Sadržaj

Poglavlje 12

TEORIJA PROCJENA

Teorija procjene sastoji se u konstrukciji metoda za ocjenu vrijednosti jednog ili više parametara poznate distribucije slučajne varijable.

U prethodnom poglavlju smo za slučajnu varijablu X (statističko obilježje) promatrali n vrijednosti $x_1, x_2, ..., x_n$ kao uzorak veličina n.
U ovom poglavlju ćemo vrijednosti $x_1, x_2, ..., x_n$ promatrati kao pojedinačne vrijednosti niza od n nezavisnih slučajnih varijabli $X_1, X_2, ..., X_n$ koje imaju istu distribuciju kao i slučajna varijable X.

Definicija 12.1 *(SLUČAJNI UZORAK veličine n)*

Neka je X slučajna varijabla (statističko obilježje populacije) s funkcijom distribucije F(x). Slučajni uzorak veličine n za slučajnu varijablu X je slučajni vektor $(X_1, X_2, ..., X_n)$*, gdje su sve slučajne varijable* X_i*, i=1,...,n, nezavisne sa zajedničkom funkcijom distribucije vjerojatnosti F(x).*
Vrijednost slučajnog uzorka je uređena n-torka $(x_1, x_2, ..., x_n)$ *ako je izmjerena vrijednost slučajnih varijabli* X_i *jednaka* $x_i \in R(X)$*, i=1,...,n.*
Ako je X diskretna slučajna varijabla ($R(X)$ *je konačan ili prebrojiv), onda je* $(X_1, X_2, ..., X_n)$ *diskretni slučajni uzorak , a ako je X kontinuirana slučajna varijabla (*$R(X) \subseteq R$*), onda je* $(X_1, X_2, ..., X_n)$ *kontinuirani slučajni uzorak.*

Definicija 12.2 *(STATISTIKA)*

Ako je $Y = h(X_1, X_2, ..., X_n)$*, gdje je h funkcija od n varijabli, onda se slučajna varijabla Y naziva statistika.*

NAPOMENA 12.1 *Odabrani elementi uzorka veličine n iz populacije trebaju biti izabrani slučajno. Trebamo koristiti tablicu slučajnih brojeva za izbor n slučajnih brojeva ili program za generiranje slučajnih brojeva.*

PRIMJER 12.1 *Zadana je diskretna slučajna varijabla X s funkcijom vjerojatnosti*

x_i	0	1	2
p_i	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Što je uzorak veličine 2 za ovu slučajnu varijablu?
Odredi sve moguće vrijednosti slučajnog uzorka veličine 2 za X.

Rješenje:
Slučajni uzorak veličine 2 za slučajnu varijablu X je slučajni vektor (X_1, X_2), gdje su sve slučajne varijable X_1 i X_2 nezavisne i jednake funkcije distribucije kao i X. Slika slučajne varijable X je $R(X) = \{0, 1, 2\}$. Slučajne varijable X_1 i X_2 mogu poprimiti iste vrijednosti kao i X. Vrijednost slučajnog uzorka je uređena dvojka (x_1, x_2) elemenata iz $R(X)$, tj. to je varijacija s ponavljanjem $r = 2$-og razreda od $n = 3$ elemenata. Broj svih takvih varijacija je $V_3^{(2)} = 3^2 = 9$.
Sve moguće vrijednosti slučajnog uzorka veličine 2 za slučajnu varijablu X:
(0,0), (0,1), (0,2),
(1,0), (1,1), (1,2),
(2,0), (2,1), (2,2).

12.1 TOČKASTE PROCJENE PARAMETARA

MOTIV 12.1 *Koliki uzorak iz normalne razdiobe s varijancom 81 treba biti da bi s vjerojatnošću 0.9544 apsolutna razlika uzoračke aritmetičke sredine i očekivanja bila manja od 5.5?*

Slučajna varijabla je određena svojom funkcijom distribucije. Mnoga statistička obilježja imaju zajedničku teorijsku funkciju distribucije pa govorimo o poznatim distribucijama (razdiobama): binomna, uniformna, normalna, Poissonova,....
Svaka razdioba karakterizirana je svojim parametrima $n, p, a, b, \mu, \sigma^2, \lambda, ...$:
$X \sim B(n, p)$,
$X \sim U(a, b)$,
$X \sim N(\mu, \sigma^2)$,
$X \sim Po(\lambda), ...$
Ako želimo odrediti vezu između teorijske i statističke razdiobe postavljaju se dva zadatka:
1. parametarske procjene, kada pretpostavimo teorijsku razdiobu i moramo odrediti (procijeniti) parametre te razdiobe.
2. neparametaske procjene, kada moramo odabrati razdiobu.

Definicija 12.3 *(PROCJENITELJ ILI ESTIMATOR)*
Procjenitelj nepznatog parametra t je funkcija slučajnog uzorka
$\widehat{T} = h(X_1, X_2, ..., X_n)$.
Procjenitelj je statistika.

Zadatak je odrediti procjenitelj $\widehat{T}$ za parametar t koji će "najbolje" procijeniti t.
Za procjenu jednog parametra možemo izabirati razne procjenitelje (funkcije h).

Definicija 12.4 *(NEPRISTRANI PROCJENITELJ)*
Procjenitelj $\widehat{T}$ je nepristran za parametar t ako je očekivanje od $\widehat{T}$ jednako vrijednosti parametra t: $E(\widehat{T}) = t$.

Definicija 12.5 *(ASIMPTOTSKI NORMALAN PROCJENITELJ)*
Procjenitelj $\widehat{T}$ je asimptotski normalan za parametar t ako slučajnoj varijabli $\dfrac{\widehat{T} - t}{\sqrt{Var(\widehat{T})}}$ asimptotski, kad $n \to \infty$, pripada standardna normalna razdioba (distribucija) N(0,1).

Definicija 12.6 *(UZORAČKA ARITMETIČKA SREDINA)*

Statistika $\overline{X} = h(X_1, X_2, ..., X_n) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ *zove se uzoračka aritmetička sredina. Vrijednost uzoračke aritmetičke sredine računa pomoću*
$\overline{x} = h(x_1, x_2, ..., x_n) = \frac{1}{n}\sum\limits_{i=1}^{n} x_i,\ \overline{x} = \frac{1}{n}\sum\limits_{k=1}^{r} x_k^* f_k.$

TEOREM 12.1 *(Svojstva uzoračke aritmetičke sredine)*

(i) *Neka je X slučajna varijabla s teorijskim očekivanjem* μ*, i varijancom* σ^2*,* $0 < \sigma^2 < \infty$*, koju ispitujemo pomoću slučajnog uzorka* $(X_1, X_2, ..., X_n)$*. Uzoračka aritmetička sredina* $\overline{X}$ *je pouzdan procjenitelj za* μ*:*
$E(\overline{X}) = \mu.$

(ii) $Var(\overline{X}) = \frac{1}{n}\sigma^2.$

(iii) $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$ *ako je normalna distribucija.*

(iv) *Uzoračka aritmetička sredina*$\overline{X}$ *je asimptotski normalan procjenitelj za* μ*:* $\overline{X}^* = \frac{\overline{X}-\mu}{\sqrt{Var(\overline{X})}} \sim N(0,1).$

Dokaz:*tko želi znati više*
(i) $E(\overline{X}) = E(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} E(X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} \mu = \mu.$
(ii)

$$\begin{aligned} Var(\overline{X}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}n \cdot Var(X_i) = \frac{1}{n}\sigma^2. \end{aligned}$$

(iii) prema (i) i (ii).
(iv) Prisjetimo se centralnog graničnog teorema:
Neka je $S_n = X_1 + X_2 + ... + X_n$,tada slučajna varijabla $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ konvergira k $N(0,1)$.

$$\overline{X}^* = \frac{\overline{X}-\mu}{\sqrt{Var(\overline{X})}} = \frac{\frac{S_n}{n}-\mu}{\sqrt{\frac{1}{n}\sigma^2}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \text{konvergira } (n \to \infty) \text{ k } N(0,1).$$

PRIMJER 12.2 *Izračunati $P(69 < \overline{X} < 75)$, ako je $\overline{X}$ uzoračka aritmetička sredina uzorka veličine n=36 iz normalne razdiobe $X \sim N(70, 144)$.*

Rješenje:
Ako je $X \sim N(70, 144)$ i $n = 36$, onda je $\overline{X} \sim N(70, 4)$,
$\overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \frac{\overline{X}-70}{2} \sim N(0, 1)$.

$$\begin{aligned} P(69 < \overline{X} < 75) &= F^*(\sqrt{n}\frac{75-\mu}{\sigma}) - F^*(\sqrt{n}\frac{59-\mu}{\sigma}) \\ &= F^*(\frac{75-70}{2}) - F^*(\frac{69-70}{2}) = F^*(2.5) - F^*(-0.5) = 0.68. \end{aligned}$$

Prisjetimo se motivacijskog primjera:

PRIMJER 12.3 *motiv*

Koliki uzorak iz normalne razdiobe s varijancom 81 treba biti da bi s vjerojatnošću 0.9544 apsolutna razlika uzoračke aritmetičke sredine i očekivanja bila manja od 5.5?

Rješenje:
Neka je $X \sim N(\mu, 81)$ i $P(|\overline{X} - \mu| < 5.5) \geq 0.9544$.
Trebamo odrediti veličinu uzorka n.
Znamo da je $\overline{X} \sim N(\mu, \frac{81}{n})$, a $\overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \sqrt{n}\frac{\overline{X}-\mu}{9} \sim N(0, 1)$.

$$\begin{aligned} P(|\overline{X} - \mu| < 5.5) &= P(|\sqrt{n}\frac{\overline{X}-\mu}{\sigma}| < \sqrt{n}\frac{5.5}{\sigma}) = P(|\overline{X}^*| \leq \sqrt{n}\frac{5.5}{\sigma}) \\ &= 2F^*(\sqrt{n}\frac{5.5}{\sigma}) - 1. \end{aligned}$$

Iz zadane vjerojatnosti dobivamo: $2F^*(\sqrt{n}\frac{5.5}{\sigma}) - 1 \geq 0.9544$,

$$F^*(\sqrt{n}\frac{5.5}{\sigma}) \geq 0.9772 \;\Rightarrow\; \sqrt{n}\frac{5.5}{\sigma} \geq 2 \;\Rightarrow\; n \geq 11.$$

Definicija 12.7 *(UZORAČKA VARIJANCA)*

Statistika $\widehat{\Sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ zove se uzoračka varijanca.

$$\widehat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Vrijednost uzoračke varijance računa se formulom

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2.$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 f_k = \frac{1}{n}\sum_{k=1}^{r}(x_k^*)^2 f_k - \overline{x}^2.$$

TEOREM 12.2 *(Svojstva uzoračke varijance)*

Neka je X slučajna varijabla s teorijskim očekivanjem μ i varijancom σ^2, koju ispitujemo pomoću slučajnog uzorka $(X_1, X_2, ..., X_n)$. Uzoračka varijanca $\widehat{\Sigma}^2$ nije pouzdan procjenitelj za σ^2: $E(\widehat{\Sigma^2}) = \frac{n-1}{n}\sigma^2$.

Dokaz:*tko želi znati više*
Prisjetimo se da je $Var(X)$=$E(X^2) - E(X)^2$ i $Var(\overline{X}) = \frac{1}{n}\sigma^2$.

$$\begin{aligned} E(\widehat{\Sigma^2}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}[Var(X_i) + E(X_i)^2] - [Var(\overline{X}) + E(\overline{X})^2] \\ &= [Var(X_i) + E(X_i)^2] - [Var(\overline{X}) + E(\overline{X})^2] \\ &= [\sigma^2 + \mu^2] - [\frac{1}{n}\sigma^2 + \mu^2] = \frac{n-1}{n}\sigma^2. \end{aligned}$$

Definicija 12.8 *(KORIGIRANA UZORAČKA VARIJANCA)*

Statistika $\widehat{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ zove se korigirana uzoračka varijanca.

$$\widehat{S^2} = \frac{n}{n-1}\widehat{\Sigma^2} = \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2).$$

Vrijednost korigirane uzoračke varijance računa se formulom

$$\widehat{s^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2).$$

$$\widehat{s^2} = \frac{1}{n-1}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 f_k = \frac{1}{n-1}(\sum_{k=1}^{r}(x_k^*)^2 f_k - n\overline{x}^2).$$

TEOREM 12.3 *(Svojstva korigirane uzoračke varijance)*

Neka je X slučajna varijabla s teorijskim očekivanjem μ i varijancom σ^2, koju ispitujemo pomoću slučajnog uzorka $(X_1, X_2, ..., X_n)$.
Korigirana uzoračka varijanca $\widehat{S^2}$ je pouzdan procjenitelj za σ^2: $E(\widehat{S^2}) = \sigma^2$.

Dokaz: Prisjetimo se da je $E(\widehat{\Sigma^2}) = \frac{n-1}{n}\sigma^2$.

$$E(\widehat{S^2}) = E(\frac{n}{n-1}\widehat{\Sigma^2}) = \frac{n}{n-1}E(\widehat{\Sigma^2}) = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2.$$

TEOREM 12.4 *(O VEZI $\widehat{S^2}, \widehat{\Sigma^2}$ I DISTRIBUCIJA $\chi^2(n-1), t(n-1)$)*

Neka su $\overline{X}$, $\widehat{S^2}$, $\widehat{\Sigma^2}$ statistike slučajnog uzorka $(X_1, X_2, ..., X_n)$ iz normalne razdiobe $X \sim N(\mu, \sigma^2)$.

Tada vrijedi:

(i) Statistika $\frac{n-1}{\sigma^2}\widehat{S^2} = \frac{n}{\sigma^2}\widehat{\Sigma^2} \sim \chi^2(n-1)$,

(ii) Statistika $\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} = \sqrt{n-1}\frac{\overline{X}-\mu}{\widehat{\Sigma}} \sim t(n-1)$.

Dokaz: *tko želi znati više*

(i) Dokaz je složen i koristi svojstvo $\chi^2(n)$ distribucije: $Y \sim \chi^2(n)$ ako je $Y = Y_1^2 + Y_2^2 + ... + Y_n^2,\ Y_i \sim N(0,1)$.

(ii) Koristimo svojstvo Studentove distribucije s n stupnjeva slobode $t(n)$: $Z \sim t(n)$ ako je $Z = Y\sqrt{\frac{n}{U}}$, za $Y \sim N(0,1),\ U \sim \chi^2(n)$.

Računamo za $X \sim N(\mu, \sigma^2),\ \overline{X} \sim N(\mu, \frac{\sigma^2}{n}),\ \overline{X}^* = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$:

$$\begin{aligned}\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} &= \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sigma\sqrt{n-1}}{\sqrt{n-1}} \cdot \frac{1}{\sqrt{\widehat{S^2}}} = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2}}} \cdot \frac{1}{\sqrt{\widehat{S^2}}} \\ &= \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2} \cdot \widehat{S^2}}}.\end{aligned}$$

Prema tvrdnji (i) zaključujemo $\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} \sim t(n-1)$.

Koristeći $\widehat{S^2} = \frac{n}{n-1}\widehat{\Sigma^2}$ možemo dobiti i tvrdnju $\sqrt{n-1}\frac{\overline{X}-\mu}{\widehat{\Sigma}} \sim t(n-1)$.

PRIMJER 12.4 *Izračunati uzoračku aritmetičku sredinu, uzoračku varijancu i korigiranu uzoračku varijancu u primjeru "težina studenata".*

x_{ksr}	f_k
61	5
64	18
67	42
70	27
73	8
	n=100

Rješenje: $\overline{x} = \frac{1}{n}\sum\limits_{k=1}^{r} x_k^* f_k$.

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{k=1}^{r} x_{ksr}^* f_k = \frac{1}{100}(61*5+64*18+67*42+70*27+73*8)\\ &= \frac{1349}{20} = 67,45\end{aligned}$$

$\widehat{s}^2 = \frac{1}{n-1}(\sum\limits_{k=1}^{r}(x_k^*)^2 f_k - n\overline{x}^2)$

$$\widehat{s}^2 = \frac{1}{n-1}(\sum_{k=1}^{r}(x_{ksr}^*)^2 f_k - n\overline{x}^2) = \frac{379}{44} = 8,6136$$

$\widehat{\sigma}^2 = \frac{1}{n}\sum\limits_{k=1}^{r}(x_k^*)^2 f_k - \overline{x}^2$.

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{r}(x_{ksr}^*)^2 f_k - \overline{x}^2 = \frac{3411}{400} = 8,5275$$

12.2 REGRESIJSKA ANALIZA

MOTIV 12.2 *Deformacije x [mm] i Brinellova tvrdoća y [$\frac{kg}{mm^2}$] za neki tip čelika dani su tablicom*

x	06	09	11	11	13	22	26	28	33	35
y	68	67	65	53	44	40	37	28	34	32

Odredite pravce regresije, uzorački koeficijent regresije i uzorački koeficijent korelacoje. Jesu li deformacija i Brinellova tvrdoća jako korelirane?

Definicija 12.9 *(UZORAČKA KOVARIJANCA. UZORAČKI KOEFICIJENT KORELACIJE) Neka je za zadani slučajni 2-dim vektor (X, Y) dobiven slučajni uzorak $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Statistika*

$$\widehat{\mu}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

zove se uzoračka kovarijanca. Vrijednost korigirane uzoračke kovarijance računa se formulom

$$\widehat{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Neka su $\widehat{\sigma}_1$ i $\widehat{\sigma}_2$ uzoračke standardne devijacije od X i Y. Uzorački koeficijent korelacije komponenti X i Y slučajnog vektora je definiran s

$$\widehat{\rho}_{xy} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1 \cdot \widehat{\sigma}_2}.$$

$$\widehat{\rho}_{xy} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}.$$

NAPOMENA 12.2 *(regresijska analiza)*

Regresijska analiza (engl. regression analysis) je statistička metoda za odeđivanje veze među slučajnim varijablama. Promatramo u slučajnom vektoru (X, Y) jednu slučajnu varijablu (npr. X) kao nezavisnu-kontroliranu (njene vrijednosti zadajemo). Druga varijabla Y je slučajna varijabla i zanima nas kako ona ovisi o X.
Prema Napomeni **??** *u poglavlju Dvodimenzionalni slučajni vektor računamo uzoračke pravce regresije.*
Ako je X nezavisna varijabla i $Y = aX + b$, parametre a i b možemo odrediti metodom najmanjih kvadrata tako da $E((Y - (aX + b))^2)$ ima minimalnu vrijednost.

$a = \widehat{\rho}_{xy} \cdot \frac{\widehat{\sigma_2}}{\widehat{\sigma_1}} = \frac{\widehat{\mu}_{XY}}{\widehat{\sigma_1^2}}$ *je uzorački koeficijent regresije Y po X.*

$b = \overline{y} - \widehat{\rho}_{xy} \cdot \frac{\widehat{\sigma_2}}{\widehat{\sigma_1}} \overline{x} = \overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma_1^2}} \cdot \overline{x}$

$y - \overline{y} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma_1^2}} (x - \overline{x}),$

$y - \overline{y} = \rho_{XY} \cdot \frac{\sigma_2}{\sigma_1} (x - \overline{x})$ *je pravac regresije Y po X.*

Analogno, ako je X=aY+b

$a = \widehat{\rho}_{xy} \cdot \frac{\widehat{\sigma_1}}{\widehat{\sigma_2}} = \frac{\widehat{\mu}_{XY}}{\widehat{\sigma_2^2}}$ *je uzorački koeficijent regresije X po Y*

$x - \overline{x} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma_2^2}} (Y - \overline{y})$

$x - \overline{x} = \widehat{\rho}_{xy} \cdot \frac{\widehat{\sigma_1}}{\widehat{\sigma_2}} (y - \overline{y})$ *je pravac regresije X po Y.*

PRIMJER 12.5 *U tablici su zapisani uzorci visina x i y od 12 mama i njihovih kćeri.*

x	165	160	170	163	173	158	178	168	173	170	175	180
y	173	168	173	165	175	168	173	165	180	170	173	178

Oderedite uzorački pravac regresije Y u odnosu na X i uzorački pravac regresije X u odnosu na Y. Odredite uzorački koeficijent korelacije i uzorački koeficijent regresije Y po X. Jesu li visine mama i kćeri jako korelirane?

Rješenje: Trebamo odrediti a i b u jednadžbi $y = ax + b$:

$$a = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma_1^2}} = \frac{n \sum x \cdot y - (\sum x) \cdot (\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$b = \overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma_1^2}} \cdot \overline{x} = \frac{(\sum y) \cdot (\sum x^2) - (\sum x) \cdot (\sum x \cdot y)}{n \sum x^2 - (\sum x)^2}$$

Računamo: $\sum x = 2033$, $\sum y = 2061$, $\sum x^2 = 344.95$, $\sum x \cdot y = 349.42$, $\sum y^2 = 354.22$.

Uzorački koeficijent regresije Y po X je $a = 0.48$, $b = 90.9$ pa je pravac regresije Y

po X $y = 0.48x + 90.9$.

Trebamo odrediti a' i b' u jednadžbi $x = a'y + b'$:

$$a' = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$b' = \overline{x} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{y} = \frac{(\sum x) \cdot (\sum y^2) - (\sum y) \cdot (\sum x \cdot y)}{n\sum y^2 - (\sum y)^2}$$

Uzorački koeficijent regresije X po Y je $a' = 1.02$, $b' = -5.12$ pa je pravac regresije X po Y $x = 1.02x - 5.12$.

Uzorački koeficijent korelacije je

$$\widehat{\rho}_{xy} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1 \cdot \widehat{\sigma}_2}.$$

$$\widehat{\rho}_{xy} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{\sqrt{n - \sum x^2 - (\sum x)^2} \cdot \sqrt{n - \sum y^2 - (\sum y)^2}} = 0.69.$$

Prisjetimo se motivacijskog primjera:

PRIMJER 12.6 *motiv*

Deformacije x [mm] i Brinellova tvrdoća y $[\frac{kg}{mm^2}]$ za neki tip čelika dani su tablicom

x	*06*	*09*	*11*	*11*	*13*	*22*	*26*	*28*	*33*	*35*
y	*68*	*67*	*65*	*53*	*44*	*40*	*37*	*28*	*34*	*32*

Odredite pravce regresije , uzorački koeficijent korelacije i uzorački koeficijent regresije. Jesu li deformacije i Brienellova tvrdoća jako korelirane ?

Rješenje: Trebamo odrediti a i b u jednadžbi $y = ax + b$:

$$a = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{x} = \frac{(\sum y) \cdot (\sum x^2) - (\sum x) \cdot (\sum x \cdot y)}{n\sum x^2 - (\sum x)^2}$$

Računamo: $\sum x = 183$, $\sum y = 440$, $\sum x^2 = 4665$, $\sum x \cdot y = 7701$, $\sum y^2 = 23232$.
Uzorački koeficijent regresije Y po X je $a = -1.32$, $b = 75.72$ pa je pravac regresije

Y po X $y = -1.32x + 75.72$.

Trebamo odrediti a' i b' u jednadžbi $x = a'y + b'$:

$$a' = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$b' = \overline{x} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{y} = \frac{(\sum x) \cdot (\sum y^2) - (\sum y) \cdot (\sum x \cdot y)}{n\sum y^2 - (\sum y)^2}$$

Uzorački koeficijent regresije X po Y je $a' = -0.72$, $b' = 55.72$ pa je pravac regresije X po Y $x = -0.72x + 55.72$.

Uzorački koeficijent korelacije je

$$\widehat{\rho}_{xy} = \frac{n\sum x \cdot y - (\sum x) \cdot (\sum y)}{\sqrt{n - \sum x^2 - (\sum x)^2} \cdot \sqrt{n - \sum y^2 - (\sum y)^2}} = -0.97$$

Budući je $\widehat{\rho}_{xy} \approx -1$ slučajne varijable su linearnoj vezi, jako korelirane.

12.3 METODA NAJVEĆE VJEROJATNOSTI (ML) *tko želi znati više*

U ovom poglavlju istaknuli smo primjere s oznakom *važno* u kojima su dani procjenitelji za parametre osnovnih distribucija u smislu najveće vjerojatnosti.

MOTIV 12.3 *U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti:*
$64.9, 64.1, 63.8, 64.0.$
(a) Izračunajte vrijednost procjenitelja (u smislu najveće vjerojatnosti) za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.
(b) Izračunajte vrijednost nepristranih procjenitelja za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.

Definicija 12.10 *(FUNKCIJA VJERODOSTOJNOSTI) Neka je X slučajna varijabla (statističko obilježje) sa teorijskom funkcijom distribucije F(x,t) s nepoznatim parametrom t i sa funkcijom vjerojatnosti za diskretnu razdiobu i funkcijom gustoće vjerojatnosti za kontinuiranu f(x,t). Neka je* $(x_1, x_2, ..., x_n)$ *vrijednost slučajnog uzorka* $(X_1, X_2, ..., X_n)$ *za promatranu varijablu.*
Za diskretnu razdiobu funkcija vjerodostojnosti $L(t)$ *definira se kao funkcija vjerojatnosti slučajnog uzorka (slučajnog vektora):*

$$L(t) = P(X_1 = x_1, t) \cdot P(X_1 = x_2, t) \cdot ... \cdot P(X_1 = x_n, t).$$

Za kontinuiranu razdiobu funkcija vjerodostojnosti $L(t)$ *definira se kao funkcija gustoće vjerojatnosti slučajnog uzorka (slučajnog vektora):*

$$L(t) = f(x_1, t) \cdot f(x_2, t) \cdot ... \cdot f(x_n, t).$$

Metoda najveće vjerojatnosti (ML = maximum likelihood method), za određivanje procjenitelja $\widehat{T} = h(X_1, X_2, ..., X_n)$ za nepoznati parametar t sastoji se u izboru one funkcije h takve da funkcija vjerodostojnosti $L(t)$ (ili $lnL(t)$) dostiže najveću vrijednost za $t = h(x_1, x_2, ..., x_n)$.

$$\frac{d}{dt}L(t) = 0 \quad \Rightarrow \quad t = h(x_1, x_2, ..., x_n) \quad \Rightarrow \quad \widehat{T} = h(X_1, X_2, ..., X_n).$$

PRIMJER 12.7 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj za parametar λ *u populaciji s Poisonovom razdiobom* $Po(\lambda)$ *jednak* $\widehat{T} = \overline{X}$.

Rješenje: Neka je $X \sim Po(\lambda)$.
Teorijska funkcija vjerojatnosti je $f(x,\lambda)$=$P(X=x,\lambda)=\frac{\lambda^x}{x!}e^{\lambda}$.
Trebamo naći ML-procjenitelj za λ.
Funkcija vjerodostojnosti je

$$\begin{aligned} L(\lambda) &= P(X_1 = x_1,\lambda)\cdot P(X_1 = x_2,\lambda)\cdot \ldots \cdot P(X_1 = x_n,\lambda) \\ &= \frac{\lambda^{x_1}}{x_1!}e^{\lambda}\cdot\frac{\lambda^{x_2}}{x_2!}e^{\lambda}\cdot\ldots\cdot\frac{\lambda^{x_n}}{x_n!}e^{\lambda} = \frac{\lambda^{\sum x_i}}{x_1!x_2!\ldots x_n!}e^{-n\lambda}. \end{aligned}$$

Tražimo maksimum funkcije vjerodostojnosti $\ln L(\lambda)$:

$$\ln L(\lambda) = -n\lambda + \sum x_i \ln\lambda - \sum \ln(x_i!),$$

$$\frac{d}{d\lambda}\ln L(\lambda) = -n + \frac{1}{\lambda}\sum x_i.$$

$$\frac{d}{d\lambda}\ln L(\lambda) = 0 \quad\Rightarrow\quad \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

$$\lambda = h(x_1,x_2,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad\Rightarrow\quad \widehat{T} = h(X_1,X_2,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

ML-procjenitelj za λ u Poisonovoj razdiobi je $\widehat{T}=\overline{X}$.
Možemo pokazati da je $\widehat{T}$ nepristrani procjenitelj $E(\widehat{T})=\lambda$.
Prisjetimo se da je $E(X)$=λ i da je $\overline{X}$ nepristrani procjenitelj za očekivanje.
$E(\widehat{T}) = E(\overline{X}) = E(X) = \lambda$.

PRIMJER 12.8 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj
(a) za parametar μ u populaciji s Normalnom razdiobom $N(\mu,\sigma^2)$ ako je σ^2 poznato jednak $\widehat{T}=\overline{X}$
(b) za parametar σ^2 u populaciji s Normalnom razdiobom $N(\mu,\sigma^2)$ ako je μ poznato jednak $\widehat{T}=\widehat{\Sigma}^2$.

Rješenje: Neka je $X \sim N(\mu,\sigma^2)$.
Teorijska funkcija gustoće vjerojatnosti je $f(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
Trebamo naći ML-procjenitelje za μ i σ^2.
Funkcija vjerodostojnosti je:

$$L(\mu,\sigma^2) = f(x_1,t)\cdot f(x_2,t)\cdot\ldots\cdot f(x_n,t) = (\frac{1}{\sigma\sqrt{2\pi}})^n\cdot e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}.$$

$$\ln L(\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum (x_i-\mu)^2 - n\ln\sigma - n\ln\sqrt{2\pi}.$$

(a) $\frac{d}{d\mu}\ln L(\mu,\sigma^2) = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0,$

$$\frac{d}{d\mu}\ln L(\mu,\sigma^2) = 0 \quad\Rightarrow\quad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x},$$

ML-procjenitelj za očekivanje u normalnoj razdiobi je $\widehat{T} = \overline{X}$. To je nepristrani procjenitelj za μ jer je $E(\widehat{T})=\mu$.

(b) $\frac{d}{d\sigma}\ln L(\mu,\sigma^2) = -\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{\sigma} = 0,$

$$\frac{d}{d\sigma}\ln L(\mu,\sigma^2) = 0 \quad\Rightarrow\quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \widehat{\sigma},\quad \text{uzoračka varijanca.}$$

ML-procjenitelj za varijancu u normalnoj razdiobi je $\widehat{T} = \widehat{\Sigma}^2$. To nije nepristrani procjenitelj za σ^2 jer je $E(\widehat{T})=\frac{n-1}{n}\sigma^2$.

Prisjetimo se motivacijskog primjera:

PRIMJER 12.9 *motiv*

U četiri mjerenja Rockwellove tvrdoće jedne ploče radnici du dobili sljedeće vrijednosti: $64.9, 64.1, 63.8, 64.0$. *(a) Izračunajte vrijednost procjenitelja (u smislu najveće vjerojatnosti-ML-procjenitelj) za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.*
(b) Izračunajte vrijednost nepristranih procjenitelja za očekivanje i varijancu tvrdoće ako pretpostavimo normalnu distribuciju.

Rješenje: (a) ML-procjenitelj za očekivanje μ je uzoračka aritmetička sredina $\overline{X}$.
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 64.2$

ML-procjenitelj za varijancu σ^2 je uzoračka varijanca $\widehat{\Sigma}^2$.
$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = 0.175$

(b)Nepristrani procjenitelj za očekivanje μ je uzoračka aritmetička sredina $\overline{X}$.
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 64.2$

Nepristrani procjenitelj za varijancu σ^2 je korigirana uzoračka varijanca $\widehat{S}^2$.
$\widehat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i-\overline{x})^2 = 0.233.$

PRIMJER 12.10 *važno*

Metodom najveće vjerojatnosti pokažite da je procjenitelj za parametar p u populaciji s Binomnom razdiobom $B(m, p)$ uz pretpostavku da je m poznato jednak $\widehat{T} = \frac{\overline{X}}{m}$

Rješenje: Neka je $X \sim B(m, p)$. Teorijska funkcija vjerojatnosti je $f(x, m, p)$=$P(X = x, m, p) = \binom{m}{x}(1-p)^{m-x}p^x$. Trebamo naći ML-procjenitelj za p. Funkcija vjerodostojnosti je

$$\begin{aligned} L(p) &= P(X_1 = x_1, m, p) \cdot P(X_1 = x_2, m, p) \cdot \ldots \cdot P(X_1 = x_n, m, p) \\ &= \prod_{i=1}^{n} \binom{m}{x_i}(1-p)^{m-x_i}p^{x_i} \end{aligned}$$

Tražimo maksimum funkcije vjerodostojnosti $\ln L(p)$:

$$\ln L(p) = \sum[\ln \binom{m}{x_i} + (m - x_i)\ln(1-p) + x_i \ln p],$$

$$\frac{d}{dp}\ln L(p) = -\frac{1}{1-p}\sum(m - x_i) + \frac{1}{p}\sum x_i$$

$$\frac{d}{d\lambda}\ln L(\lambda) = 0 \quad \Rightarrow \quad p = \frac{1}{m} \cdot \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{m}\overline{x}.$$

$p = h(x_1, x_2, ..., x_n) = \frac{1}{m} \cdot \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \frac{1}{m}\overline{x}$

$\widehat{T} = h(X_1, X_2, ..., X_n) = \frac{1}{m} \cdot \frac{1}{n}\sum\limits_{i=1}^{n} X_i = \frac{1}{m}\overline{X}.$

ML-procjenitelj za p u Binomnoj razdiobi $B(m, p)$ s poznatim m je $\widehat{T} = \frac{1}{m}\overline{X}.$
Možemo pokazati da je $\widehat{T}$ nepristrani procjenitelj $E(\widehat{T}) = p$.
Prisjetimo se da je $E(X)$=mp i da je $\overline{X}$ nepristrani procjenitelj za očekivanje.

$$E(\widehat{T}) = E(\frac{1}{m}\overline{X}) = \frac{1}{m}E(\overline{X}) = \frac{1}{m}E(X) = \frac{1}{m}mp = p.$$

12.4 Ponovimo

PROCJENITELJI PARAMETARA ZADANE DISTRIBUCIJE s očekivanjem μ i varijancom σ^2

slučajni uzorak	$(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$
statistika	$T = h(X_1, X_2, ..., X_n)$, $T : \Omega \to \mathbb{R}$
procjenitelj	$\widehat{T} = h(X_1, X_2, ..., X_n)$
nepristrani procjenitelj za parametar t	$E(\widehat{T}) = t$
uzoračka aritm. sredina	$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$
	$E(\overline{X}) = \mu,$
	$Var(\overline{X}) = \frac{\sigma^2}{n}$
$n \to \infty$	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
uzoračka varijanca	$\widehat{\Sigma}^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$
	$\widehat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \overline{x})^2$
uzoračka stand. devijacija	$\widehat{\sigma}$
	$E(\widehat{\Sigma}^2) = \frac{n-1}{n}\sigma^2$
korigirana uzoračka varijanca	$\widehat{S}^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$
	$\widehat{s}^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$
korigirana uz. stand. devijacija	$\widehat{s}$
	$E(\widehat{S}^2) = \sigma^2$

PROCJENITELJI PARAMETARA ZADANE DISTRIBUCIJE (ML)

distribucija ; parametar	procjenitelj
$Po(\lambda); \lambda$	$\widehat{T} = \overline{X}$
$N(\mu, \sigma^2);\ \ \mu$	$\widehat{T} = \overline{X}$
$N(\mu, \sigma^2);\ \ \sigma^2$	$\widehat{T} = \widehat{\Sigma}^2$
$(B(m, p);\ \ p$	$\widehat{T} = \frac{1}{m}\overline{X}$

STATISTIKE PARAMETARA NORMALNE DISTRIBUCIJE $N(\mu, \sigma^2)$

slučajni uzorak za X	$(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$
statistika $T = h(X_1, X_2, ..., X_n)$	distribucija
$\overline{X}$	$\sim N(\mu, \frac{\sigma}{\sqrt{n}})$
$\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$	$\sim N(0,1)$
$\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}}$	$\sim t(n-1)$
$\frac{n-1}{\sigma^2}\widehat{S}^2$	$\sim \chi^2(n-1)$

REGRESIJSKA ANALIZA

slučajni uzorak za (X, Y)	$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
uzoračka kovarijanca	$\widehat{\mu}_{xy} = \frac{1}{n} \sum_i (x_i - \overline{x})(y_i - \overline{y})$
uzorački koef. korelacije	$\widehat{\rho}_{xy} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1 \cdot \widehat{\sigma}_2}$
	$\frac{n \sum x \cdot y - (\sum x) \cdot (\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}}$
$Y = aX + b$	a- uzorački koef. regresije
a	$\frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2}$
	$= \frac{n \sum x \cdot y - (\sum x) \cdot (\sum y)}{n \sum x^2 - (\sum x)^2}$
b	$\overline{y} - \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2} \cdot \overline{x}$
	$= \frac{(\sum y) \cdot (\sum x^2) - (\sum x) \cdot (\sum x \cdot y)}{n \sum x^2 - (\sum x)^2}$
pravac regresije Y po X	$y - \overline{y} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_1^2}(x - \overline{x})$
pravac regresije X po Y	$x - \overline{x} = \frac{\widehat{\mu}_{xy}}{\widehat{\sigma}_2^2}(y - \overline{y})$