

Uzorak, statistika

# Metod uzorka

- Određivanje raspodjele obilježja na datoj populaciji često nije moguće registrovanjem vrijednosti obilježja na svakom elementu populacije
- Metodom uzorka se iz populacije na određeni način uzima konačan podskup od  $n$  elemenata i na njegovim elementima se ispituju vrijednosti obilježja  $X$
- Reprezentativan uzorak – zaključci o obilježju  $X$  koji su dobijeni za uzorak prenose se na populaciju
  - Reprezentativnost se obezbjeđuje slučajnim izborom elemenata iz populacije u uzorak

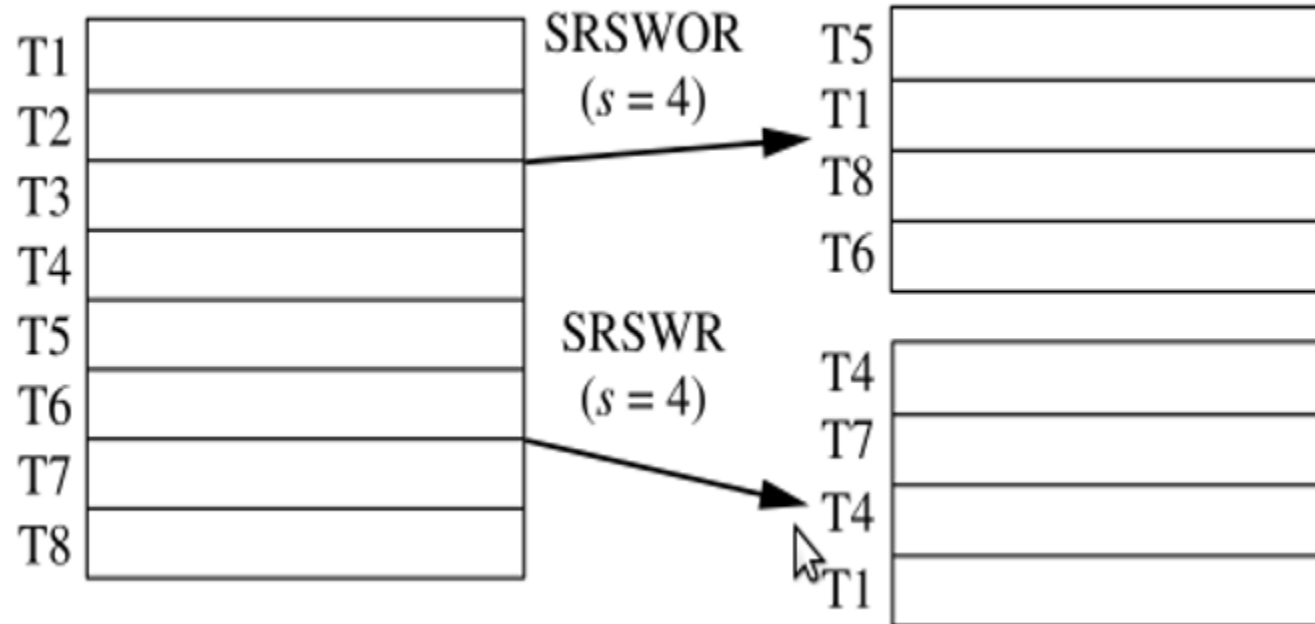
# Vrste uzorka

- Slučajni uzorak dobija se postupkom kod kojeg svaki element populacije ima istu vjerovatnoću da bude izabran u uzorak
- Kod prostog slučajnog uzorka sem što svaki element populacije ima jednaku vjerovatnoću da bude izabran, i svaki uzorak ima istu vjerovatnoću da bude izabran
- Prost slučajni uzorak sa vraćanjem dobija se ako se svaki element populacije vraća poslije izvlačenja u uzorak što znači da ma koji element populacije može biti izabran više puta u uzorak
- Prost slučajni uzorak bez vraćanja generiše se na način da se jednom izabrani element populacije više ne vraća u populaciju

## Vrste uzorka (2)

- Sistematski uzorak generiše se na način da se elementi populacije biraju u jednakim intervalima – vremena, prostora, poretka – npr. izbor svakog stotog pacijenta u ambulanti
- Stratifikovani uzorak dobija se kada se populacija podijeli na stratumе a zatim se proporcionalno i na slučajan način biraju elementi iz svakog stratuma
- Klaster uzorak dobija se kada se populacija podijeli na klasterе a u uzorak se biraju samo neki od tako dobijenih klastera

# Prost slučajan uзорak



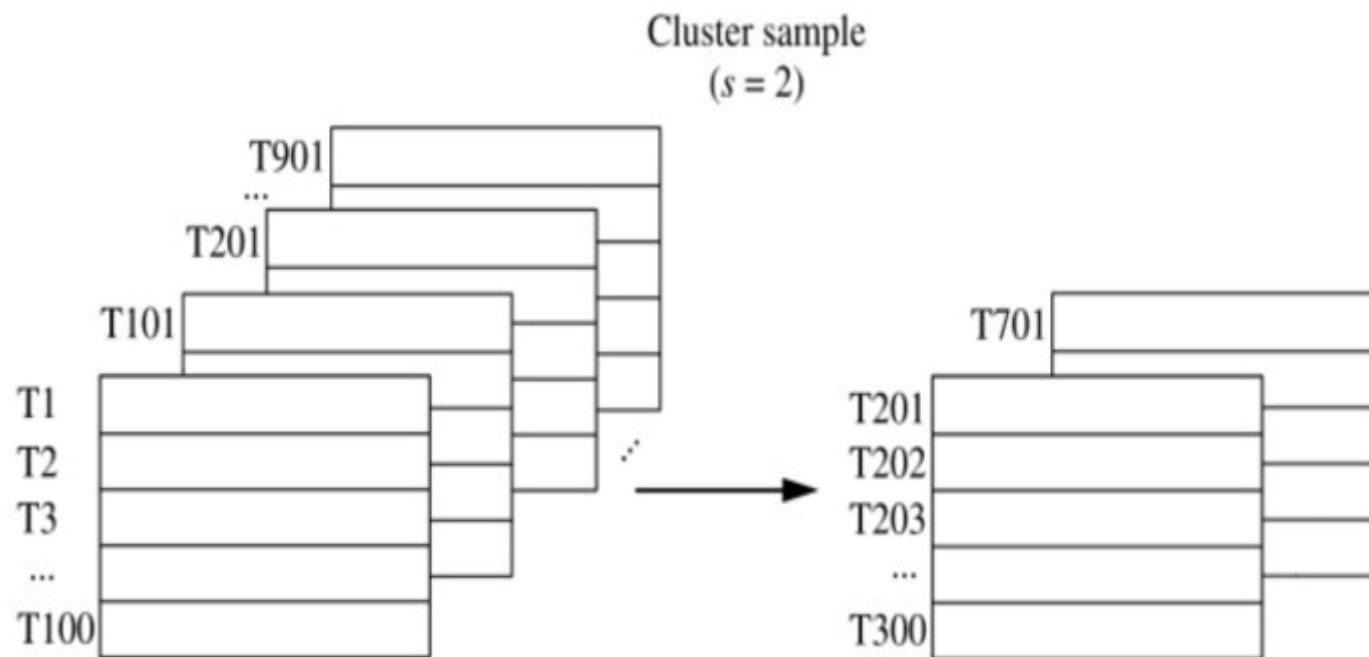
# Stratifikovani uzorak

Stratified sample  
(according to *age*)

T38	youth
T256	youth
T307	youth
T391	youth
T96	middle_aged
T117	middle_aged
T138	middle_aged
T263	middle_aged
T290	middle_aged
T308	middle_aged
T326	middle_aged
T387	middle_aged
T69	senior
T284	senior

T38	youth
T391	youth
T117	middle_aged
T138	middle_aged
T290	middle_aged
T326	middle_aged
T69	senior

# Klaster uзорak



# Statistika

- Kada smo uzabrali uzorak na svakom elementu posmatra se obilježje  $X$ . Registrovanje vrijednosti obilježja  $X$  naziva se statistički eksperiment. Niz dobijenih vrijednosti  $(x_1, x_2, \dots, x_n)$  naziva se realizovani uzorak
- Posmatramo slučajne promjenljive  $X_1, X_2, \dots, X_n$  koje imaju istu raspodjelu kao  $X$  i važi da su međusobno nezavisne
- **Definicija.** Neka se na populaciji  $\xi$  posmatra obilježje  $X$ . Prost slučajan uzorak obima  $n$  za obilježje  $X$  je  $n$ -torka nezavisnih slučajnih promjenljivih  $(X_1, X_2, \dots, X_n)$  od kojih svaka ima istu raspodjelu kao i  $X$ .



# Statistika (2)

- **Definicija.** Neka je  $(X_1, X_2, \dots, X_n)$  prost slučajan uzorak obima  $n$  za obilježje  $X$ . Slučajna promjenljiva  $Y=f(X_1, X_2, \dots, X_n)$  je statistika ako je funkcija samo promjenljivih  $X_1, X_2, \dots, X_n$ .
- Statistike koje se najčešće koriste su:
  - Sredina uzorka
  - Najmanja vrijednost
  - Najveća vrijednost
  - Medija
  - Mod
  - Disperzija, popravljena disperzija
  - Standardna devijacija

# Osobine i raspodjele nekih statistika

- Ako je  $X$  normalno distribuirana varijabla, tada aritmetička sredina uzorka ima raspodjelu

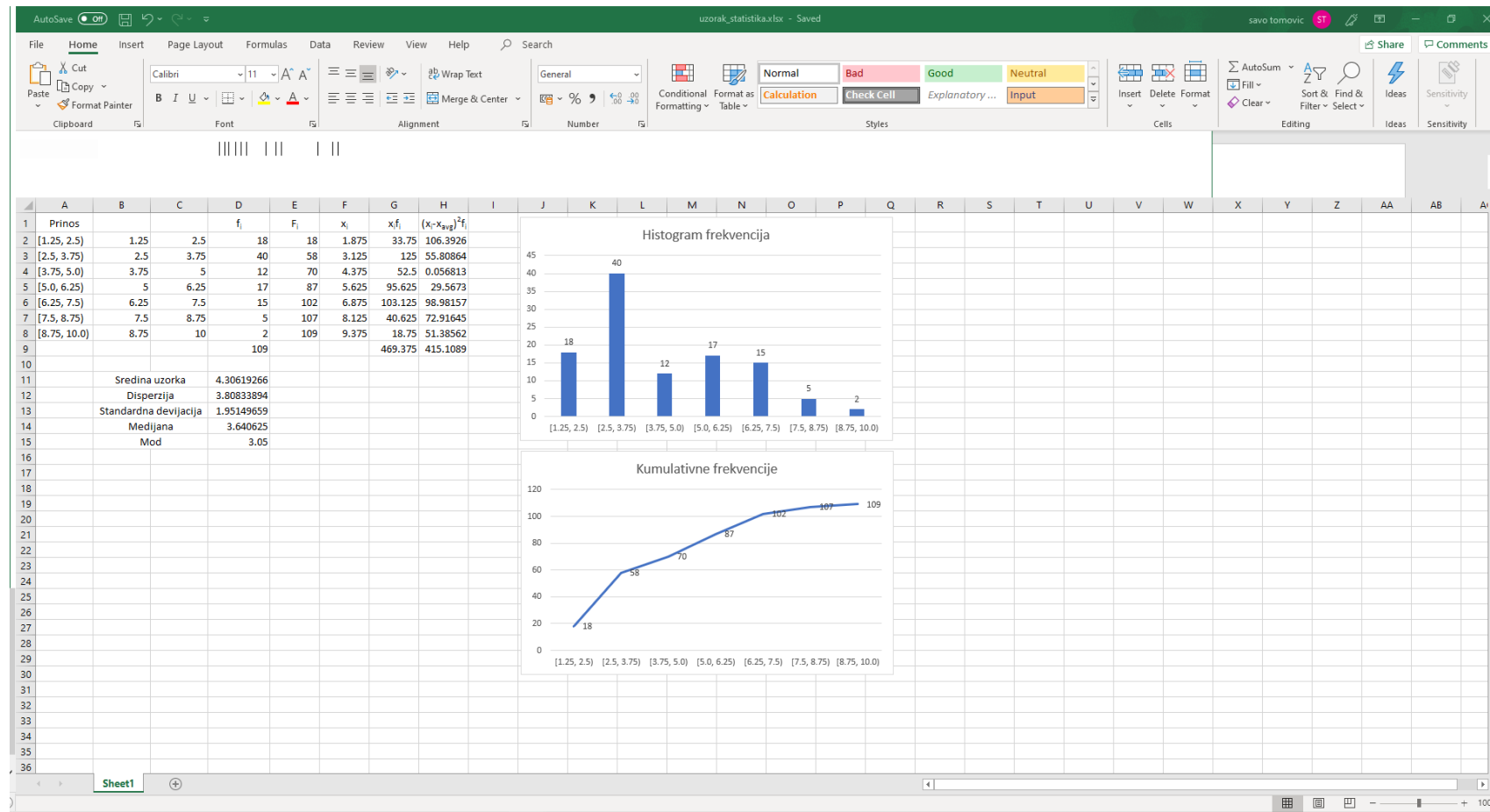
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- **Primjer 1.** Visina čovjeka ima približno normalnu raspodjelu  $N(167.5, 8^2\text{cm})$ . Ako je slučajno izabran uzorak od 64-ro ljudi, koja je vjerovatnoća da je njihova prosječna visina između 165 i 170cm?

- b)

$$\begin{aligned}\mu_{\bar{x}} &= \mu = 167.5, & \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{64}} = 1. \\ z_1 &= \frac{165 - 167.5}{1} = -2.5, & z_2 &= \frac{170 - 167.5}{1} = 2.5 \\ P\{165 \leq \bar{X} \leq 170\} &= \Phi(2.5) - \Phi(-2.5) = 0.9876.\end{aligned}$$

# Tačkasta ocjena nepoznatog parametra



# Intervalna ocjena nepoznatog parametra

- **Definicija.** Neka je dat prost slučajan uzorak  $(X_1, X_2, \dots, X_n)$  za obilježje  $X$  i neka su  $U_1 = u_1(X_1, X_2, \dots, X_n)$  i  $U_2 = u_2(X_1, X_2, \dots, X_n)$  dvije statistike takve da je  $P(U_1 \leq \theta \leq U_2) = \gamma$ . Tada se interval  $(U_1, U_2)$  naziva interval povjerenja za parametar  $\theta$  na nivou  $\gamma$ .
- Jedna od statistika  $U_1$  ili  $U_2$  može biti konstanta, pa se tada govori o jednostranom intervalu povjerenja. Uobičajene vrijednosti za  $\gamma$  su 0.9, 0.95, 0.99

# Interval povjerenja kada $X$ ima normalnu raspodjelu

- Interval povjerenja za nepoznatu srednju vrijednost obilježja
  - Disperzija  $\sigma^2$  je poznata, interval povjerenja je

$$\left( \bar{X}_n - c \frac{\sigma}{\sqrt{n}}, \bar{X}_n + c \frac{\sigma}{\sqrt{n}} \right)$$

$\Upsilon$	c
0.9	1.65
0.95	1.96
0.99	2.58

- Disperzija je nepoznata

$$\left( \bar{X}_n - c \frac{S}{\sqrt{n}}, \bar{X}_n + c \frac{S}{\sqrt{n}} \right)$$

- Broj c je kvantil reda  $(1+\Upsilon)/2$  Studentove  $t_{n-1}$  raspodjele

## Primjer 2

- Poznato je da obilježje ima disperziju  $\sigma^2=3.24$ . Na osnovu uzorka obima  $n=25$  dobijena je ocjena  $\bar{x}_{25} = 2.1$ . Odrediti 90% interval povjerenja za nepoznatu srednju vrijednost obilježja.

$$\begin{aligned} & \left( \bar{X}_n - c \frac{\sigma}{\sqrt{n}}, \bar{X}_n + c \frac{\sigma}{\sqrt{n}} \right) \\ & = \left( 2.1 - 1.65 \frac{\sqrt{3.24}}{25}, 2.1 + 1.65 \frac{\sqrt{3.24}}{25} \right) = (1.506, 2.694) \end{aligned}$$

# Primjer 3

- Menadžer firme želi da procijeni nedeljna primanja nekoliko hiljada zaposlenih sa greškom od 1800 CSD sa 99% povjerenja. Od ranije poznato da nedeljna primanja zaposlenih imaju normalnu raspodjelu sa devijacijom 3600 CSD. Koji je najmanji obim uzorka potreban?

Interval povjerenja je  $(\bar{X}_n - c \frac{\sigma}{\sqrt{n}}, \bar{X}_n + c \frac{\sigma}{\sqrt{n}})$ , pa je njegova širina  $d = 2c \frac{\sigma}{\sqrt{n}}$ . Za  $\gamma = 0.99$  je  $c = 2.58$ . Otuda je  $2 * 2.58 * \frac{3600}{\sqrt{n}} = 1800$ , odnosno  $\sqrt{n} = 2 * 2.58 * \frac{3600}{1800}$  i  $n = 26.63 \approx 27$  radnika.

# Primjer 4

- Neka su iz uzorka obima  $n=10$  dobijene ocjene  $\bar{x}_{10} = 3$  i  $s=0.25$  ( $s$  je devijacija uzorka). Odrediti 95% interval povjerenja za nepoznatu srednju vrijednost obilježja.

Interval povjerenja kada nije poznata disperzija je  $\left(\bar{X}_n - c \frac{s}{\sqrt{n}}, \bar{X}_n + c \frac{s}{\sqrt{n}}\right)$ , gdje je broj  $c$  kvantil reda  $\frac{1+\gamma}{2}$  Studentove  $t_{n-1}$  raspodjele. Za 95% interval povjerenja potrebno je u tablici Studentove raspodjele naći  $\frac{1+0.95}{2} = 0.975$  kvantil sa  $n-1=9$  stepeni slobode, odnosno  $c=2.262$ . Traženi interval je  $\left(\bar{X}_n - c \frac{s}{\sqrt{n}}, \bar{X}_n + c \frac{s}{\sqrt{n}}\right) = \left(3 - 2.262 \frac{0.25}{\sqrt{10}}, 3 + 2.262 \frac{0.25}{\sqrt{10}}\right) = (2.821, 3.178)$



# Interval povjerenja za nepoznatu disperziju

- Jednostrani interval povjerenja

Slučajna promjenljiva  $\frac{(n-1)S^2}{\sigma^2}$  ima  $\chi_{n-1}^2$  (hi-kvadrat) raspodjelu, pa se interval određuje na sljedeći način  $\left(0, \frac{(n-1)S^2}{c}\right)$  gdje je broj  $c$  kvantil reda  $1-\gamma$  raspodjele  $\chi_{n-1}^2$ .

- Dvostrani interval povjerenja

$\left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$  gdje je broj  $c_2$  kvantil reda  $\frac{1+\gamma}{2}$ , a  $c_1$  kvantil reda  $\frac{1-\gamma}{2}$  raspodjele  $\chi_{n-1}^2$

# Primjer 5

- Za dati uzorak je  $\bar{x}_{25} = 3.5, s^2 = 0.71, s = 0.85$ . Odrediti jednostrani interval povjerenja na nivou  $\Upsilon=0.9$  i dvostrani nivo povjerenja na nivou  $\Upsilon=0.95$  za nepoznatu disperziju obilježja.

Kvantil reda  $1-\Upsilon=0.1$  raspodjele  $\chi_{24}^2$  je  $c = 15.6587$  pa je

jednostrani interval povjerenja  $\left(0, \frac{(n-1)s^2}{c}\right) =$

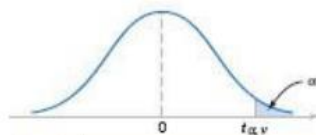
$$\left(0, \frac{(25-1) \cdot 0.71}{15.6587}\right) = (0, 1.088)$$

Kvantil reda  $\frac{1+\Upsilon}{2} = \frac{1+0.95}{2} = 0.972$  raspodjele  $\chi_{24}^2$  je  $c_2 =$

$39.364$ , a kvantil reda  $\frac{1-\Upsilon}{2} = \frac{1-0.95}{2} = 0.025$  raspodjele  $\chi_{24}^2$  je

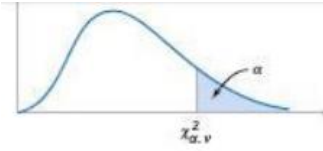
$c_1 = 12.4012$  pa je dvostrani interval povjerenja

$$\left(\frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1}\right) = \left(\frac{(25-1) \cdot 0.71}{39.364}, \frac{(25-1) \cdot 0.71}{12.4012}\right) = (0.433, 1.374)$$



# Studentova raspodjela

$\alpha \backslash \nu$	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
$\infty$	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



# Hi-kvadrat raspodjela

$\alpha \backslash v$	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67