

Statistika



VJEROJATNOST I DISTRIBUCIJE VJEROJATNOSTI

Doc. dr. sc. Igor Štoković

VJEROJATNOST

- **osnova zaključivanja u statistici**
- **osnova povezanosti između uzorka i populacije**

VJEROJATNOST VALJANIH ZAKLJUČAKA O POPULACIJI NA TEMELJU UZORKA

- statističko promatranje rijetko obuhvaća cijeli skup - populaciju
- najčešće je promatranjem obuhvaćen dio populacije - uzorak
- jedan od najznačajnijih ciljeva statističke analize je donositi zaključke o populaciji na temelju uzorka

VJEROJATNO - SLUČAJNO

- koliko je vjerojatno da rezultati utvrđeni u uzorku vrijede u populaciji ?
- u kojoj mjeri su rezultati utvrđeni u uzorku posljedica slučaja

Aritmetička sredina uzorka i populacije

- primjer:
 - farma “x” ima ukupno 5000 spolno zrelih ovaca
 - u uzorku od 50 slučajno odabranih ovaca izmjerena je tjelesna masa u kg
 - aritmetička sredina za tjelesnu masu iznosila je 42,8 kg, a standardna devijacija 6,1 kg
- pitanje: da li vrijednosti za tjelesnu masu utvrđene na uzorku od 50 ovaca vrijede za cijelu populaciju od 5000 ovaca ?

VJEROJATNOST - definicija

- **tri osnovna pristupa:**

1. Osobni pristup - razina osobne procjena da će se neki događaj dogoditi (primjer: postoji 80% vjerojatnosti da će se krava MB 100 oteliti 28. travnja 2006.)

2. A priori (model) pristup - vjerojatnost se procijenjuje na temelju provjerenih modela koje nije potrebno provjeravati (primjer: postoji 50% vjerojatnosti da će bačeni novčić pasti na “pismo” i 50% vjerojatnosti da će pasti na “glavu”)

3. Pristup na osnovi (relativne) učestalosti - vjerojatnost je udio mogućeg rezultata (događaja) koji možemo očekivati u velikom broju promatranja provedenih pod istim uvjetima (primjer: postoji 12% vjerojatnosti da će krmača oprasiti 13 odojaka)

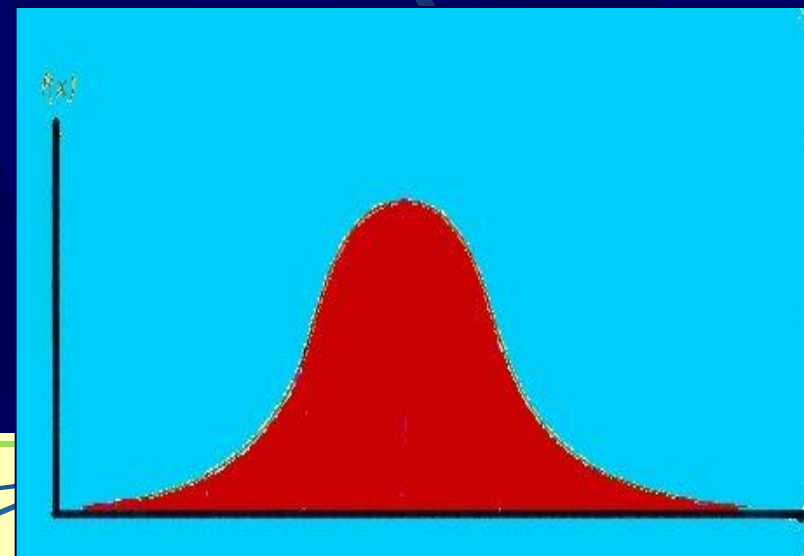
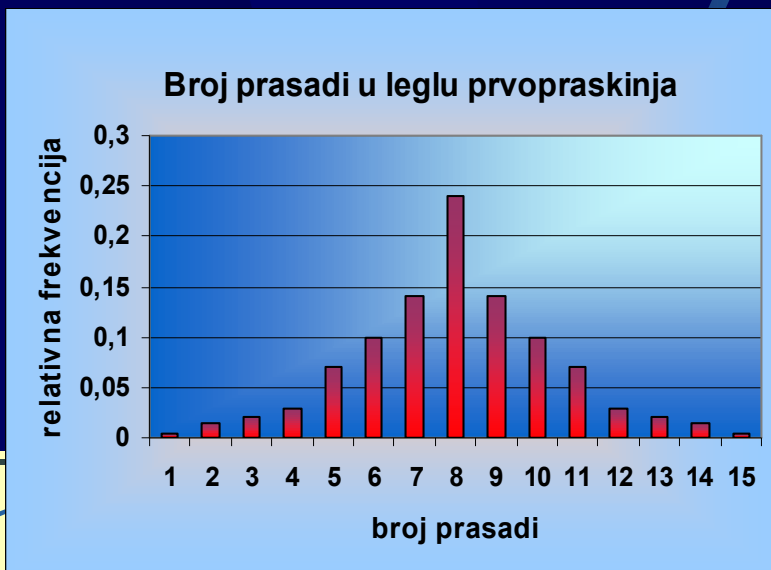
IZRAŽAVANJE VJEROJATNOSTI

- ukoliko vjerojatnost definiramo kao relativnu učestalost ili udio tada:
 - numerička vrijednost vjerojatnosti iznosi između 0 i 1
 - vjerojatnost 0 znači nemogući događaj (“životinja će doživjeti 2000 godina”)
 - vjerojatnost 1 znači siguran događaj (“životinja će uginuti”)
 - često vjerojatnost izražavamo u % (od 0 do 100%) ili kao udio (šansa za događaj je 1 od 3)

DISTRIBUCIJE VJEROJATNOSTI

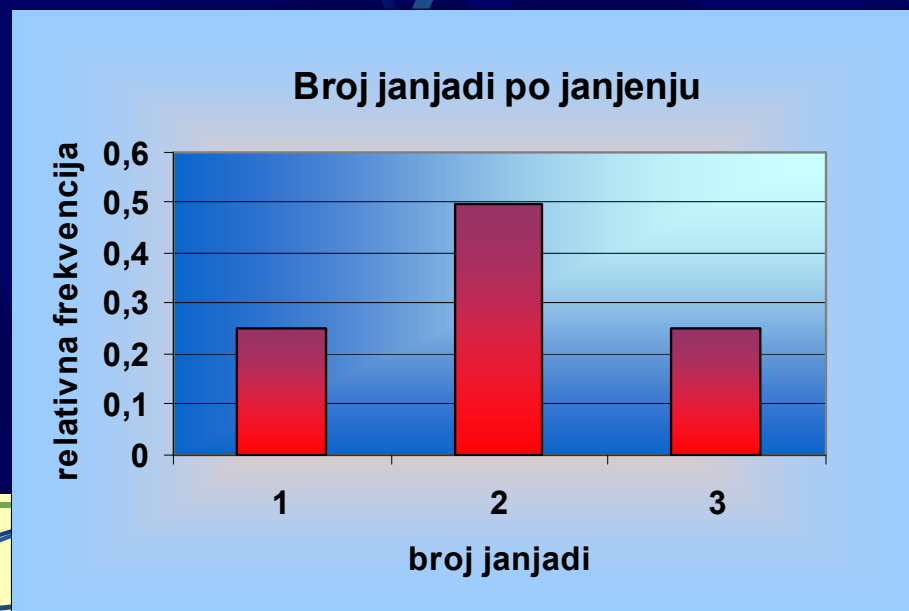
- teorijski modeli za izračunavanje vjerojatnosti nekog ishoda (događaja)
- podjela prema tipu varijable distribucije vjerojatnosti mogu biti:
 - diskontinuirane
 - kontinuirane

GRAFIČKI PRIKAZ DISKONTINUIRANIH I KONTINUIRANIH DISTRIBUCIJA VJEROJATNOSTI



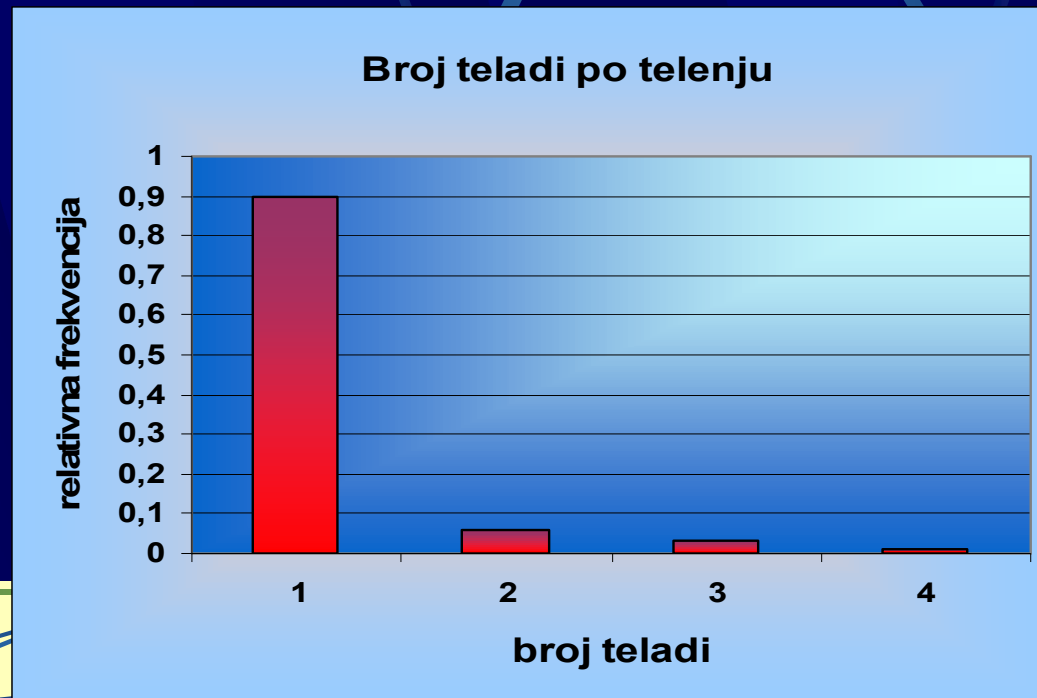
a) PRIMJER JEDNOSTAVNE DISKONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

- varijabla ima samo 3 moguća ishoda (napr: plodnost ovaca – jedno janje, dva janjeta ili tri janjeta u leglu)
- svi mogući ishodi varijable prikazani su na apscisi, a visina stupaca označava vjerojatnost pojave za određeni ishod
- zbroj vjerojatnosti svih ishoda (zbroj visine svih stupaca) iznosi 1



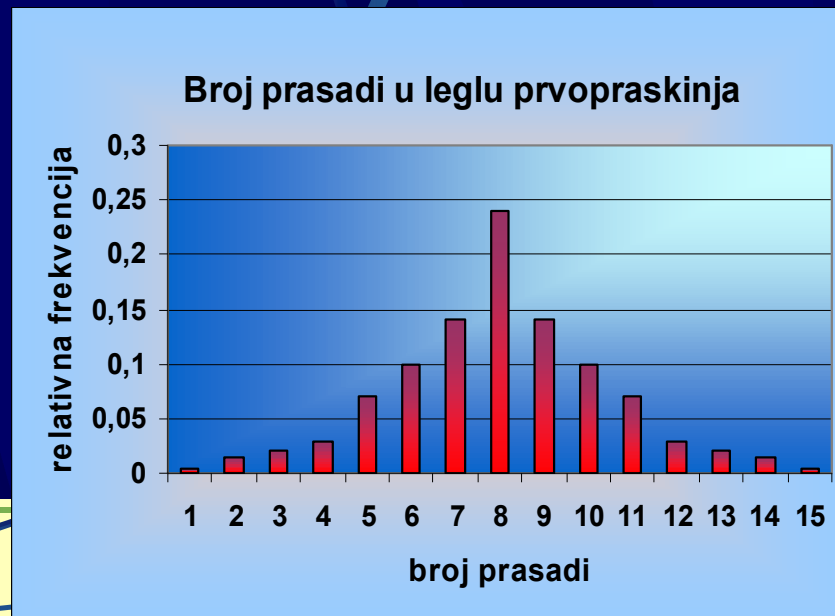
b) PRIMJER JEDNOSTAVNE DISKONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

- varijabla ima 4 moguća ishoda (napr: broj teladi u leglu - 1, 2, 3 ili 4)
- svi mogući ishodi varijable prikazani su na apscisi, a visina stupaca označava vjerojatnost pojave za određeni ishod
- zbroj vjerojatnosti svih ishoda (zbroj visine svih stupaca) iznosi 1



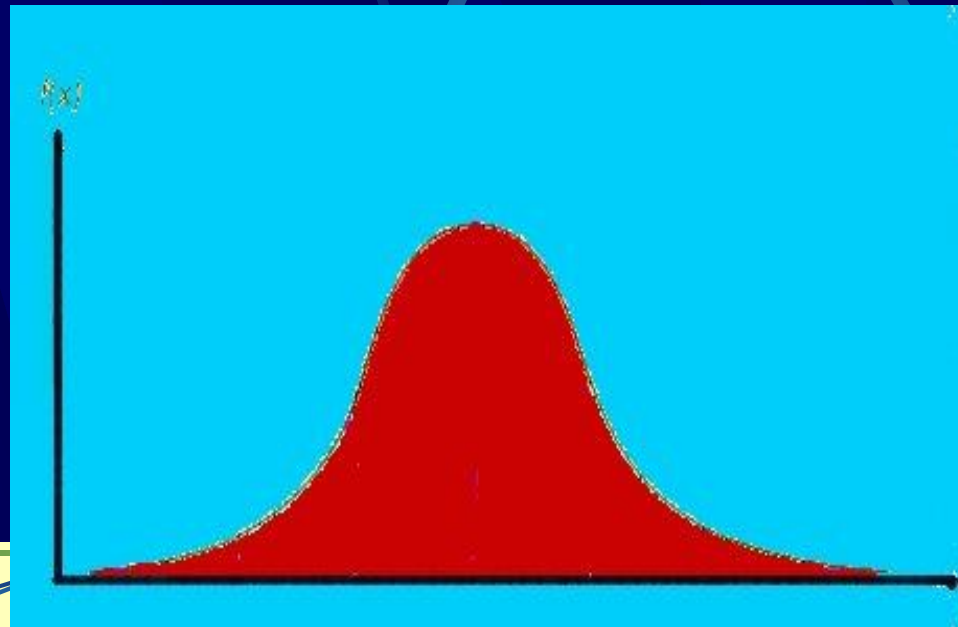
c) PRIMJER DISKONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI S VEĆIM (ALI KONAČNIM) BROJEM ISHODA

- varijabla ima 15 mogućih ishoda (napr. plodnost krmača u legu može biti 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ili 15 odojaka u leglu)
- svi mogući ishodi varijable prikazani su na apscisi, a visina stupaca označava vjerojatnost pojave za određeni ishod
- zbroj vjerojatnosti svih ishoda (zbroj visine svih stupaca) iznosi 1



d) PRIMJER KONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

- varijabla ima beskonačni broj ishoda koji ovise o preciznosti mjerenja (napr: tjelesna masa)
- nije moguće nacrtati pojedinačne stupce za svaki ishod
- zbroj vjerojatnosti svih ishoda iznosi 1, a to je ukupna površina ispod krivulje



DISKONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

- **Binomna distribucija**
- **Poisson distribucija**

KONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

- **Normalna distribucija**
- **Studentova t - distribucija**
- **Hi - kvadrat distribucija**
- **F - distribucija**

STATISTIKA

NORMALNA DISTRIBUCIJA

NORMALNA DISTRIBUCIJA

– značenje i oblik

- osnova za razumijevanje pojmova statističke vjerojatnosti
- Gauss-ova (njemački matematičar, 1777.-1858.)
- zvonasta krivulja

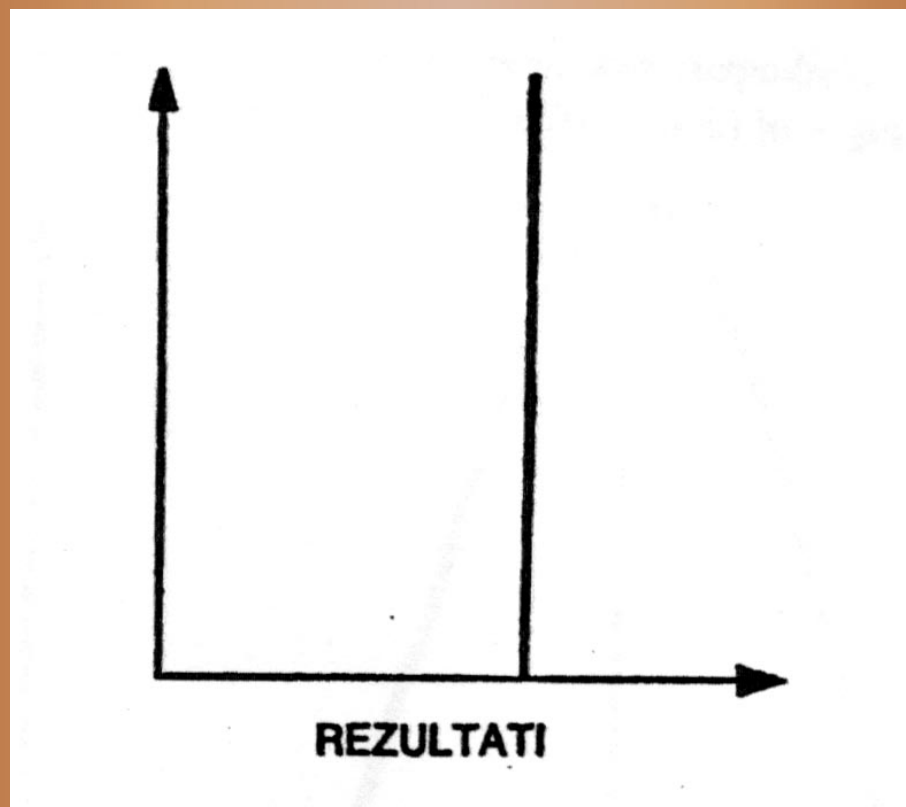


NORMALNA DISTRIBUCIJA – uvjeti nastanka

- najveći broj svojstava u prirodi očituje normalnu distribuciju ukoliko su ispunjeni uvjeti:
 - veliki broj rezultata (mjerjenja)
 - da su sva mjerjenja provedena jednakom metodom i u što sličnijim uvjetima
 - skupina u kojoj mjerimo mora biti homogena po svim svojstvima, a heterogena po svojstvu koje mjerimo

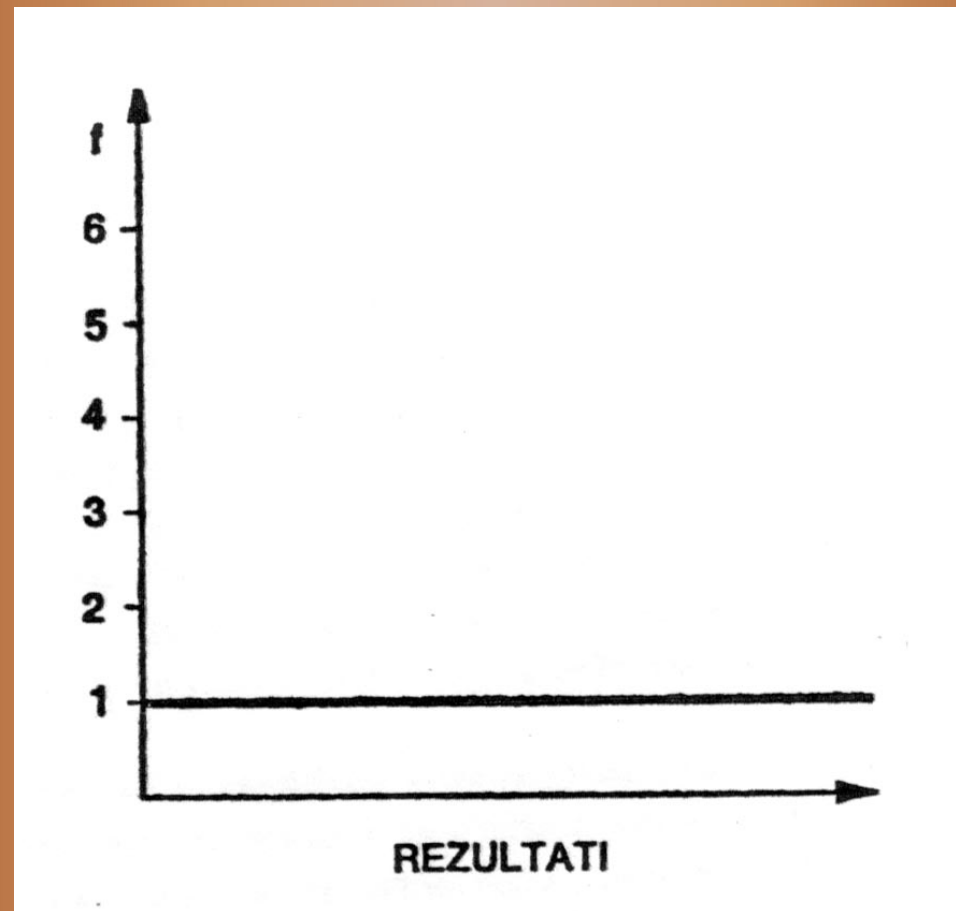
Distribucija vrijednosti varijable

- distribucija kada se pri svakom mjerenju dobije isti rezultat



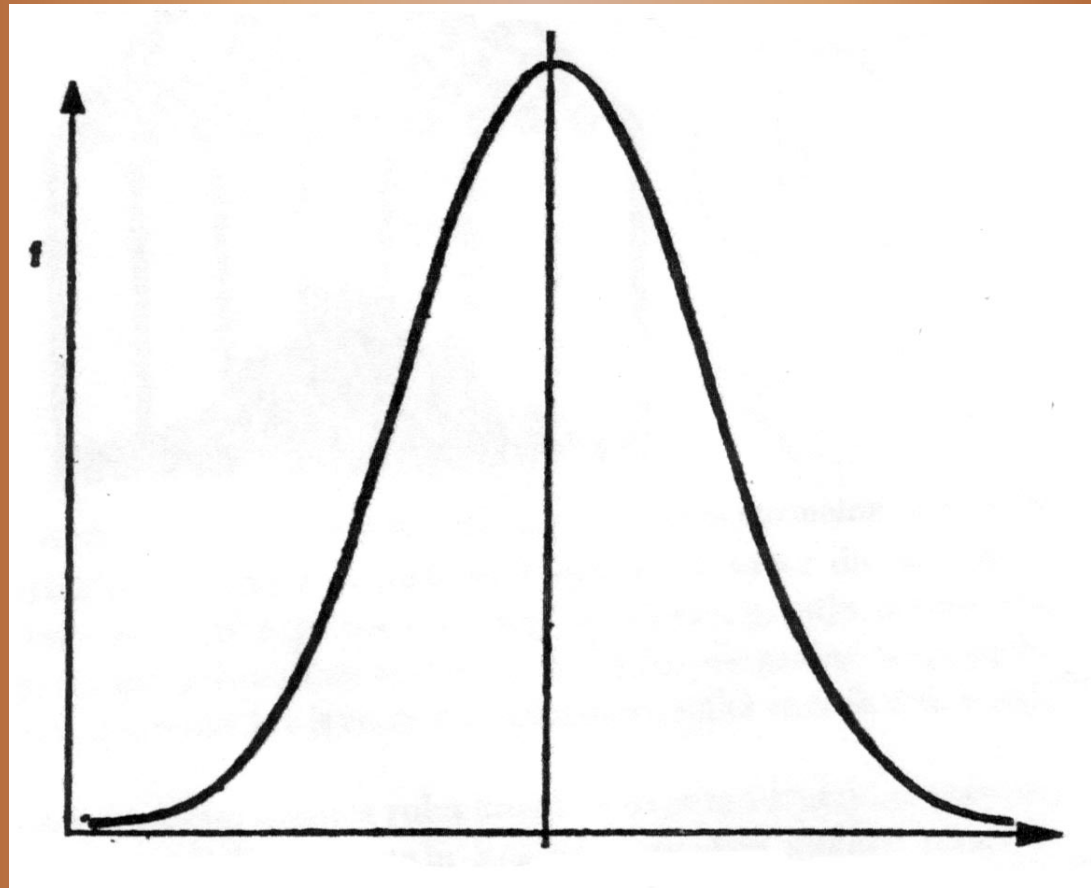
Distribucija vrijednosti varijable

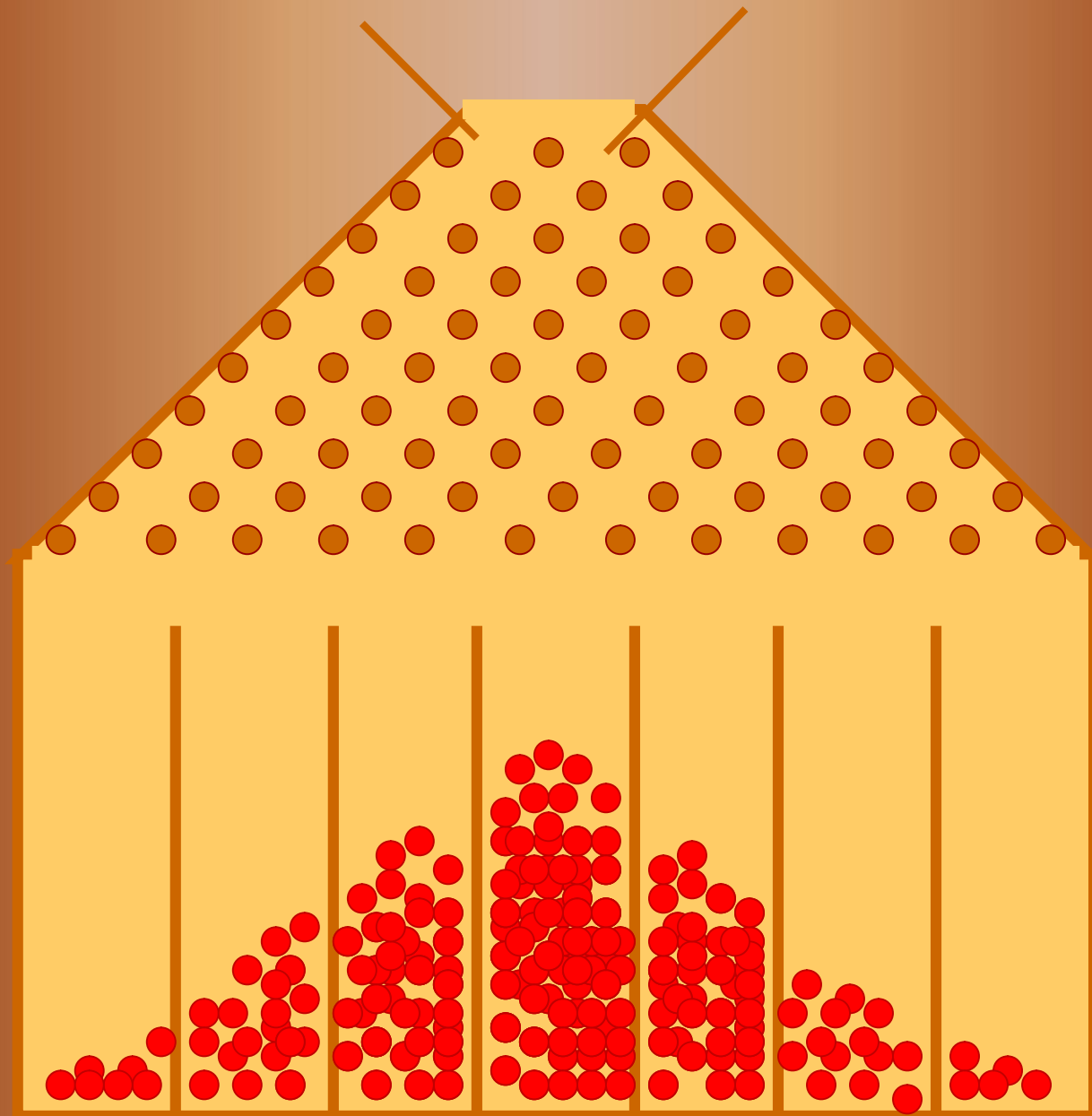
- distribucija kada se pri svakom mjerenju dobije drugačiji rezultat



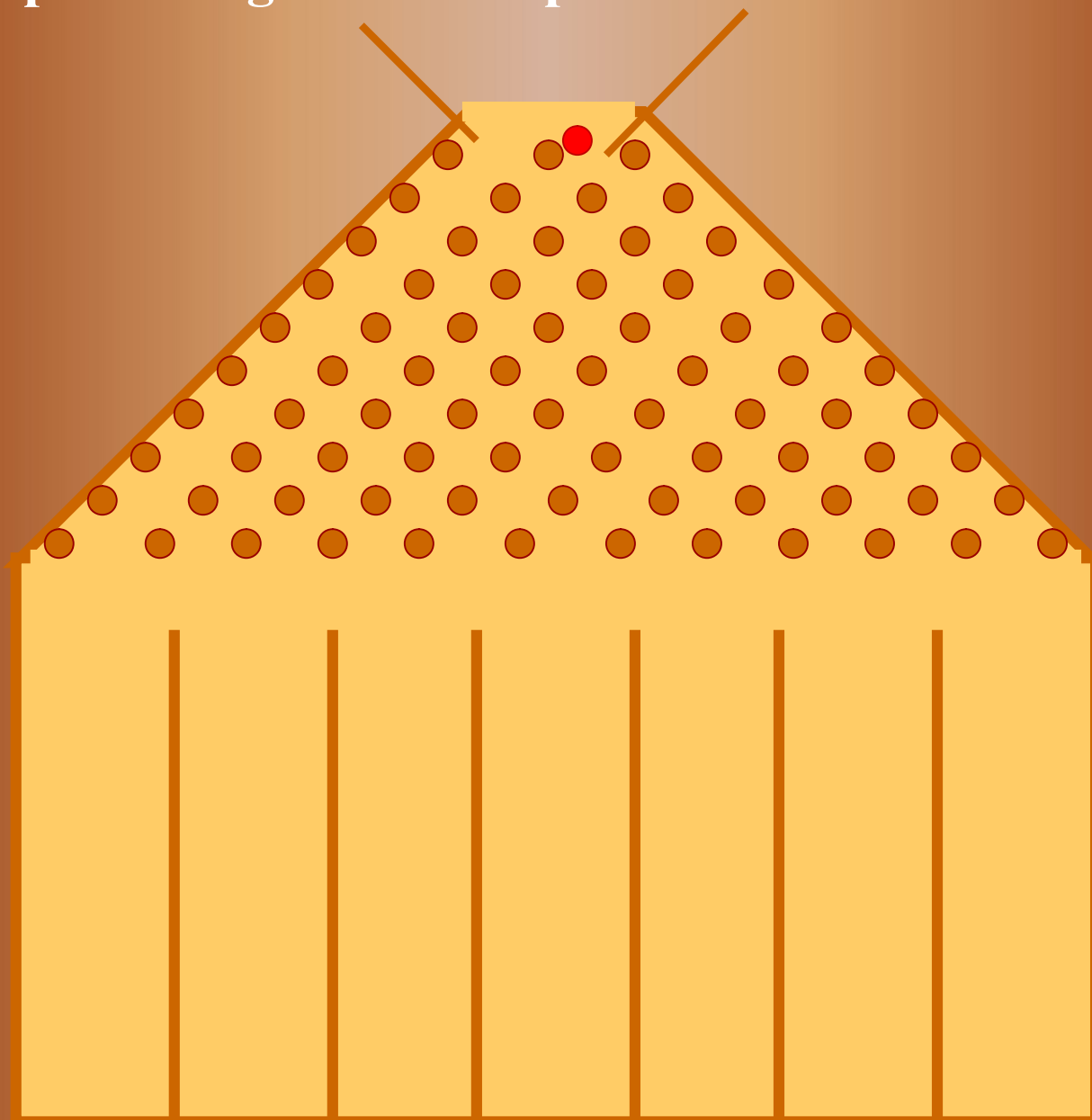
Distribucija vrijednosti varijable

- normalna distribucija



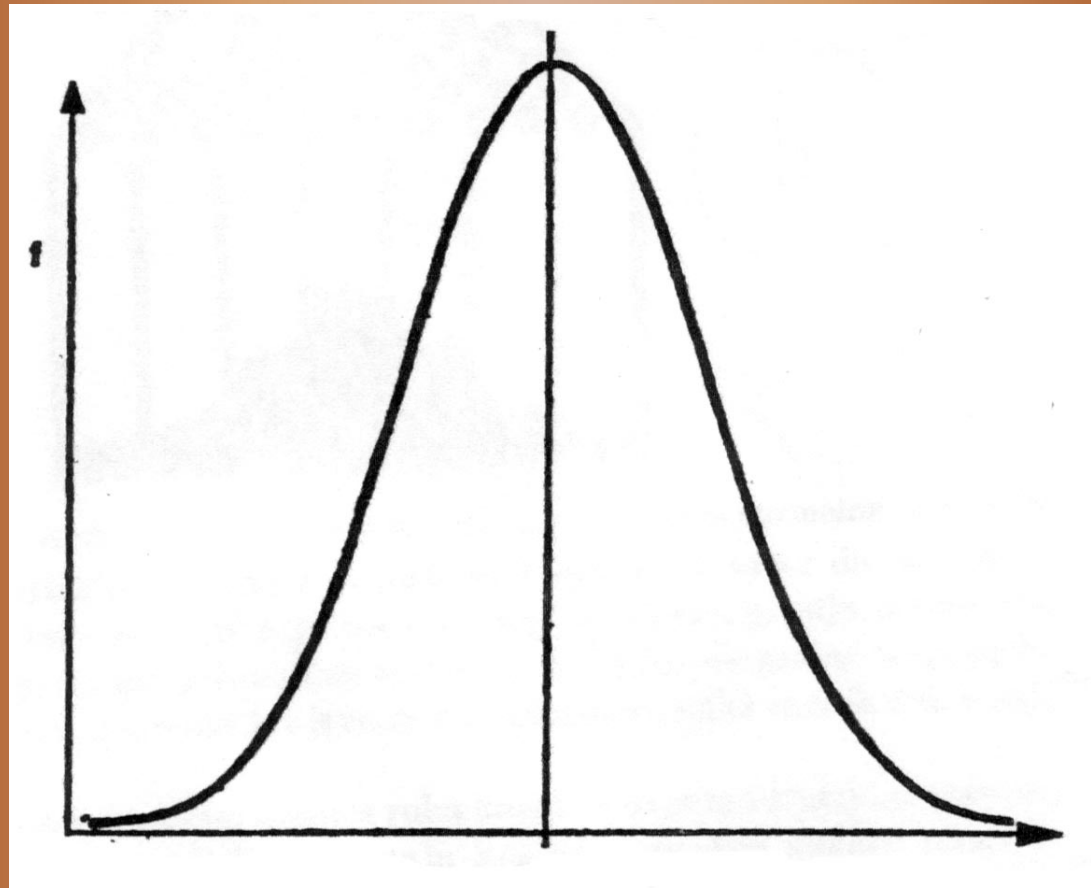


Pokus: raspored kuglica nakon prolaska kroz redove čavlića



Distribucija vrijednosti varijable

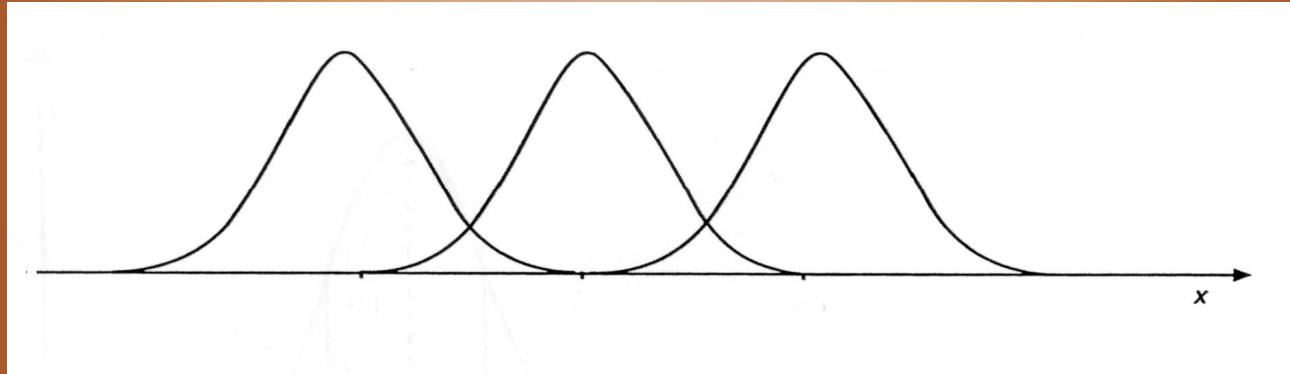
- normalna distribucija



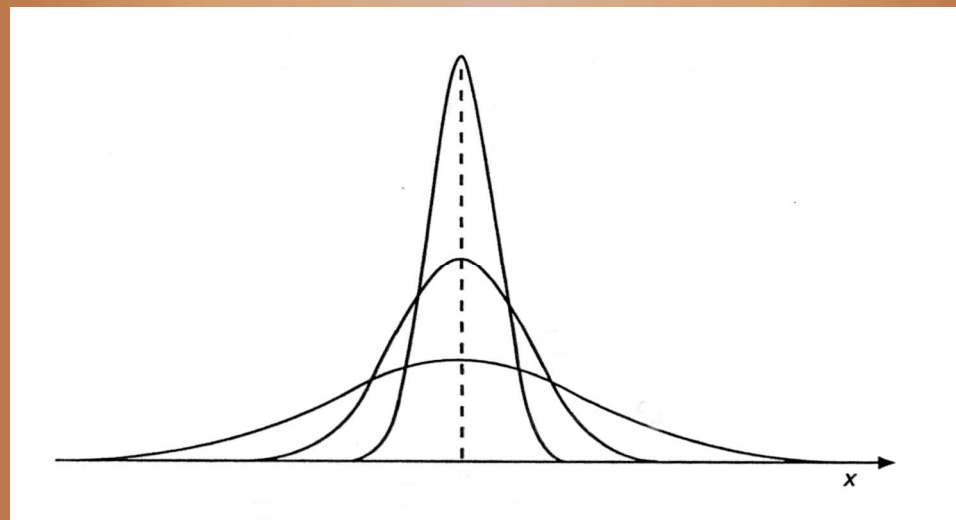
NORMALNA DISTRIBUCIJA – promjena oblika ovisno o veličini standardne devijacije aritmetičke sredine

- ukoliko standardna devijacija (s) ostaje nepromjenjena, povećavanjem srednje vrijednosti (\bar{x}) dolazi do horizontalnog pomicanja krivulje udesno. Smanjivanje srednje vrijednosti (\bar{x}) “pomiče” krivulju horizontalno ulijevo.
- smanjenje standardne devijacije (s) dovodi do uske i visoke krivulje, a povišenje standardne devijacije (s) do široke i niske

NORMALNA DISTRIBUCIJA – promjena oblika ovisno o veličini standardne devijacije aritmetičke sredine



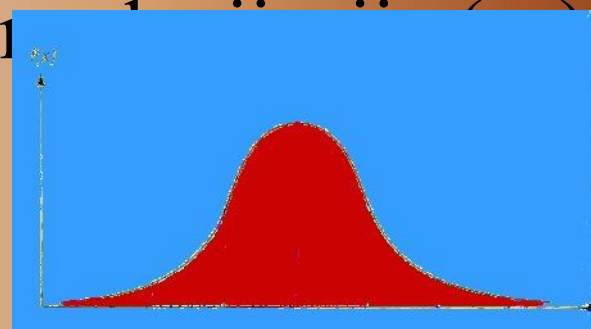
različita srednja vrijednost, jednaka standardna devijacija



jednaka srednja vrijednost, različita standardna devijacija

NORMALNA DISTRIBUCIJA – oblik i karakteristike

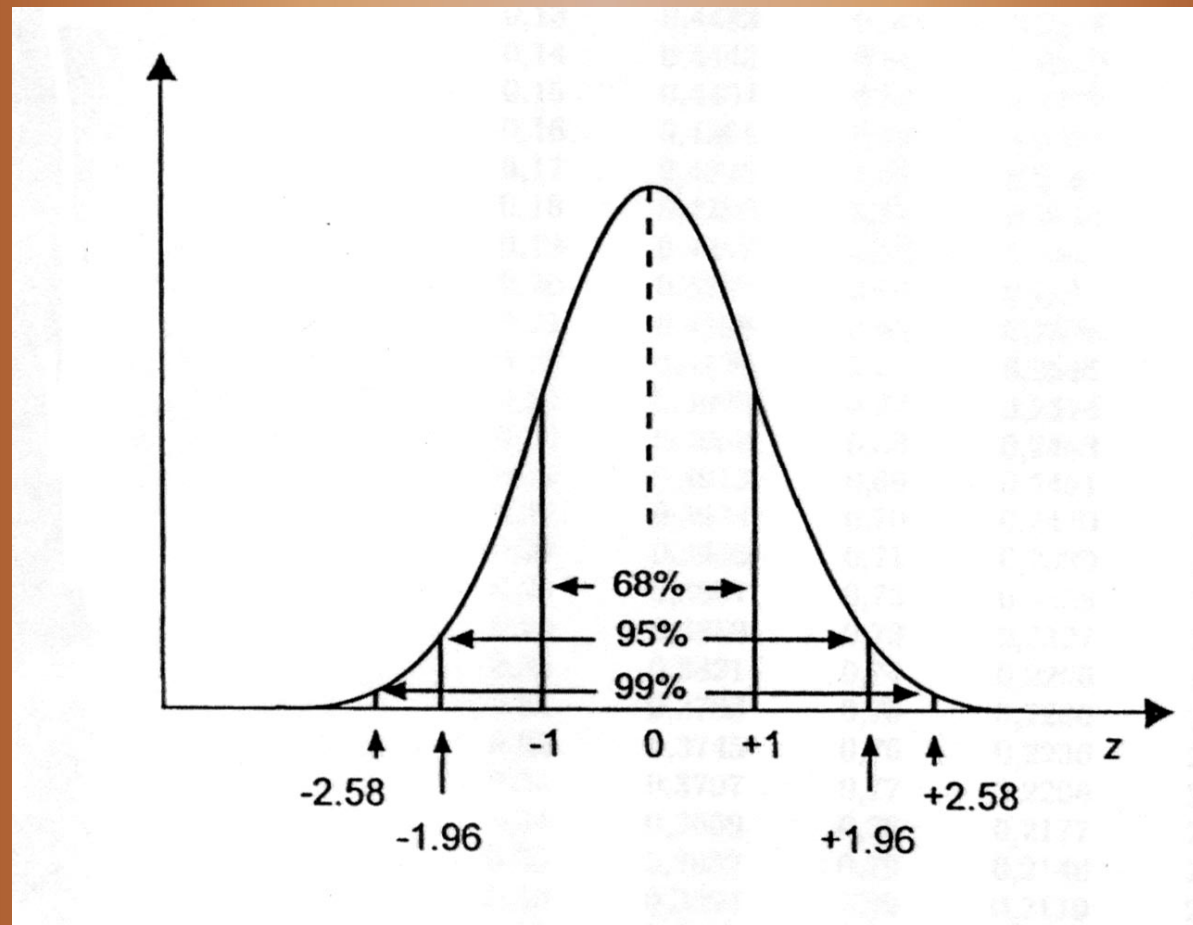
- određena je s dva pokazatelja: srednja vrijednost (\bar{x}) i standardna devijacija (σ)
- unimodalna je
- simetrična je s obzirom na srednju vrijednost (\bar{x}). Krivulja s desne strane od \bar{x} je zrcalna slika krivulje s lijeve strane od \bar{x} . (“zvonasti oblik”)



NORMALNA DISTRIBUCIJA – raspored i položaj pojedinih jedinica skupa

- unutar normalne distribucije nalazi se uvijek isti postotak jedinica statističkog skupa, a udaljenost pojedine jedinice od aritmetičke sredine izračunava se u jedinicama standardne devijacije
- unutar granica $(\bar{x} - s)$ i $(\bar{x} + s)$ nalazi se 68,3 % od svih rezultata
- unutar granica $(\bar{x} - 1,96 s)$ i $(\bar{x} + 1,96 s)$ nalazi se 95,4 % od svih rezultata - raspon 95% rezultata ; “reference range“

NORMALNA DISTRIBUCIJA – raspored i položaj pojedinih jedinica skupa



NORMALNA DISTRIBUCIJA – praktična primjena

- normalna distribucija je osnova za izračunavanje vjerojatnosti određenog rezultata u nizu mjerenja

VJEROJATNOSTI ODREĐENOG REZULTATA (uz pretpostavku normalne distribucije)

- osniva se na
 - STANDARDIZIRANOM ODSTUPANJU (z)
 - TABLICI KOJA POKAZUJE POVRŠINU ISPOD NORMALNE KRIVULJE ZA ODREĐENO STANDARDIZIRANO ODSTUPANJE (z)

Tablica

površine ispod
normalne krivulje

P = površina standardne
normalne raspodjelne
krivulje od zadane
vrijednosti X do kraja
krivulje. Površina cijele
krivulje = 1

$$X - \bar{X}$$

$$Z = \frac{\quad}{\quad}$$

S

z	P	z	P	z	P
0,00	0,5000	0,44	0,3300	0,88	0,1894
0,01	0,4960	0,45	0,3264	0,89	0,1867
0,02	0,4920	0,46	0,3228	0,90	0,1841
0,03	0,4880	0,47	0,3192	0,91	0,1814
0,04	0,4840	0,48	0,3156	0,92	0,1788
0,05	0,4801	0,49	0,3121	0,93	0,1762
0,06	0,4761	0,50	0,3085	0,94	0,1736
0,07	0,4721	0,51	0,3050	0,95	0,1711
0,08	0,4681	0,52	0,3015	0,96	0,1685
0,09	0,4641	0,53	0,2981	0,97	0,1660
0,10	0,4602	0,54	0,2946	0,98	0,1635
0,11	0,4562	0,55	0,2912	0,99	0,1611
0,12	0,4522	0,56	0,2877	<u>1,00</u>	<u>0,1587</u>
0,13	0,4483	0,57	0,2843	1,05	0,1469
0,14	0,4443	0,58	0,2810	1,10	0,1357
0,15	0,4404	0,59	0,2776	1,15	0,1251
0,16	0,4364	0,60	0,2743	1,20	0,1151
0,17	0,4325	0,61	0,2709	1,25	0,1056
0,18	0,4286	0,62	0,2676	1,30	0,0968
0,19	0,4247	0,63	0,2643	1,35	0,0885
0,20	0,4207	0,64	0,2611	1,40	0,0808
0,21	0,4168	0,65	0,2578	1,45	0,0735
0,22	0,4129	0,66	0,2546	1,50	0,0668
0,23	0,4090	0,67	0,2514	1,55	0,0606
0,24	0,4052	0,68	0,2483	1,60	0,0548
0,25	0,4013	0,69	0,2451	1,65	0,0495
0,26	0,3974	0,70	0,2420	1,70	0,0446
0,27	0,3936	0,71	0,2389	1,75	0,0401
0,28	0,3897	0,72	0,2358	1,80	0,0359
0,29	0,3859	0,73	0,2327	1,85	0,0322
0,30	0,3821	0,74	0,2296	1,90	0,0287
0,31	0,3783	0,75	0,2266	1,95	0,0256
0,32	0,3745	0,76	0,2236	2,00	0,0228
0,33	0,3707	0,77	0,2206	2,10	0,0179
0,34	0,3669	0,78	0,2177	2,20	0,0139
0,35	0,3632	0,79	0,2148	2,30	0,0107
0,36	0,3594	0,80	0,2119	2,40	0,00820
0,37	0,3557	0,81	0,2090	2,50	0,00621
0,38	0,3520	0,82	0,2061	2,60	0,00466
0,39	0,3483	0,83	0,2033	2,70	0,00347
0,40	0,3446	0,84	0,2005	2,80	0,00256
0,41	0,3409	0,85	0,1977	2,90	0,00187
0,42	0,3372	0,86	0,1949	3,00	0,00135
0,43	0,3336	0,87	0,1922	3,50	0,000233
				4,00	0,0000317

- primjer:

Prosječna masa teladi namijenjene za tov iznosi 170 kg, a standardna devijacija 10 kg. Utvrdite:

1. S kojom proporcijom se pojavljuje telad čija je masa veća od 180 kg?

2. Koliki je udio (broj) teladi čija je masa tijela između 170 i 180 kg?

$$N = 250; \quad X_1 = 180; \quad \bar{X} = 170; \quad s = 10$$

$$Z = \frac{X_1 - \bar{X}}{s} = \frac{180 - 170}{10} = 1$$

primjer: nastavak

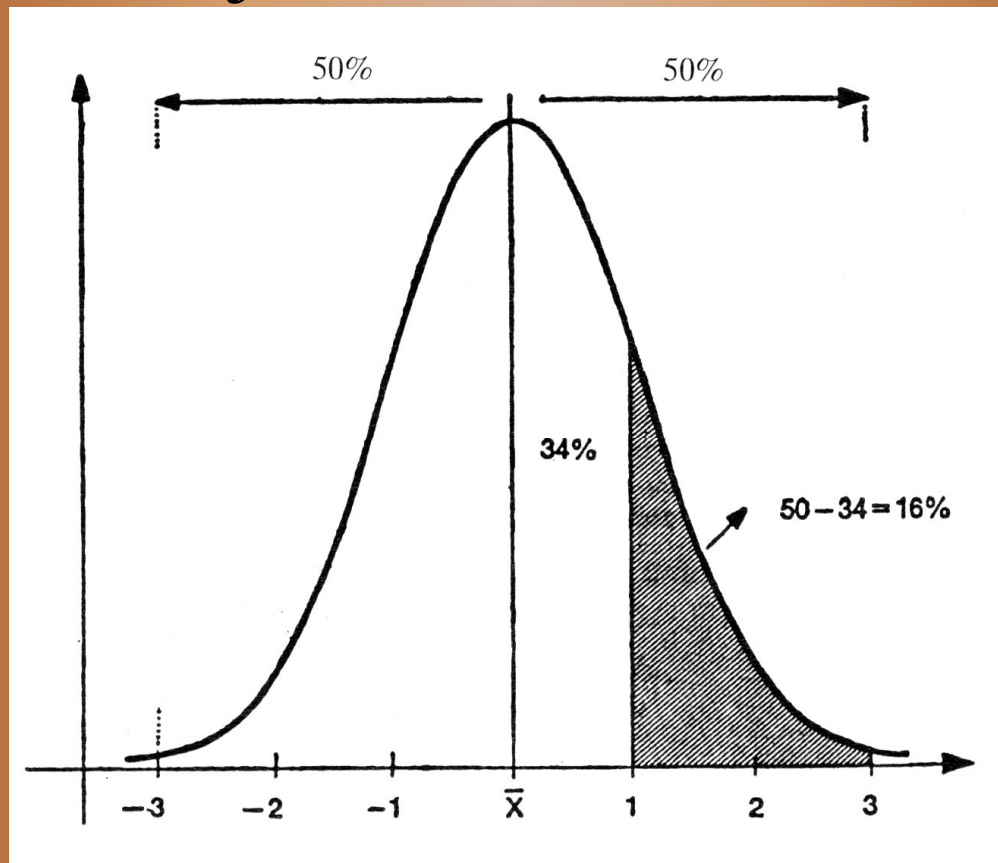
1. Vrijednost koju u “Tablici normalne distribucije” nalazimo na $z = 1$ iznosi 0,1587 što znači da je to površina pod normalnom krivuljom, odnosno da udio teladi s tjelesnom masom većom od 180 kg iznosi približno 16% ili 40 teladi.

2. Udio teladi čija je tjelesna masa između 170 i 180 kg iznosi 0,3413% ili 85 teladi, jer $x + 1s$ uvijek sadrži 34,13% podataka.

(Polovica površine ispod normalne krivulje iznosi 0,500 od čega se odbije 0,3413, odnosno udio teladi do 180 kg mase te tako ostaje 0,1587 ili priližno 16% onih s tjelesnom masom preko 180 kg.)

primjer: nastavak

ako neki rezultat pada točno na $+1s$, onda u čitavoj populaciji ima oko 84% slabijih i oko 16% boljih rezultata



PROVJERA DA LI DISTRIBUCIJA KONKRETNIH PODATAKA ODGOVARA NORMALNOJ DISTRIBUCIJI POČETNI JE KORAK SVAKE STATISTIČKE OBRADE !!!

Postupci provjere:

- subjektivno
 - histogram
 - normal plot
- testovi
 - mjere asimetrije i zaobljenosti
 - Shapiro-Wilk W test
 - Kolmogorov-Smirnov test