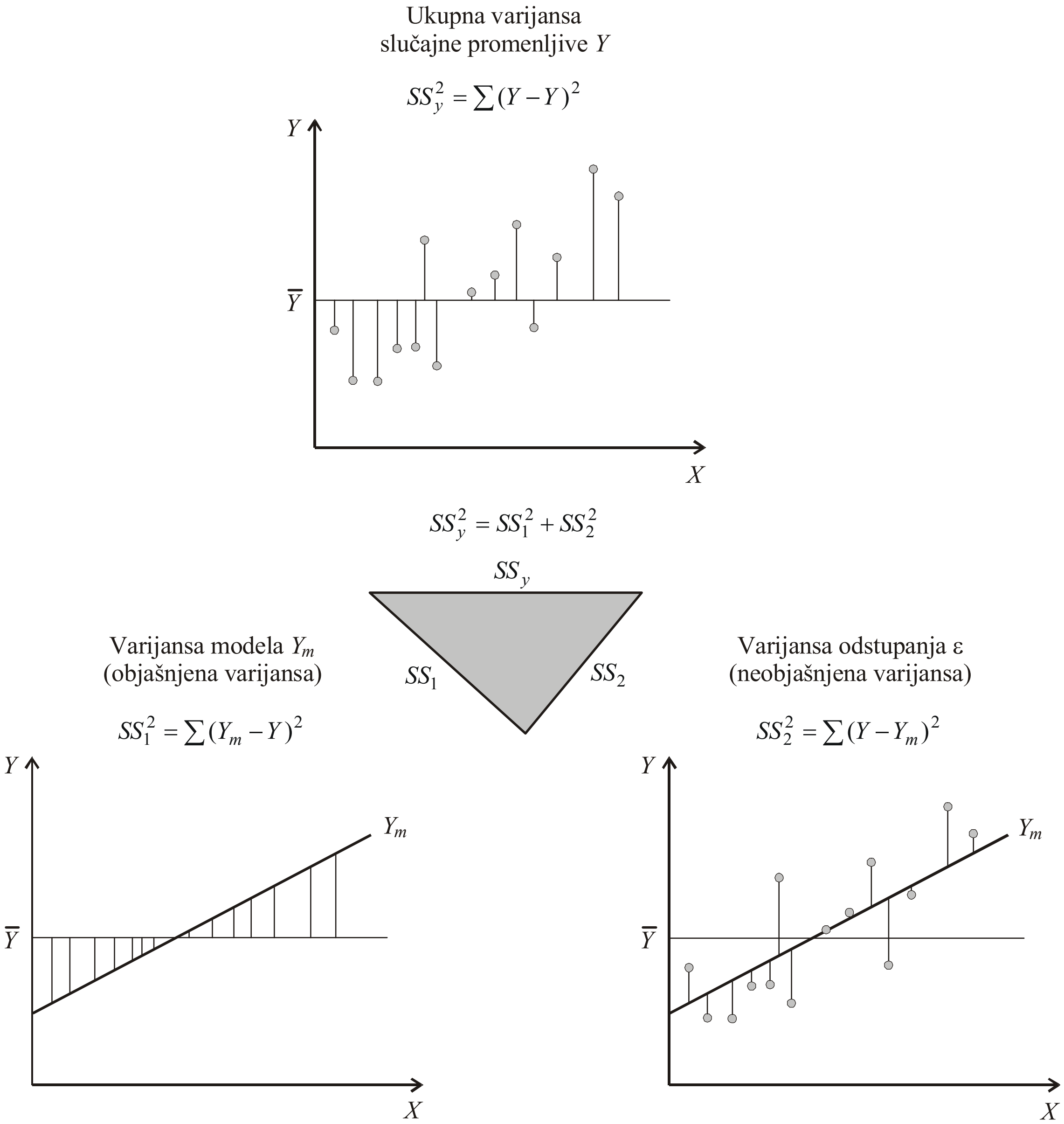
3.4 Veza između korelacionog odnosa i standardne greške

Na osnovu jednačine (5) sledi:

$$R^2 = \frac{\sum(Y_m-\overline{Y})^2}{\sum(Y-\overline{Y})^2} = \frac{\sum(Y-\overline{Y})^2 - \sum(Y-Y_m)^2}{\sum(Y-\overline{Y})^2} = 1 - \frac{(N-M-1)S_{Y|X}^2}{(N-1)S_Y^2} \qquad (6)$$

I standardna greška ocene se može izraziti preko korelacionog odnosa:

$$S_{Y|X}^2 = \frac{N-1}{N-M-1}S_y^2(1-R^2) \qquad (7)$$



Slika 2. Objašnjena i neobjašnjena varijansa u linearnoj regresiji.

Linearna regresija

Linearna regresija predstavlja najjednostavniji model veze između dve promenljive:

$$Y_m = aX + b$$

Već je pokazano da se koeficijenti a i b određuju metodom najmanjih kvadrata na osnovu statistika nizova X i Y:

$$\hat{a} = r\frac{S_y}{S_x}, \qquad \hat{b} = \bar{y} - \hat{a}\bar{x} \tag{8}$$

gde je r koeficijent korelacije određen iz uzorka po poznatoj formuli:

$$r = \frac{C_{xy}}{S_x S_y} = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y}$$

Odstupanja u linearnoj regresiji su jednaka:

$$\varepsilon_i = y_i - Y_m(x_i) = y_i - (ax_i + b) = y_i - \bar{y} - r\frac{S_y}{S_x}(x_i - \bar{x})$$

a suma njihovih kvadrata iznosi:

$$\sum_{i=1}^{N}\varepsilon_i^2 = \sum_{i=1}^{N}(y_i - \bar{y})^2 - 2r\frac{S_y}{S_x}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) + r^2\frac{S_y^2}{S_x^2}\sum_{i=1}^{N}(x_i - \bar{x})^2 =$$

$$= (N-1)S_y^2 - 2r\frac{S_y}{S_x}\cdot(N-1)rS_xS_y + r^2\frac{S_y^2}{S_x^2}\cdot(N-1)S_x^2 =$$

$$= (N-1)(1-r^2)S_y^2$$

Standardna greška ocene, na osnovu jednačine (4), tada je:

$$S_{Y|X}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\varepsilon_i^2 = \frac{N-1}{N-2}(1-r^2)S_y^2$$

odnosno

$$S_{Y|X} = S_y\sqrt{\frac{N-1}{N-2}(1-r^2)}$$

Koeficijent determinacije je jednak koeficijentu korelacije (iz izraza (6) za $M = 1$):

$$R^2 = 1 - \frac{(N-2)S_{Y|X}^2}{(N-1)S_y^2} = 1 - \frac{(N-1)(1-r^2)S_y^2}{(N-1)S_y^2} = r^2$$

Dakle, r^2 predstavlja procenat objašnjene varijanse linearnim regresionim modelom.

4.1 Intervali poverenja u linearnoj regresiji

Linearni regresioni model definisan je parametrima a i b. Ocene ovih parametara mogu se dobiti pomoću izraza (5), pa je ocena modela:

$$\hat{Y}_m = \hat{a}X + \hat{b} \tag{9}$$

Međutim, ocene $\hat{a}$ i $\hat{b}$ su dobijene na osnovu uzorka, tako da ocena modela (9) ne mora da odgovara tačnom modelu linearne veze između Y i X (pod tačnim modelom podrazumeva se model koji važi za populaciju). Označimo tačan model na sledeći način:

$$Y_m = aX + b \tag{10}$$

gde su a i b tačne vrednosti parametara.

Sada se mogu definisati različiti intervali poverenja za linearnu regresiju.

(a) Interval poverenja zavisno promenljive Y za fiksiranu vrednost nezavisno promenljive X ($X = x_0$) u odnosu na tačan model

Već je rečeno da odstupanja slučajne promenljive Y od tačnog modela prate normalnu raspodelu sa srednjom vrednošću 0 i varijansom σ_ε^2, :

$$\varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Stoga i slučajna promenljiva Y za fiksirano X prati normalnu raspodelu sa srednjom vrednošću $Y_m(x_0)$ i varijansom σ_ε^2:

$$\{Y \mid X = x_0\} \sim N(Y_m(x_0), \sigma_\varepsilon^2)$$

Na osnovu ovoga sledi da interval poverenja od β% (odnosno za prag poverenja $\alpha = 1 - \beta$) za vrednost slučajne promenljive Y za fiksiranu vrednost nezavisno promenljive X ($X = x_0$) u odnosu na tačan model glasi:

$$Y_m(x_0) - |z_{\alpha/2}|\sigma_\varepsilon < Y(x_0) < Y_m(x_0) + |z_{\alpha/2}|\sigma_\varepsilon$$

gde je $z_{\alpha/2}$ standardizovana normalna promenljiva za koju je $F(z_{\alpha/2}) = \alpha/2$.

U praktičnim proračunima σ_ε se zamenjuje svojom ocenom, tj. sa $S_{Y|X}$, a tačan model ocenjenim modelom, tako da se prethodni izraz svodi na:

$$\hat{Y}_m(x_0) - |z_{\alpha/2}| S_{Y|X} < Y(x_0) < \hat{Y}_m(x_0) + |z_{\alpha/2}| S_{Y|X}$$

Tako npr. za $z = \pm 1$ dobija se pojas od $\pm S_{Y|X}$ oko modela u koji, prema normalnoj raspodeli, ulazi oko 68% tačaka, dok se za $z = \pm 2$ dobija pojas od $\pm 2S_{Y|X}$ oko modela u koji ulazi oko 95% tačaka.

(b) Interval poverenja regresionog modela Y_m za fiksiranu vrednost nezavisno promenljive X ($X = x_0$)

Može se pokazati da je varijansa ocenjenog modela (6) za fiksiranu vrednost nezavisno promenljive $X = x_0$ jednaka:

$$\mathrm{var}[\hat{Y}_m(x_0)] = \sigma_\varepsilon^2 \left[\frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right]$$

U proračunima gornji izraz se svodi na:

$$S_{\hat{Y}_m} = S_{Y|X} \sqrt{\frac{1}{N} + \frac{(x_0 - \bar{x})^2}{(N-1)S_x^2}}$$

Interval poverenja od β% (odnosno za prag poverenja $\alpha = 1 - \beta$) regresionog modela Y_m za fiksiranu vrednost nezavisno promenljive X ($X = x_0$) glasi:

$$\hat{Y}_m(x_0) - |t_{\alpha/2}| S_{\hat{Y}_m} < Y_m(x_0) < \hat{Y}_m(x_0) + |t_{\alpha/2}| S_{\hat{Y}_m}$$

gde je $t_{\alpha/2}$ Studentova promenljiva sa $N - 2$ stepeni slobode.

(c) Interval poverenja prognozirane vrednosti slučajne promenljive Y za fiksiranu vrednost nezavisno promenljive X ($X = x_0$)

Prognoza vrednosti slučajne promenljive Y uz pomoć regresionog modela za zadatu vrednost slučajne promenljive $X = x_0$ daje:

$$Y(x_0) = \hat{Y}_m(x_0) + \varepsilon_0$$

Pošto se radi o prognozi buduće vrednosti Y, model $\hat{Y}_m$ (koji je već ustanovljen) i greška ε_0 su međusobno nezavisni, pa je tada:

$$\mathrm{var}[Y(x_0)] = \mathrm{var}[\hat{Y}_m(x_0) + \varepsilon_0] = \mathrm{var}[\hat{Y}_m(x_0)] + \mathrm{var}[\varepsilon_0] =$$

$$= \sigma_\varepsilon^2 \left[\frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \right] + \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \right]$$

U proračunima gornji izraz se svodi na:

$$S_{Y(x_0)} = S_{Y|X} \sqrt{1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{(N-1)S_x^2}}$$

Interval poverenja od β% (odnosno za prag poverenja $\alpha = 1 - \beta$) regresionog modela Y_m za fiksiranu vrednost nezavisno promenljive X ($X = x_0$) glasi:

$$\hat{Y}_m(x_0) - |t_{\alpha/2}| S_{Y(x_0)} < Y_m(x_0) < \hat{Y}_m(x_0) + |t_{\alpha/2}| S_{Y(x_0)}$$

gde je $t_{\alpha/2}$ Studentova promenljiva sa $N - 2$ stepeni slobode.

DODATAK A

Teorijske raspodele u hidrologiji za kontinualne slučajne promenljive

Date su teorijske funkcije raspodele koje se najčešće koriste u hidrologiji. Pored definicija raspodela, dati su izrazi za određivanje parametara po metodi momenata i funkcije iz programa EXCEL kojima se mogu računati vrednosti funkcija raspodele i kvantila.

Normalna raspodela

Funkcija raspodele:

$$F(x) = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{(x-a)^2}{2b^2}\right\} dx$$

Funkcija standardizovane normalne raspodele:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{z^2}{2}\right\} dz$$

gde je:

$$z = \frac{x-a}{b}$$

Parametri:

$$a = \bar{x}\,, \quad b = S_x$$

Funkcija raspodele u Excelu:

F = NORMDIST(x,a,b,TRUE)   ili
F = NORMSDIST(z)

Inverzna funkcija raspodele u Excelu:

x = NORMINV(F,a,b)   ili
x = a + b*NORMSINV(F)

Log-normalna raspodela

Funkcija raspodele:

$$F(x) = \frac{1}{b\sqrt{2\pi}} \int_{0}^{x} \frac{1}{x} \exp\left\{-\frac{(\ln x-a)^2}{2b^2}\right\} dx$$

Sa smenom:

$$y = \ln x$$

log-normalna raspodela se svodin na normalnu raspodelu promenljive Y. Sa smenom

$$z = \frac{y-a}{b}$$

log-normalna raspodela se svodi na standardizovanu normalnu raspodelu.

Parametri na osnovu metode momenata:

$$a = \ln \bar{x} - \frac{b^2}{2}, \quad b^2 = \ln(1 + C_{vx}^2)$$

Parametri na osnovu metode maksimalne verodostojnosti:

$$a = \bar{y}, \quad b = S_y$$

Funkcija raspodele u Excelu:

F = LOGNORMDIST(x,a,b) ili
F = NORMSDIST(z)

Inverzna funkcija raspodele u Excelu:

x = LOGINV(F,a,b) ili
x = EXP(a + b*NORMSINV(F))

Napomena: U lognormalnoj raspodeli može da se radi i sa dekadnim logaritmima umesto sa prirodnim (smena $y = \log x$). Medjutim, Excelove funkcije LOGNORMDIST i LOGINV rade samo sa prirodnim logaritmima, tako da kada se radi sa transformacijom preko dekadnih logaritama treba koristiti funkcije za standardizovanu normalnu raspodelu NORMSDIST i NORMSINV.

Gumbelova raspodela

Funkcija raspodele:

$$F(x) = \exp\left\{ -\exp\left[-\frac{x-u}{\alpha} \right] \right\}$$

Inverzna funkcija raspodele:

$$x = u + \alpha[-\ln(-\ln F)]$$

Smenom

$$y = \frac{x-u}{\alpha}$$

dolazi se do funkcije standardizovane Gumbelove raspodele:

$$F(x) = \exp\{-\exp[-y]\}$$

čija je inverzna forma

$$y = -\ln(-\ln F)$$

Parametri:

$$u = \bar{x} - 0.45 S_x, \quad \alpha = 0.78 S_x$$

Pirson 3 raspodela

Funkcija raspodele:

$$F(x) = \int_c^x \frac{1}{b\Gamma(\alpha)} \left(\frac{x-c}{b} \right)^{a-1} \exp\left\{ -\frac{x-c}{b} \right\} dx$$

Ako se Pirson 3 raspodela i njena inverzna funkcija odredjuju iz tablica, tada se koristi faktor frekvencije

$$k_P(T) = \frac{x(T) - \bar{x}}{S_x}$$

koji je tabulisan u zavisnosti od vrednosti funkcije raspodele i koeficijenta asimetrije.

Parametri:

$$a = \frac{4}{C_{sx}^2}, \quad b = \frac{S_x C_{sx}}{2}, \quad c = \bar{x} - ab$$

Funkcija raspodele u Excelu: u Excelu figuriše samo dvoparametarska gama raspodela GAMMADIST(x,a,b,kum), čija je primena ograničena samo za pozitivne vrednosti koeficijenta asimetrije C_{sx} i parametra b. Ovu funkciju onda treba koristiti na sledeći način:

– za $C_{sx} > 0$ i $b > 0$:

F = GAMMADIST((x-c)/b,a,1,TRUE)

– za $C_{sx} < 0$ i $b < 0$:

F = 1 – GAMMADIST((x-c)/b,a,1,TRUE)

Inverzna funkcija raspodele u Excelu:

– za $C_{sx} > 0$ i $b > 0$:

x = c + b * GAMMAINV(F,a,1)

– za $C_{sx} < 0$ i $b < 0$:

x = c + b * GAMMAINV(1-F,a,1)

Log-Pirson 3 raspodela

Log-Pirson 3 raspodela se koristi analogno Pirson 3 raspodeli uz transformaciju:

$$y = \ln x \text{ ili } y = \log x$$

Eksponencijalna raspodela

Funkcija raspodele:

$$F(x) = 1 - \exp\left\{-\frac{x}{a}\right\}$$

Inverzna funkcija raspodele:

$$x = -a \ln(1 - F)$$

Parametri:

$$a = \bar{x}$$

Funkcija raspodele u Excelu:

F = EXPONDIST(x,a,TRUE)

Dvoparametarska Vejbulova raspodela

Funkcija raspodele:

$$F(x) = 1 - \exp\left\{-\left(\frac{x}{a}\right)^b\right\}$$

Inverzna funkcija raspodele:

$$x = a[-\ln(1-F)]^{1/b}$$

Parametri: parametar b se računa numerički iz izraza

$$f(b) = \frac{\Gamma(1+2/b)}{\Gamma^2(1+1/b)} = 1 + C_{vx}^2$$

a parametar a kao:

$$a = \frac{\bar{x}}{\Gamma(1+1/b)}$$

Gama funkcija u Excelu:

$\Gamma(x)$ = EXP(GAMMALN(x))

Funkcija raspodele u Excelu:

F = WEIBULL(x,b,a,TRUE)

Studentova t-raspodela

Funkcija raspodele:

$$F(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{t} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt$$

Parametri:

ν = broj stepeni slobode = $N-1$

Kada $\nu \to \infty$, Studentova raspodela teži normalnoj raspodeli.

Funkcija raspodele u Excelu:

– za $t > 0$:

F = 1 – TDIST(t, ν, 1)

– za $t < 0$:

F = TDIST(| t |, ν, 1)

Inverzna funkcija raspodele u Excelu:

t = TINV(2(1-F), ν)

Za potrebe proračuna kritičnih vrednosti $t_{\alpha/2}$ i $t_{1-\alpha/2}$ u okviru t-testa, gornji izraz se svodi na:

$t_{1-\alpha/2}$ = TINV(α, ν)

dok je $t_{\alpha/2} = -t_{1-\alpha/2}$ zbog simetrije ove raspodele.

Fišerova F-raspodela

Promenljiva F u Fišerovoj raspodeli označiće se sa x radi preglednije notacije.

Funkcija raspodele:

$$F(x) = \nu_1^{\nu_1/2}\nu_2^{\nu_2/2}\,\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \int_0^x \frac{x^{(\nu_1-2)/2}}{(\nu_1 x+\nu_2)^{(\nu_1+\nu_2)/2}}\,dx$$

Parametri:

ν_1 = broj stepeni slobode brojioca
ν_2 = broj stepeni slobode imenioca

Funkcija raspodele u Excelu:

F = 1 – FDIST(x, ν_1, ν_2)

Inverzna funkcija raspodele u Excelu:

x = FINV(1-F, ν_1, ν_2)

Za potrebe proračuna kritične vrednosti statistike F-testa za zadati prag značajnosti α, gornji izraz se svodi na:

x = FINV(α, ν_1, ν_2)

Hi-kvadrat raspodela

Funkcija raspodele:

$$F(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^x x^{\nu/2-1} e^{-x/2} dx$$

Parametri:

ν = broj stepeni slobode

Funkcija raspodele u Excelu:

F = 1 – CHIDIST(x, ν)

Inverzna funkcija raspodele u Excelu:

x = CHIINV(1-F, ν)

Za potrebe proračuna kritičnih vrednosti χ^2 u okviru hi-kvadrat testa za zadati prag značajnosti α, gornji izraz se svodi na:

$\chi^2_{\nu,1-\alpha}$ = CHIINV(α, ν)

DODATAK B

Teorijske raspodele u hidrologiji za diskretne slučajne promenljive

Date su teorijske raspodele diskretnih slučajnih promenljivih koje se najčešće koriste u hidrologiji. Pored definicija raspodela, dati su izrazi za odredjivanje parametara i funkcije iz programa EXCEL kojima se mogu računati vrednosti vervatnoća po ovim raspodelama.

Poasonova raspodela

Zakon raspodele:

$$p_i = P\{X = i\} = \frac{\lambda^i e^{-\lambda}}{i!}$$

Svojstva raspodele:

$$EX = DX = \lambda$$

$$I = \frac{DX}{EX} = 1$$

Parametar:

$$\lambda = \bar{x}$$

Zakon raspodele u Excelu:

p = POISSON(i, lambda, FALSE)

Binomna raspodela

Zakon raspodele:

$$p_i = P\{X = i\} = \binom{a}{i} p^i (1-p)^{a-1}$$

Svojstva raspodele:

$$EX = ap$$

$$DX = ap(1-p)$$

$$I = 1 - p$$

Parametri:

– ako je poznat parametar *a* (broj opita):

$$p = \frac{\bar{x}}{a}$$

– ako je parametar *a* nepoznat najpre se računa:

$$a = \frac{\bar{x}}{p} = \frac{\bar{x}}{1-I}$$

a zatim se tako sračunat parametar *a* zaokružuje na celobrojnu vrednost i računa parametar *p*:

$$p = \frac{\bar{x}}{a}$$

Zakon raspodele u Excelu:

p = BINOMDIST(i, a, p, FALSE)

Negativna binomna raspodela

Zakon raspodele:

$$p_i = P\{X = i\} = \binom{b-1+i}{b-1} p^b (1-p)^i$$

Svojstva raspodele:

$$EX = \frac{b(1-p)}{p}, \qquad DX = \frac{b(1-p)}{p^2}$$

$$I = \frac{1}{p}$$

Parametri: najpre se računa:

$$b = \frac{p\overline{x}}{1-p} = \frac{\overline{x}}{I-1}$$

a zatim se tako sračunat parametar *b* zaokružuje na celobrojnu vrednost i računa parametar *p*:

$$p = \frac{b}{\overline{x}+b}$$

Zakon raspodele u Excelu:

p = NEGBINOMDIST(i, b, p)

DODATAK C

Testiranje saglasnosti empirijske i teorijske funkcije raspodele

U testovima saglasnosti porede se empirijske (osmotrene) i teorijske frekvencije slučajne promenljive koje pripadaju određenim klasama, ili se porede empirijske i teorijske kumulativne relativne frevencije.

Nulta hipoteza u testovima saglasnosti jeste da su empirijska i teorijska raspodela saglasne, a alternativna hipoteza je da ove dve raspodele nisu saglasne.

Hi-kvadrat test

Test je primenljiv i na diskretne i na kontinualne slučajne promenljive. Razmatra se statistika

$$\chi^2 = \sum_{k=1}^{K} \frac{(f_{e,k} - f_{t,k})^2}{f_{t,k}}$$

gde su $f_{e,k}$ empirijske frekvencije, $f_{t,k}$ teorijske frekvencije, a K broj klasa vrednosti slučajne promenljive. Statistika χ^2 prati χ^2-raspodelu sa $\nu = K - \rho - 1$ stepeni slobode (ρ je broj parametara u teorijskoj raspodeli). Smatra se da je χ^2-test pouzdan ako u svakoj klasi ima bar 5 elemenata.

Oblast prihvatanja nulte hipoteze (da su raspodele saglasne) za usvojen prag značajnosti α je:

$$\chi^2 < \chi^2_{\nu,1-\alpha}$$

Naglašava se da je χ^2-test jednostrani test (statistika χ^2 može da ima samo pozitivne vrednosti).

Tablica χ^2-raspodele data je u prilogu.

Test Kolmogorova-Smirnova

Ovaj neparametarski test je primenljiv samo na kontinualne slučajne promenljive. Nulta hipoteza da se teorijska raspodela ne razlikuje od empirijske formuliše se kao:

$$H_o: \quad \lim_{N \to \infty} P\{F_e(x) - F_t(x)\} = 1$$

Kriterijum testa je statistika koja predstavlja maksimalnu razliku između teorijske i empirijske raspodele:

$$D_{\max} = \max_x |F_e(x) - F_t(x)|$$

Za velike vrednosti N, statistika

$$\lambda = D_{\max}\sqrt{N}$$

ima sledeću funkciju raspodele:

$$\lim_{N \to \infty} P\{\lambda \le u\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2u^2}$$

Pomoću gornjeg izraza mogu se računati kritične vrednosti statistike $D_{\max}$ za velike uzorke (npr. za $N > 35$). Za manje uzorke, kritične vrednosti ove statistike daju se u tablici u prilogu.

Test Kramera-Mizesa

Test Kramera-Mizesa (Cramer-von Mises) takođe poredi empirijsku i teorijsku raspodelu kroz statistiku:

$$\omega^2 = \int_{-\infty}^{\infty} [F_e(x) - F_t(x)]^2 \, dF_t(x)$$

Ako je niz uređen u rastućem redosledu i $F_e(x_i) = i/N$, gornja statistika postaje:

$$\omega^2 = \frac{1}{12N^2} + \frac{1}{N} \sum_{i=1}^{N} \left[F_t(x_i) - \frac{i - 0.5}{N} \right]^2$$

Za velike uzorke ($N > 40$) statistika $N\omega^2$ ima definisanu raspodelu koja ne zavisi od dužine uzorka N i može se koristiti za određivanje kritičnih vrednosti. Tablica sa kritičnim vrednostima $N\omega^2$ data je u prilogu.

Test Andersona-Darlinga

Ovaj test daje veću težinu graničnim oblastima funkcije raspodele (malim i velikim verovatnoćama). Posmatra se statistika:

$$A^2 = \int_{-\infty}^{\infty} \frac{[F_e(x) - F_t(x)]^2}{F_t(x)[1 - F_t(x)]} \, dF_t(x)$$

Može se pokazati da je gornji izraz ekvivalentan izrazu:

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^{N} (2i - 1) \ln F_t(x_i)[1 - F_t(x_{N-i+1})]$$

gde su x_i i x_{n-i+1} elementi niza uređenog u rastući redosled. Smatra se su kritične vrednosti statistike A^2 za pragove značajnosti 5% i 1%

$$A^2_{0.05} = 2.492 \quad A^2_{0.01} = 3.857$$

dovoljno tačne za uzorke sa $N > 10$.

DODATAK D

Intervali poverenja funkcije raspodele: varijanse (standardne greške) kvantila

Interval poverenja funkcije raspodele određuje se tako što se za veličinu slučajne promenljive zadatog povratnog perioda, odnosno kvantil $X(T)$, računaju gornja i donja granica intervala poverenja:

- gornja granica:

  $$X_g(T) = X(T) + |z_\alpha| S_{X(T)}$$

- donja granica:

  $$X_d(T) = X(T) - |z_\alpha| S_{X(T)}$$

gde je $S_{X(T)}$ standardna grešku kvantila $X(T)$, a $z_\alpha = z(\alpha) = -z(1-\alpha)$ standardizovana normalna promenljiva za prag značajnosti α koji odgovara intervalu poverenja β:

$$\alpha = \frac{1-\beta}{2}$$

Na primer, za $\beta = 90\%$, $\alpha = 5\%$.

Standardna greška kvantila

Standardna greška kvantila $S_{X(T)}$ predstavlja kvadratni koren varijanse kvantila:

$$S^2_{X(T)} = \mathrm{var}[X(T)]$$

i ona se razlikuje za pojedine teorijske raspodele.

- *Normalna raspodela i Pirson 3 raspodela*

Može se pokazati da standardna greška kvantila za normalnu raspodelu približno iznosi:

$$S_{X(T)} = \frac{S_x}{\sqrt{N}} \sqrt{1 + \frac{1}{2} K^2(T)}$$

gde je $K(T)$ faktor frekvencije za odgovarajuću raspodelu (videti odeljak o faktorima frekvencije). Gornji izraz se koristi i kao aproksimacija za standardna greška kvantila za Pirson 3 raspodelu, s obzirom da je tačan izraz za ovu raspodelu veoma komplikovan.

- *Gumbelova raspodela*

Za Gumbelovu raspodelu standardna greška kvantila je:

$$S_{X(T)} = \frac{S_x}{\sqrt{N}} \sqrt{1 + 1.1396 K(T) + 1.1 K^2(T)}$$

gde je $K(T) = K_G(T)$ faktor frekvencije za Gumbelovu raspodelu.

- *Log-normalna raspodela i log-Pirson 3 raspodela*

Za log-normalnu i log-Pirson 3 raspodelu radi se analogno kao i za normalnu odnosno Pirson 3 raspodelu, samo za logaritmovani niz $Y = \log X$. Ako je $Y(T)$ logaritamovan kvantil:

$$Y(T) = \log X(T)$$

tada je standardna greška logaritmovanog kvantila:

$$S_{Y(T)} = \frac{S_y}{\sqrt{N}} \sqrt{1 + \frac{1}{2} K^2(T)}$$

gde je *K*(*T*) faktor frekvencije za odgovarajuću raspodelu. Tada su grenice intervala poverenja:

$$Y_g(T) = Y(T) + | z_\alpha | S_{Y(T)}$$

$$Y_d(T) = Y(T) - | z_\alpha | S_{Y(T)}$$

odnosno

$$X_g(T) = 10^{Y_g(T)}$$

$$X_d(T) = 10^{Y_d(T)}$$

Faktori frekvencije

Faktor frekvencije *K*(*T*) se definiše kao:

$$K(T) = \frac{X(T) - \bar{x}}{S_x}$$

- *Normalna raspodela*

Kako se kvantil u normalnoj raspodeli određuje na osnovu standardizovane normalne promenljive *z*:

$$X(T) = \bar{x} + S_x \cdot z(T)$$

faktor frekvencije za normalnu raspodelu je očigledno:

$$K_N(T) = \frac{X(T) - \bar{x}}{S_x} = \frac{\bar{x} + S_x \cdot z(T) - \bar{x}}{S_x} = z(T)$$

- *Log-normalna raspodela*

S obzirom da se kvantili po log-normalnoj raspodeli određuju tako što se na lgaritmovan niz $Y = \log X$ primeni normalna raspodela:

$$Y(T) = \bar{y} + S_y \cdot z(T)$$

faktor frekvencije za log-normalnu raspodelu je:

$$K_{LN}(T) = \frac{Y(T) - \bar{y}}{S_y} = z(T)$$

- *Gumbelova raspodela*

Kako se kvantil po Gumbelovoj raspodeli određuje pomoću standardizovane Gumbelove promenljive y_G:

$$X(T) = \alpha_G y_G(T) + u = 0.78 S_x \cdot y_G(T) + \bar{x} - 0.45 S_x$$

gde su α_G i u parametri Gumbelove raspodele, faktor frekvencije za Gumbelovu raspodelu može se izraziti kao:

$$K_G(T) = \frac{X(T) - \bar{x}}{S_x} = \frac{\alpha_G y_G(T) + u - \bar{x}}{S_x} = 0.78 y_G(T) - 0.45$$

Kako je

$$y_G(T) = -\ln(-\ln F) = -\ln\left(\ln\frac{T}{T-1}\right)$$

faktor frekvencije Gumbelove raspodele može se izraziti kao funkcija samo povratnog perioda:

$$K_G(T) = -0.78\ln\left(\ln\frac{T}{T-1}\right) - 0.45$$

- *Pirson 3 raspodela*

Faktor frekvencije za Pirson 3 raspodelu najjednostavnije se računa direktno:

$$K_P(T) = \frac{X(T) - \bar{x}}{S_x}$$

Gornji izraz je moguće prikazati i u drugačijoj formi. Kvantil za Pirson 3 raspodelu određuje se na osnovu parametara a, b i c i inverzne funkcije jednoparametarske gama raspodele koja zavisi samo od parametra a:

$$X(T) = c + bF^{-1}(T, a)$$

Kako je:

$$a = \frac{4}{C_{sx}^2}, \quad b = \frac{S_x C_{sx}}{2}, \quad c = \bar{x} - ab$$

može se napisati:

$$b = \frac{S_x}{\sqrt{a}}, \quad c = \bar{x} - S_x\sqrt{a}$$

pa sledi

$$K_P(T) = \frac{X(T) - \bar{x}}{S_x} = \frac{c + bF^{-1}(T, a) - \bar{x}}{S_x} = \frac{\bar{x} - S_x\sqrt{a} + \frac{S_x}{\sqrt{a}}F^{-1}(T, a) - \bar{x}}{S_x} = \sqrt{a} + \frac{F^{-1}(T, a)}{\sqrt{a}}$$

ili

$$K_P(T) = \frac{2}{C_{sx}} + \frac{C_{sx}}{2}F^{-1}(T, C_{sx})$$

Drugim rečima, faktor frekvencije za Pirson 3 raspodelu ne zavisi samo od povratnog perioda kao kod dvoparametarskih raspodela, već i od koeficijenta asimetrije.

- *Log-Pirson 3 raspodela*

Kako se log-Pirson 3 raspodela dobija primenom Pirson 3 raspodele na logaritmovan niz Y = log X, faktor frekncencije se računa na osnovu logaritmovanog kvantila:

$$K_{LP}(T) = \frac{Y(T) - \bar{y}}{S_y}$$

Analogno Pirson 3 raspodeli, faktor frekvencije za log-Pirson 3 raspodelu je funkcija povratnog perioda i koeficijenta asimetrije logaritmovanog niza C_{sy}.

DODATAK E

Statističke funkcije u programu Excel

Funkcija	Opis
AVERAGE	srednja vrednost niza
BINOMDIST	binomna raspodela
CHIDIST	hi-hvadrat raspodela
CHIINV	inverzna hi-kvadrat raspodela
CHITEST	hi-kvadrat test
CORREL	koeficijent korelacije
COUNT	broj članova niza
COVAR	kovarijansa
EXPONDIST	eksponencijalna raspodela
FDIST	F raspodela
FINV	inverzna F raspodela
FREQUENCY	frekvencije (učestalosti)
FTEST	F-test
GAMMADIST	dvoparametarska gama raspodela
GAMMAINV	inverzna dvoparametarska gama raspodela
GAMMALN	prirodni logaritam gama funkcije
GEOMEAN	geometrijska sredina
HARMEAN	harmonijska sredina
HYPGEOMDIST	hipergeometrijska raspodela
INTERCEPT	koeficijent linearne regresije (konstanta)
KURT	koeficijent zakrivljenosti (4. momenat)
LOGINV	inverzna log-normalna raspodela
LOGNORMDIST	log-normalna raspodela
MAX	maksimum miza
MEDIAN	medijana niza
MIN	minimum niza
MODE	mod niza (najučestalija vrednost)
NEGBINOMDIST	negativna binomna raspodela
NORMDIST	normalna raspodela
NORMINV	inverzna normalna raspodela
NORMSDIST	standardizovana normalna raspodela
NORMSINV	inverzna standardizovana normalna raspodela
POISSON	Poasonova raspodela
RANK	rang u nizu
SKEW	koeficijent asimetrije
SLOPE	koeficijent linearne regresije (nagib linije)
STANDARDIZE	standardizovana vrednost
STDEV	standardna devijacija (nepristrasna)
STDEVP	standardna devijacija populacije
STEYX	standardna greška regresije Y na X
TDIST	Studentova t raspodela
TINV	inverzna Studentova t raspodela
TTEST	t-test
VAR	varijansa (nepristrasna)
VARP	varijansa populacije
WEIBULL	Vejbulova raspodela
ZTEST	z-test

DODATAK F

Neke statističke tablice

Test Kolmogorova-Smirnova

Vrednosti statistike D_N za zadatu vrednost praga značajnosti α

	α				
N	0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.829
4	0.494	0.525	0.564	0.624	0.734
5	0.446	0.474	0.510	0.563	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.369	0.409	0.489
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.449
13	0.284	0.302	0.325	0.361	0.432
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.327	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.279	0.309	0.371
19	0.237	0.252	0.271	0.301	0.361
20	0.231	0.246	0.265	0.294	0.352
25	0.210	0.220	0.238	0.264	0.317
30	0.190	0.200	0.218	0.242	0.290
35	0.180	0.190	0.202	0.224	0.269
40	0.170	0.180	0.189	0.210	0.252
> 40	$1.07/\sqrt{N}$	$1.14/\sqrt{N}$	$1.22/\sqrt{N}$	$1.36/\sqrt{N}$	$1.63/\sqrt{N}$

Hi-kvadrat raspodela

Vrednosti χ^2 za zadatu vrednost funkcije raspodele F,
odnosno za zadatu vrednost praga značajnosti α

				F			
	0.5	0.75	0.9	0.95	0.975	0.99	0.995
				α			
ν	0.5	0.25	0.1	0.05	0.025	0.01	0.005
1	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	29.3	34.8	40.3	43.8	47.0	50.9	53.7
31	30.3	35.9	41.4	45.0	48.2	52.2	55.0
32	31.3	37.0	42.6	46.2	49.5	53.5	56.3
33	32.3	38.1	43.7	47.4	50.7	54.8	57.6
34	33.3	39.1	44.9	48.6	52.0	56.1	59.0
35	34.3	40.2	46.1	49.8	53.2	57.3	60.3
40	39.3	45.6	51.8	55.8	59.3	63.7	66.8
50	49.3	56.3	63.2	67.5	71.4	76.2	79.5

Test Kramera-Mizesa

Vrednosti statistike $N\omega^2$ za zadatu vrednost praga značajnosti α

α	0.5	0.4	0.3	0.25	0.2	0.1	0.05	0.025	0.01	0.005
Nω²	0.120	0.148	0.184	0.210	0.241	0.348	0.462	0.581	0.744	0.870

Vrednosti praga značajnosti α za zadate vrednosti statistike $N\omega^2$

Nω²	**α**	***Nω²***	**α**	***Nω²***	**α**	***Nω²***	**α**	***Nω²***	**α**
0.010	1.000	0.230	0.216	0.450	0.053	0.670	0.015	0.890	0.004
0.015	1.000	0.235	0.209	0.455	0.052	0.675	0.015	0.895	0.004
0.020	0.997	0.240	0.202	0.460	0.050	0.680	0.014	0.900	0.004
0.025	0.990	0.245	0.195	0.465	0.049	0.685	0.014	0.905	0.004
0.030	0.978	0.250	0.188	0.470	0.048	0.690	0.013	0.910	0.004
0.035	0.963	0.255	0.182	0.475	0.046	0.695	0.013	0.915	0.004
0.040	0.941	0.260	0.176	0.480	0.045	0.700	0.013	0.920	0.004
0.045	0.912	0.265	0.170	0.485	0.043	0.705	0.012	0.925	0.004
0.050	0.879	0.270	0.165	0.490	0.042	0.710	0.012	0.930	0.004
0.055	0.847	0.275	0.159	0.495	0.041	0.715	0.012	0.935	0.003
0.060	0.814	0.280	0.154	0.500	0.040	0.720	0.011	0.940	0.003
0.065	0.783	0.285	0.149	0.505	0.039	0.725	0.011	0.945	0.003
0.070	0.752	0.290	0.144	0.510	0.038	0.730	0.011	0.950	0.003
0.075	0.721	0.295	0.140	0.515	0.036	0.735	0.010	0.955	0.003
0.080	0.692	0.300	0.135	0.520	0.035	0.740	0.010	0.960	0.003
0.085	0.664	0.305	0.131	0.525	0.034	0.745	0.010	0.965	0.003
0.090	0.636	0.310	0.127	0.530	0.033	0.750	0.010	0.970	0.003
0.095	0.611	0.315	0.123	0.535	0.032	0.755	0.009	0.975	0.003
0.100	0.586	0.320	0.119	0.540	0.032	0.760	0.009	0.980	0.003
0.105	0.563	0.325	0.115	0.545	0.031	0.765	0.009	0.985	0.003
0.110	0.541	0.330	0.112	0.550	0.030	0.770	0.009	0.990	0.003
0.115	0.520	0.335	0.108	0.555	0.029	0.775	0.008	0.995	0.003
0.120	0.500	0.340	0.105	0.560	0.028	0.780	0.008	1.000	0.002
0.125	0.480	0.345	0.101	0.565	0.027	0.785	0.008	1.005	0.002
0.130	0.461	0.350	0.098	0.570	0.027	0.790	0.008	1.010	0.002
0.135	0.443	0.355	0.095	0.575	0.026	0.795	0.008	1.015	0.002
0.140	0.425	0.360	0.092	0.580	0.025	0.800	0.007	1.020	0.002
0.145	0.409	0.365	0.090	0.585	0.024	0.805	0.007	1.025	0.002
0.150	0.392	0.370	0.087	0.590	0.024	0.810	0.007	1.030	0.002
0.155	0.377	0.375	0.084	0.595	0.023	0.815	0.007	1.035	0.002
0.160	0.362	0.380	0.082	0.600	0.022	0.820	0.007	1.040	0.002
0.165	0.348	0.385	0.079	0.605	0.022	0.825	0.006	1.045	0.002
0.170	0.335	0.390	0.077	0.610	0.021	0.830	0.006	1.050	0.002
0.175	0.322	0.395	0.074	0.615	0.021	0.835	0.006	1.055	0.002
0.180	0.310	0.400	0.072	0.620	0.020	0.840	0.006	1.060	0.002
0.185	0.299	0.405	0.070	0.625	0.019	0.845	0.006	1.065	0.002
0.190	0.288	0.410	0.068	0.630	0.019	0.850	0.006	1.070	0.002
0.195	0.278	0.415	0.066	0.635	0.018	0.855	0.005	1.075	0.002
0.200	0.268	0.420	0.064	0.640	0.018	0.860	0.005	1.080	0.002
0.205	0.258	0.425	0.062	0.645	0.017	0.865	0.005	1.085	0.002
0.210	0.249	0.430	0.060	0.650	0.017	0.870	0.005	1.090	0.002
0.215	0.240	0.435	0.058	0.655	0.016	0.875	0.005	1.095	0.001
0.220	0.232	0.440	0.057	0.660	0.016	0.880	0.005	1.100	0.001
0.225	0.224	0.445	0.055	0.665	0.015	0.885	0.005		