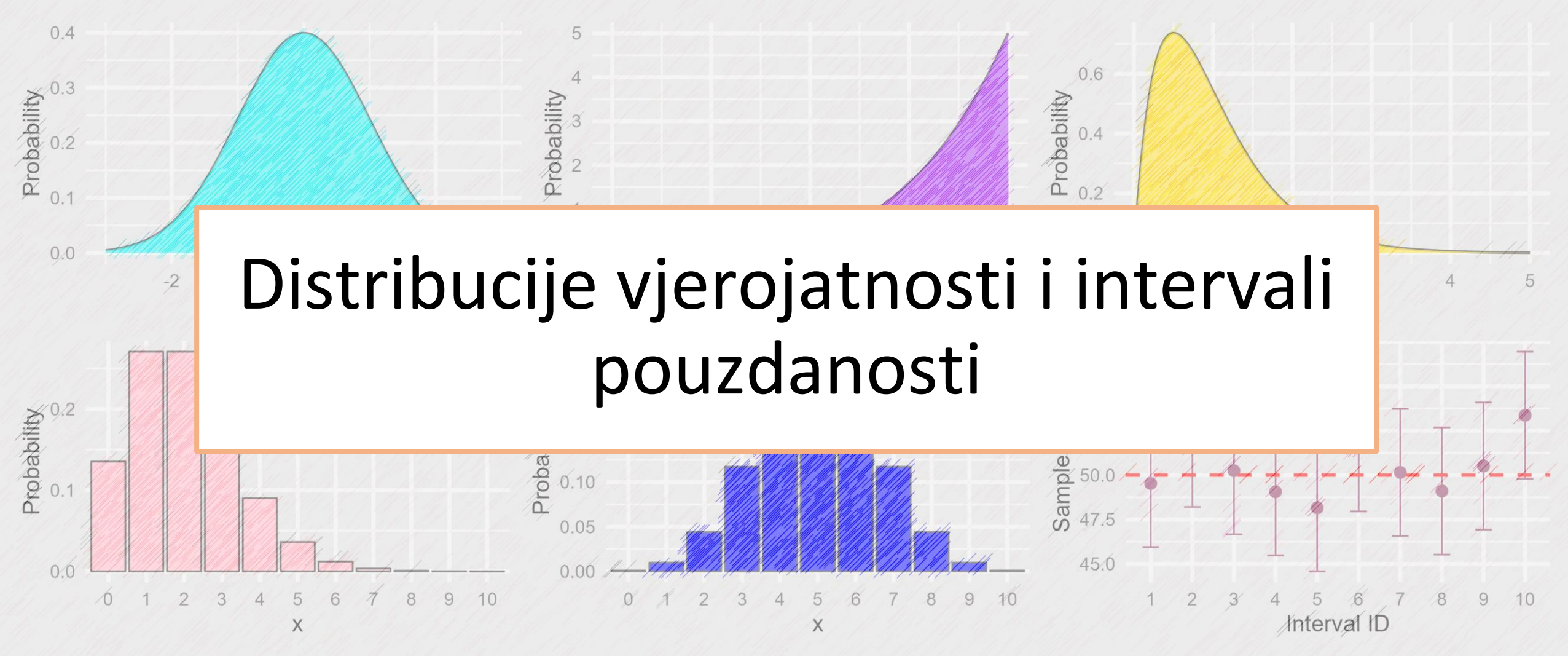


Distribucije vjerojatnosti i intervali pouzdanosti



Izv.prof. Rosa Karlić
Predavanje 6, MZIRuB 2025/2026
10.12.2025.

Populacije i uzorci

- Populacija (od interesa) : cijela skupina ispitanika o kojima želimo nešto zaključiti
- Uzorak – podskup populacije
- **Parametar** – karakteristika populacije (npr. srednja vrijednost populacije, μ)
- **Statistika** – bilo koja funkcija ispitanika u nasumičnom uzorku (npr. srednja vrijednost uzorka, $\bar{x}$)
- Poželjno je odabrati **nasumičan uzorak** iz populacije kako u analizu ne bismo uveli **pristranost**

Slučajna varijabla

- funkcija koja dodjeljuje numeričke vrijednosti različitim događajima u prostoru uzorka
- npr. studija u kojoj istraživači izlažu bakterijski soj mutagenu i zatim mjere broj mutacija koje se javljaju u određenom genu nakon određenog broja generacija
- može poprimiti samo pozitivne vrijednosti

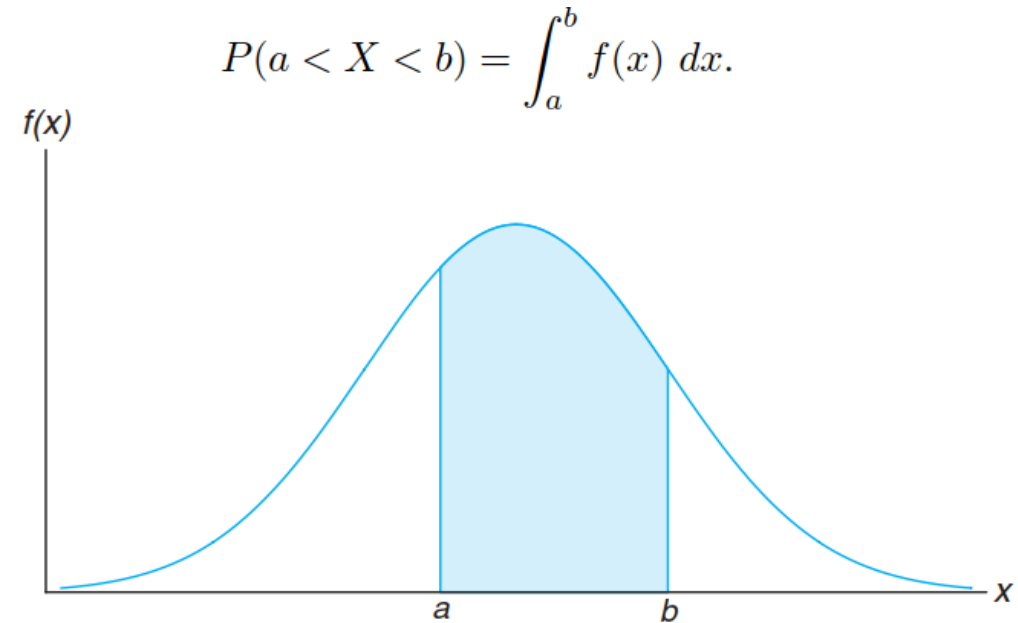
Ishod	Pridružena numerička vrijednost (x)
0	5
1	12
2	7
...	...

Slučajna varijabla

- Slučajne varijable mogu biti:
 - Diskretne
 - Starost pacijenata u godinama $X = \{0, 1, 2, 3, \dots, N\}$
 - Broj mutacija u DNA lancu $X = \{0, 1, 2, 3, \dots, N\}$
 - Kontinuirane
 - Visina pacijenta $X = (0, M)$
 - Tjelesna temperatura pacijenta $X = (M, N)$
- Opisujemo ih distribucijama vjerojatnosti – vjerojatnost da će slučajna varijabla poprimiti određenu vrijednost ili se nalaziti u određenom intervalu

Kontinuirane slučajne varijable

- Mogu poprimiti beskonačno mnogo vrijednosti
- Područje vrijednosti – interval na brojevnom pravcu ili cijeli brojevni pravac
- Određujemo vjerojatnost da će se vrijednost kontinuirane slučajne varijable nalaziti unutar nekog intervala (vjerojatnost da će poprimiti točno određenu vrijednost je 0)
- Opisuju se funkcijama gustoće vjerojatnosti



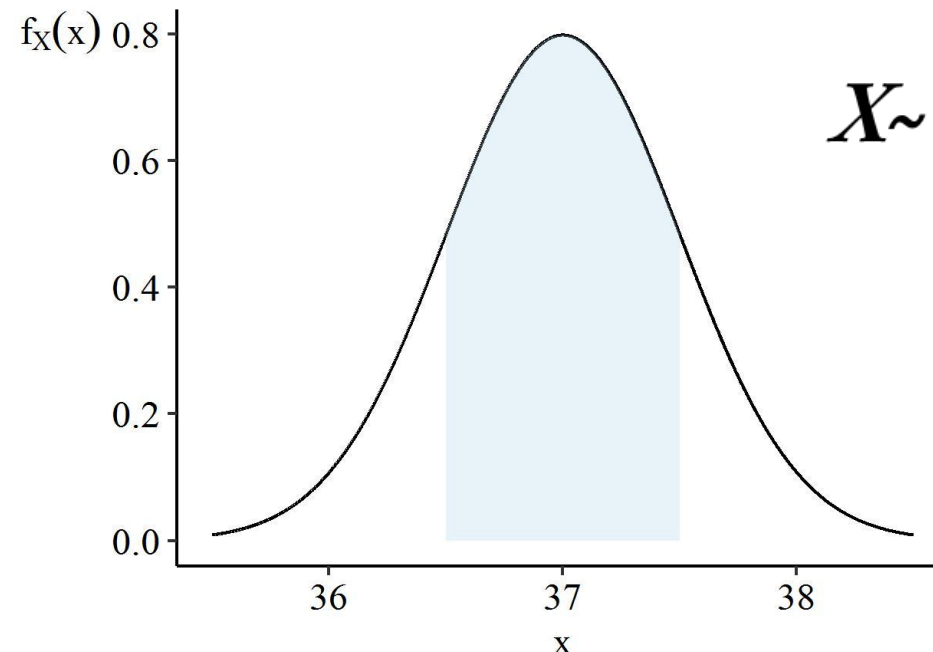
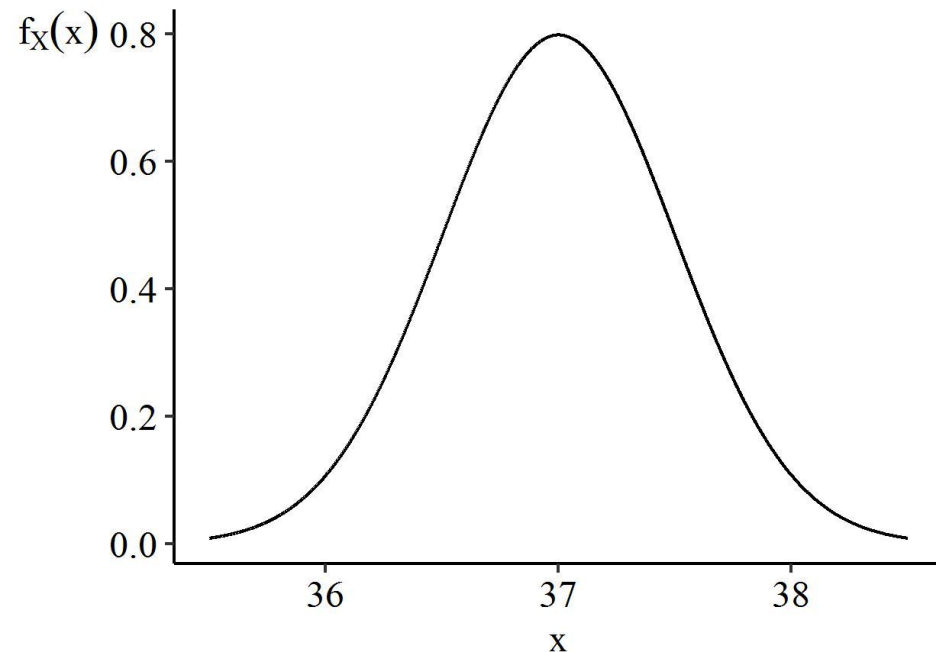
Walpole, R. E., Myers, R. H., Myers, S. L. and Ye, K. (2012)
Probability and Statistics for Engineers and Scientists

Kontinuirane slučajne varijable - primjer

- Tjelesna temperatura ispitanika
- μ (srednja vrijednost) and σ (standardna devijacija) određuju lokaciju i oblik distribucije

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$X \sim N(\mu, \sigma^2)$$



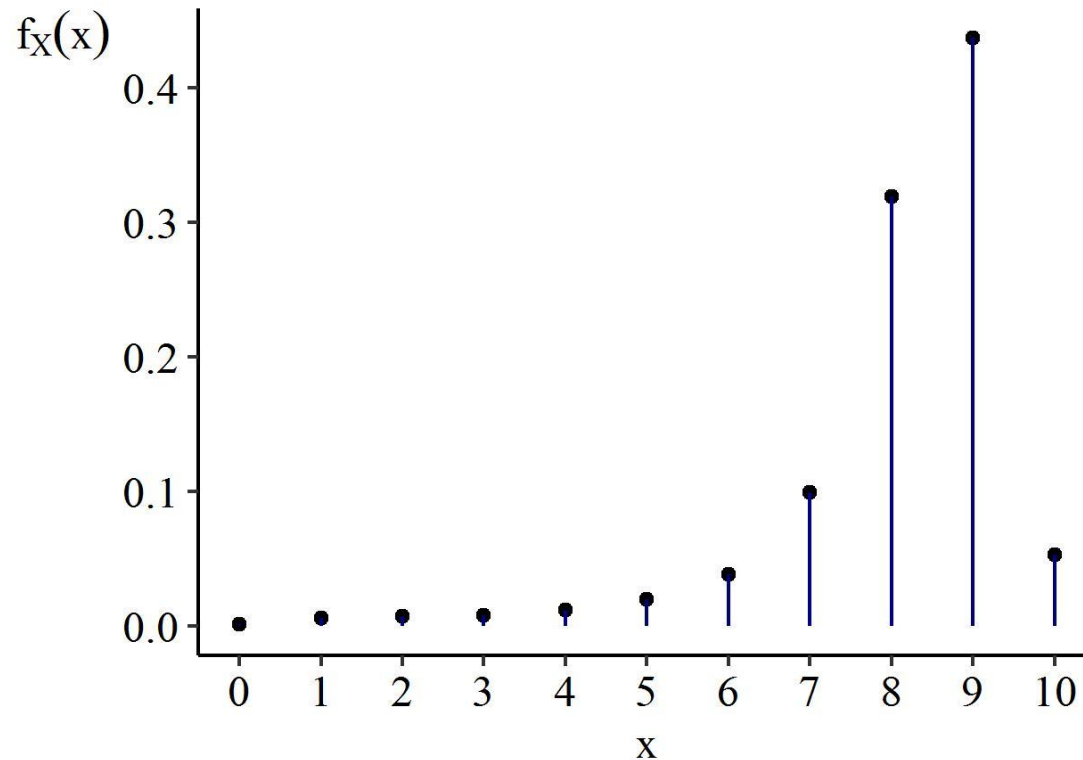
Diskretne slučajne varijable

- Mogu poprimiti prebrojivo mnogo diskretnih vrijednosti
- Svaka vrijednost ima konačnu vjerojatnost
- Opisuju se funkcijama mase vjerojatnosti

$$f_X(x) = P(X = x)$$

Diskretne slučajne varijable - primjer

- Apgar ocjena



$F_X(x)$	x
0.001	0
0.006	1
0.007	2
0.008	3
0.012	4
0.020	5
0.038	6
0.099	7
0.319	8
0.437	9
0.053	10

Kumulativna funkcija distribucije

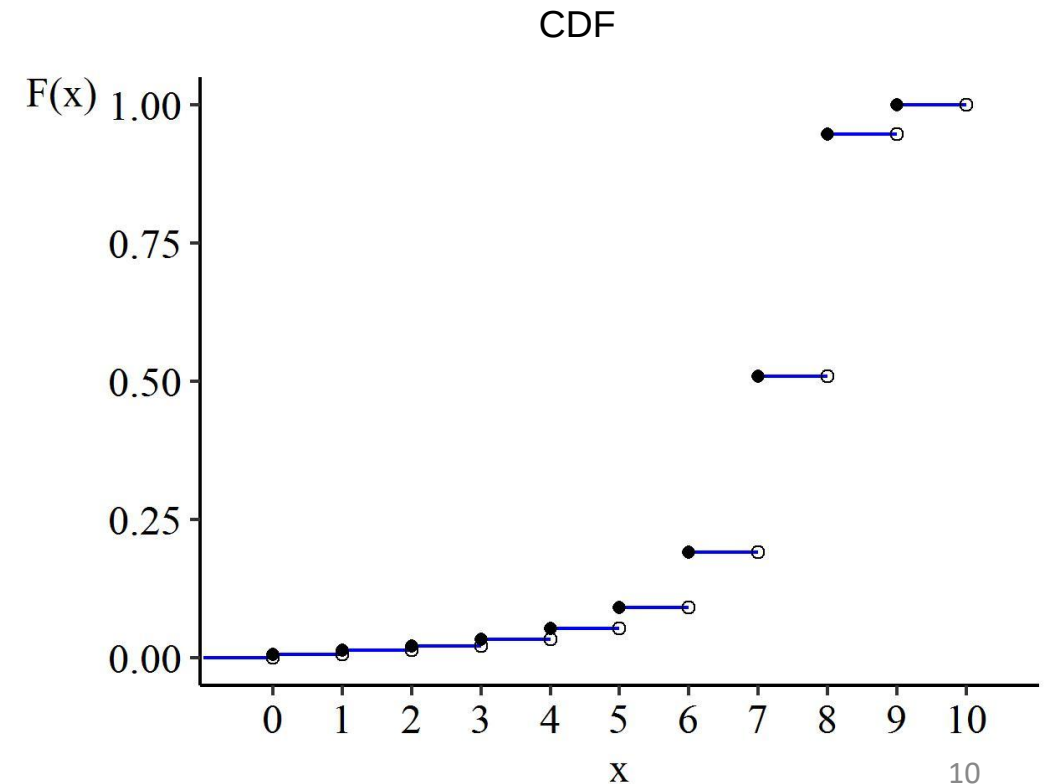
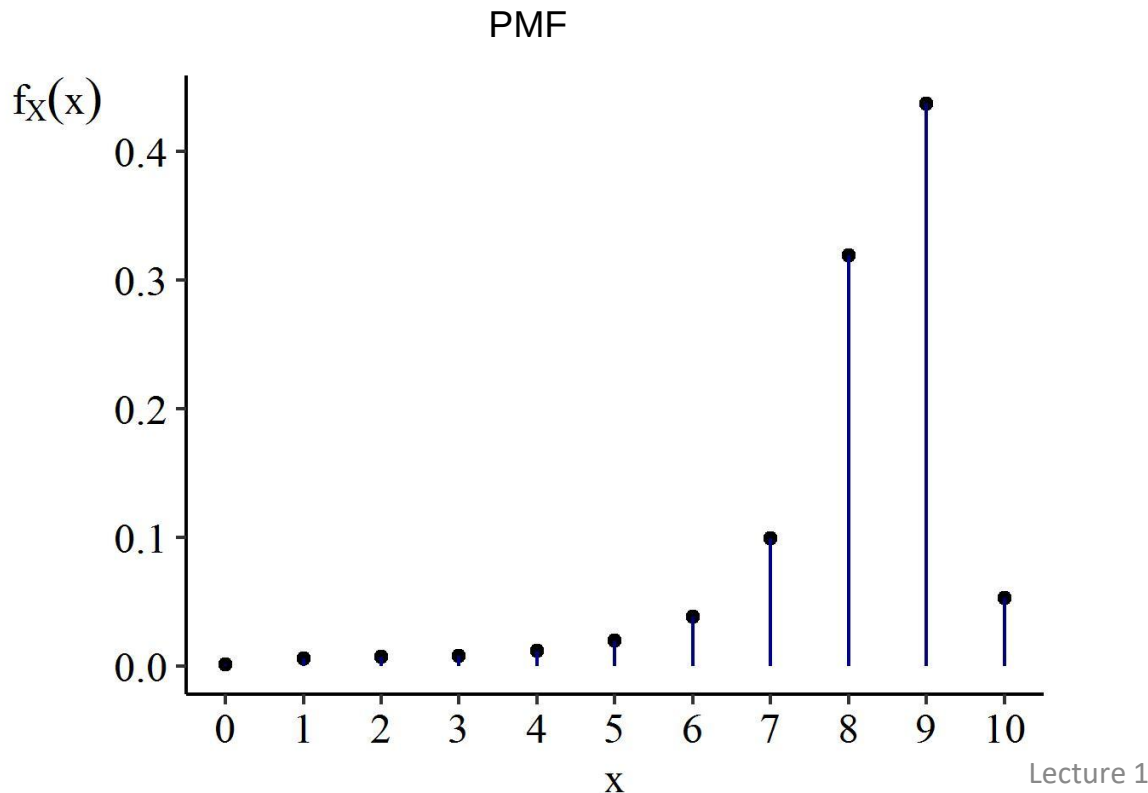
- Cumulative distribution function (CDF)
- Vjerojatnost da je vrijednost slučajne varijable X manja ili jednaka od x

$$F(x) = P(X \leq x)$$

Kumulativna funkcija distribucije

- Diskretne slučajne varijable
- Monotono se povećava od 0 do 1

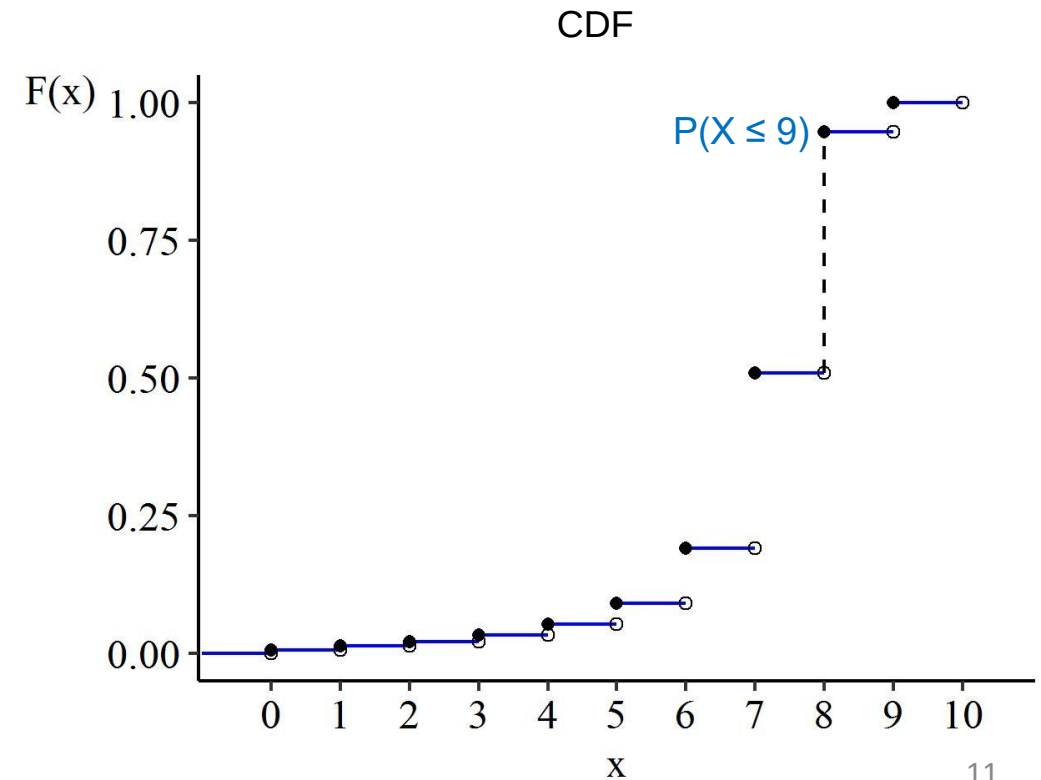
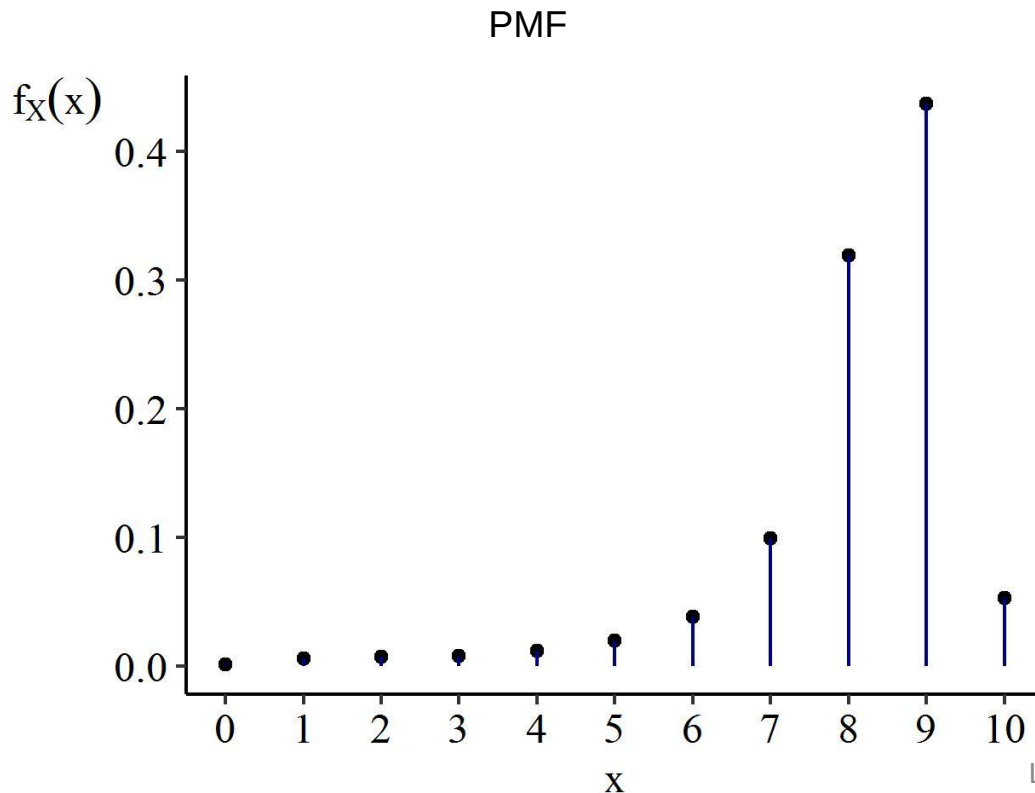
$$F(x) = P(X \leq x)$$



Kumulativna funkcija distribucije

- Diskretne slučajne varijable
- Monotono se povećava od 0 do 1

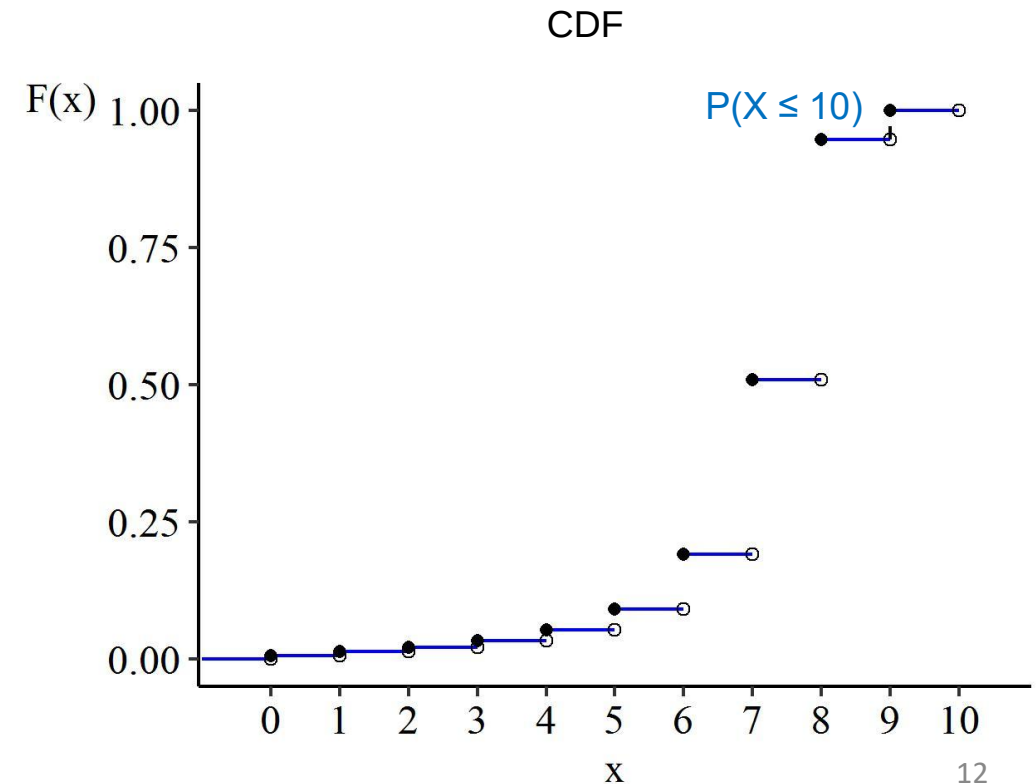
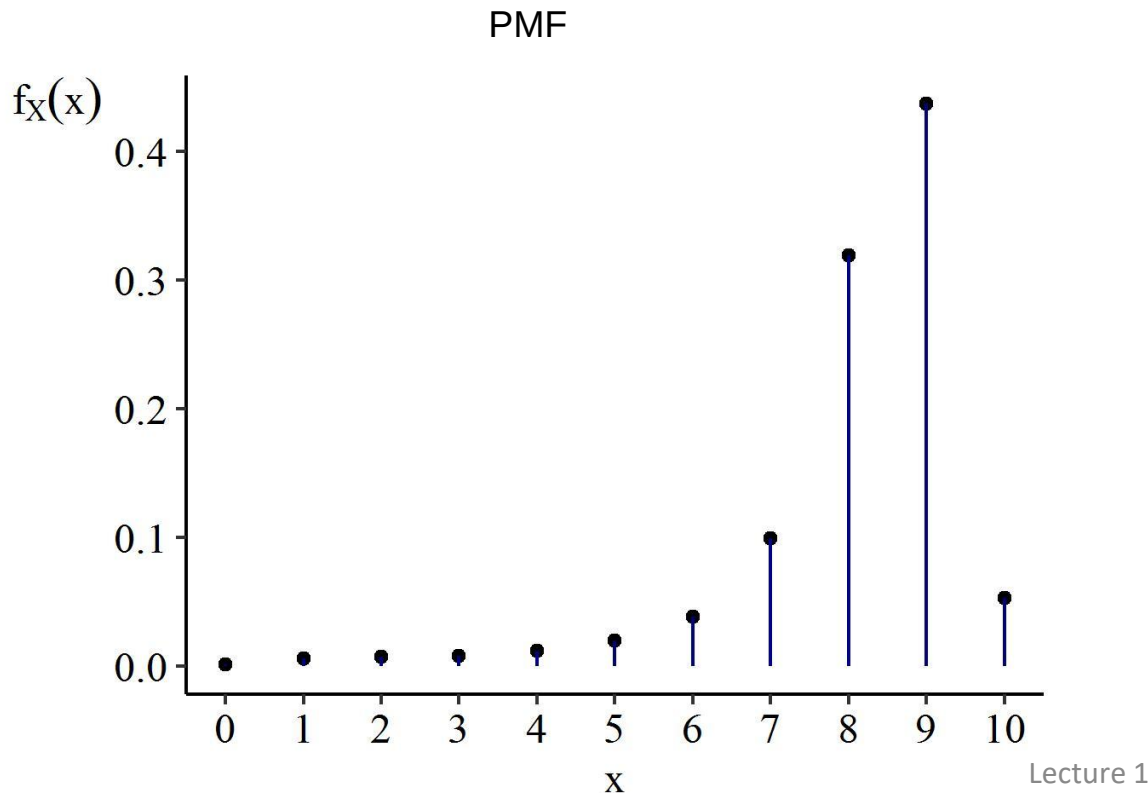
$$F(x) = P(X \leq x)$$



Kumulativna funkcija distribucije

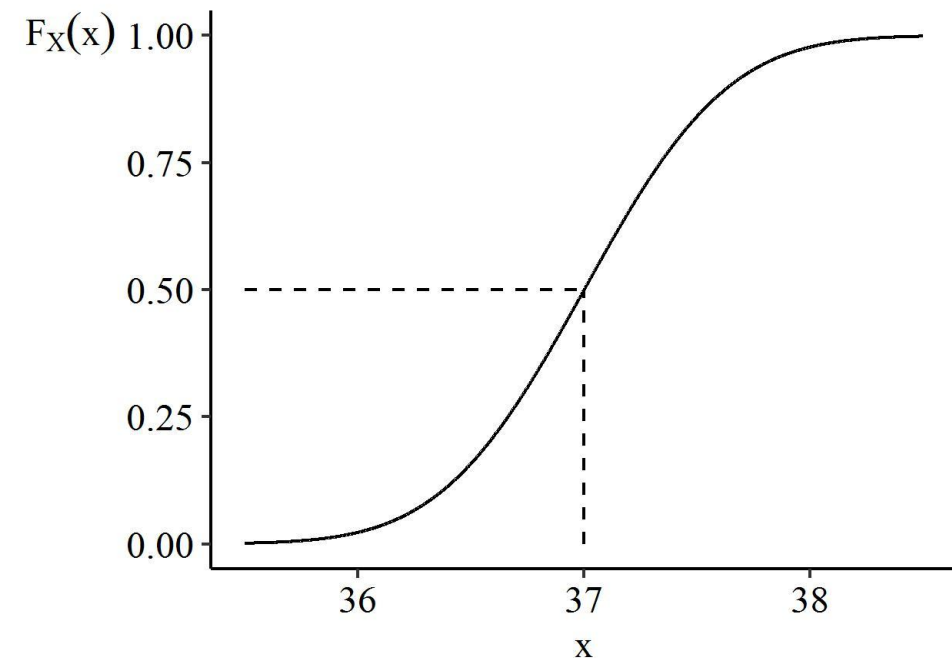
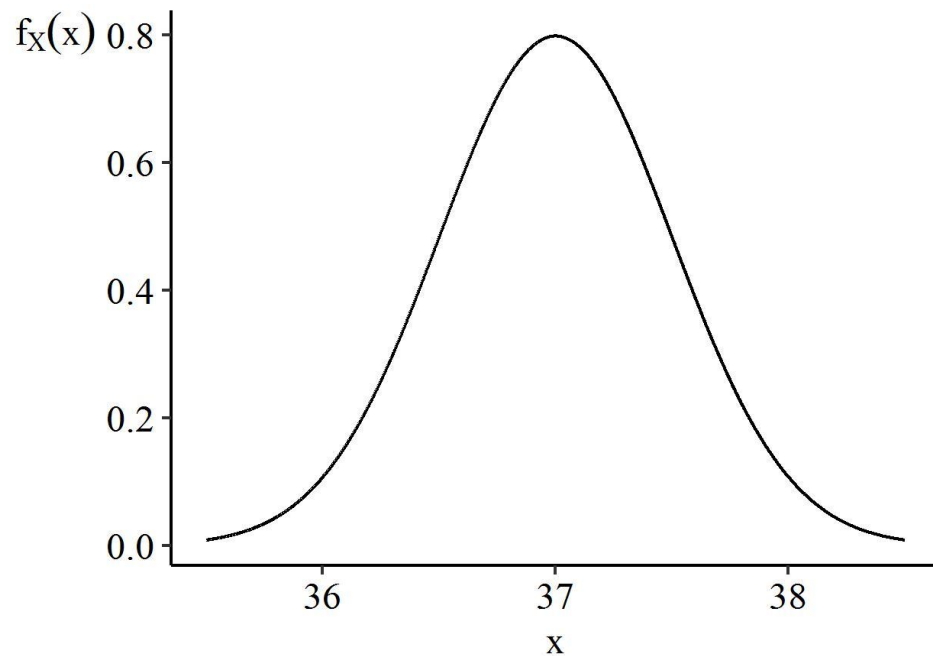
- Diskretne slučajne varijable
- Monotono se povećava od 0 do 1

$$F(x) = P(X \leq x)$$



Kumulativna funkcija distribucije

- Kontinuirane slučajne varijable $F(x) = P(X \leq x)$
- Monotono se povećava od 0 do 1



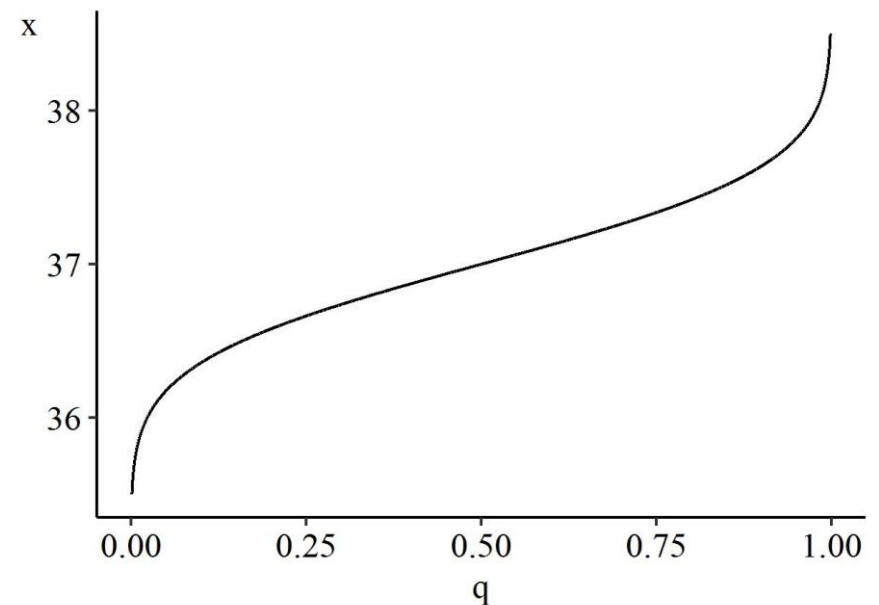
Kvantilna funkcija (Inverse CDF)

- Ako je X slučajna varijabla sa CDF F . Kvantilna funkcija (inverse CDF) je definirana kao:

$$\text{za} \quad F^{-1}(q) = \inf \{x : F(x) \geq q\}$$

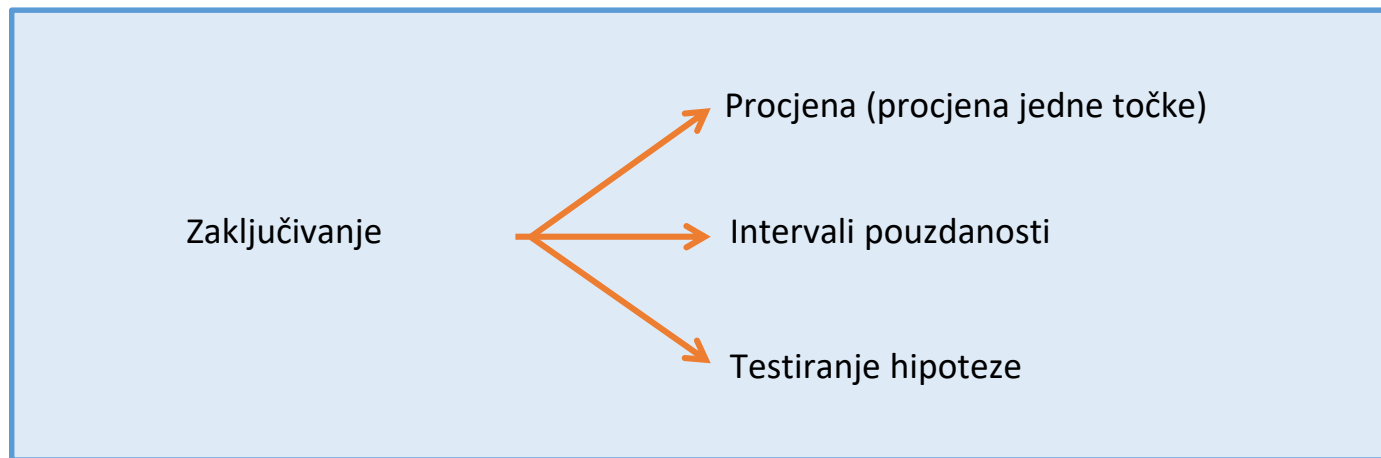
$$q \in [0,1]$$

- $F^{-1}(1/4)$ – prvi kvartil
- $F^{-1}(1/2)$ - medijan
- $F^{-1}(3/4)$ – treći kvartil



Statističko zaključivanje

- Proces korištenja podataka za donošenje zaključaka o distribuciji koja je generirala podatke ili nekoj značajki te distribucije, kao što je srednja vrijednost
- Dominantni pristupi: frekventističko zaključivanje i Bayesovo zaključivanje



Procjena parametara

Zaključujemo nešto o populaciji na temelju informacija dobivenih iz uzorka

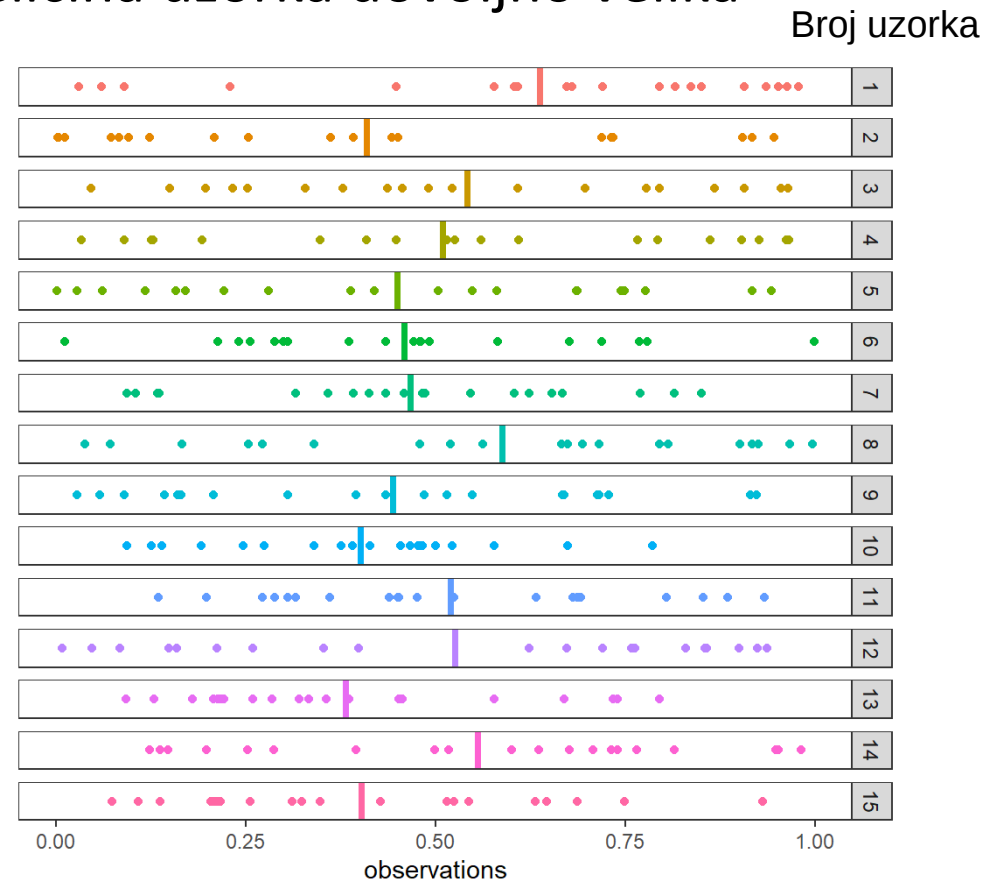
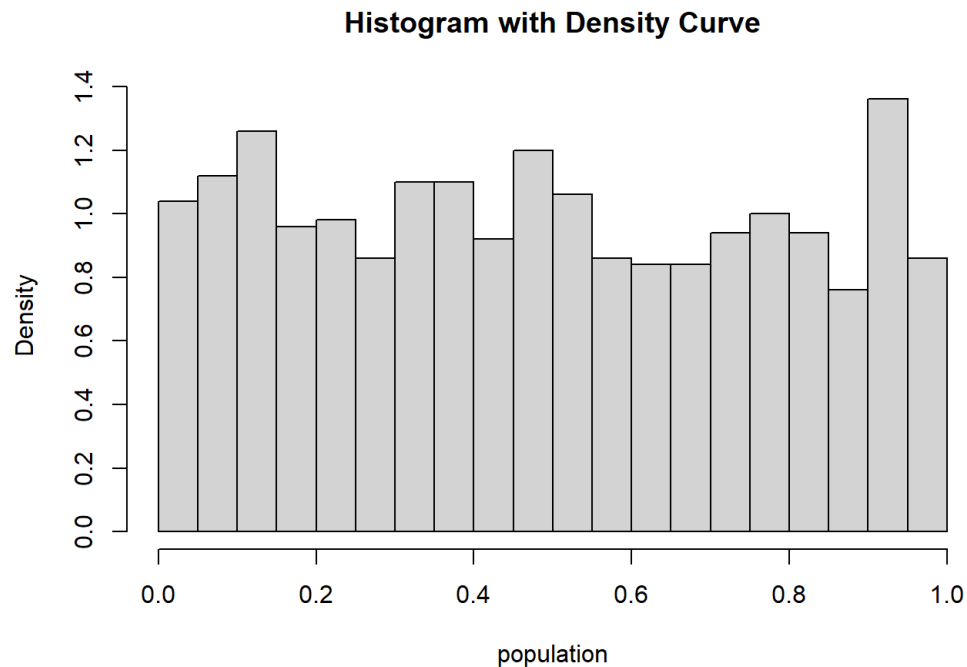
- Statistike se koriste kao procjene parametara
- Procjena u jednoj točki
- Procjena u obliku intervala

Intervali pouzdanosti

- Koristi se za izražavanje preciznosti i nesigurnosti povezanih s određenom metodom uzorkovanja.
 - Sastoji se od tri dijela:
 - Razina pouzdanosti - opisuje nesigurnost metode uzorkovanja
 - Statistika
 - Margina greške
- Definiraju procjenu intervala koji opisuje preciznost metode
- Razina pouzdanosti - koliko čvrsto vjerujemo da će neka metoda uzorkovanja proizvesti interval pouzdanosti koji uključuje stvarni parametar populacije.

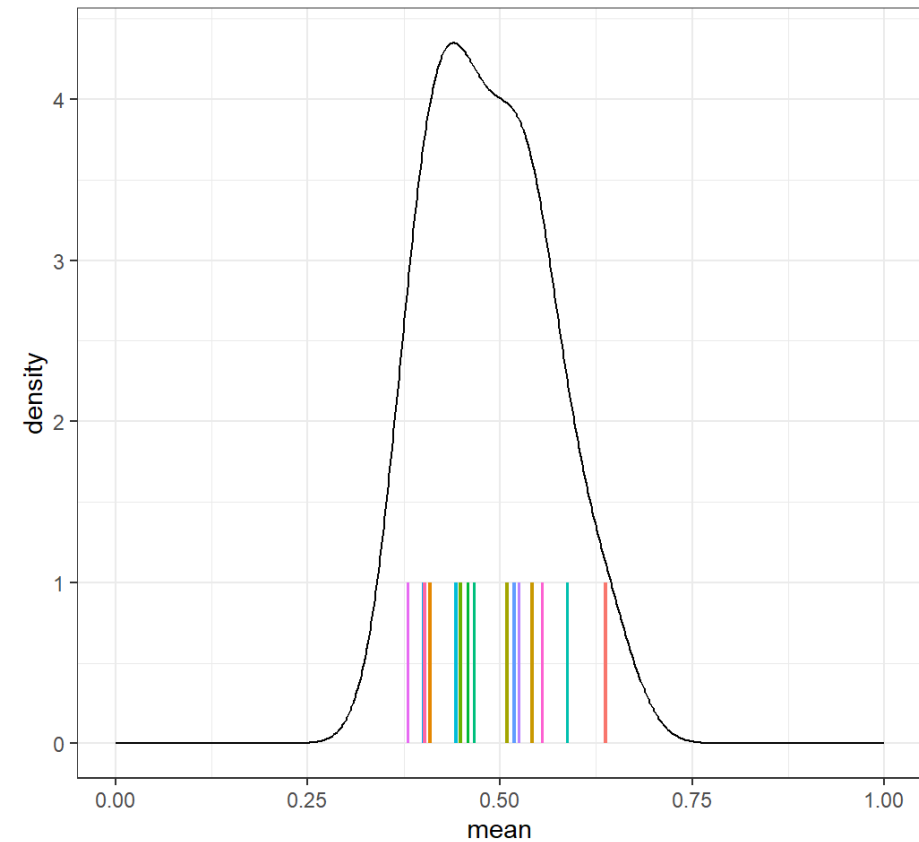
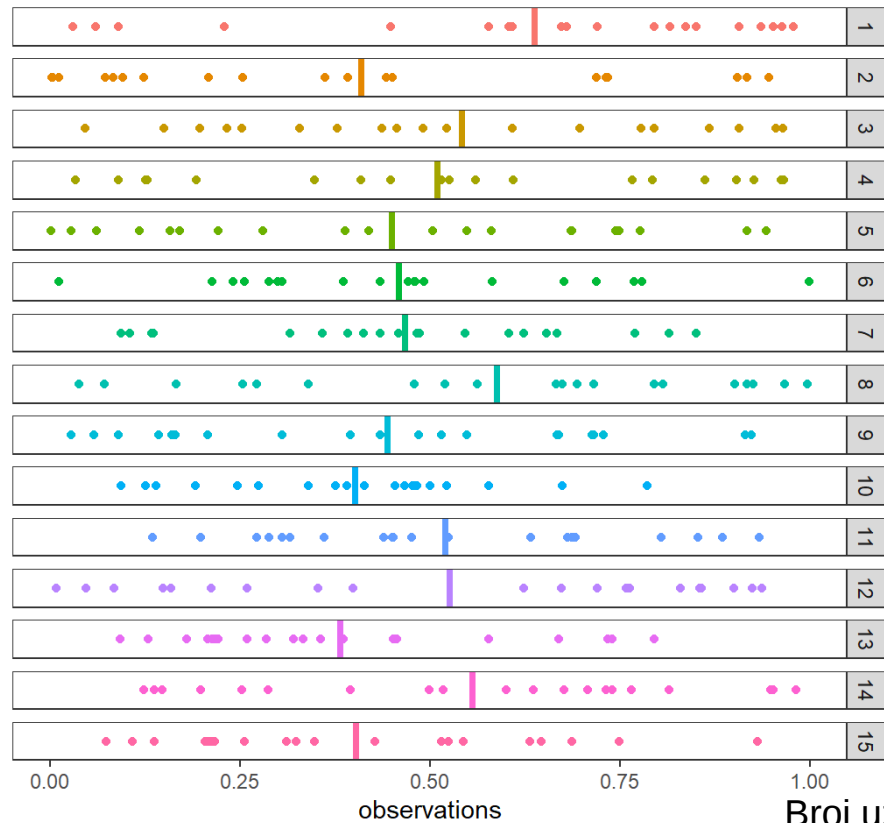
Centralni granični teorem

- Distribucija procjena statistika zbroja ili prosjeka i.i.d. slučajne varijable bit će normalne ili gotovo normalne ako je veličina uzorka dovoljno velika

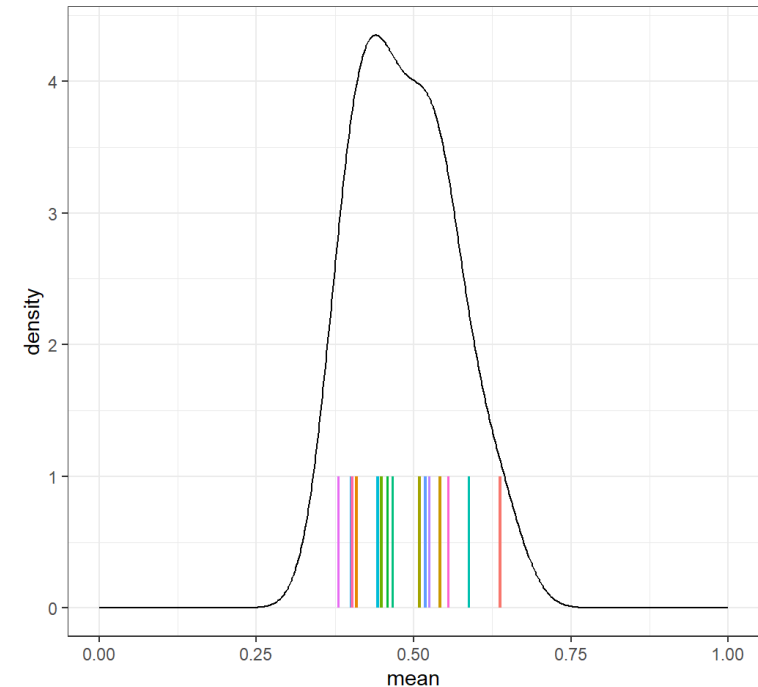


Centralni granični teorem

- Distribucija procjena statistika zbroja ili prosjeka i.i.d. slučajne varijable bit će normalne ili gotovo normalne ako je veličina uzorka dovoljno velika



Centralni granični teorem

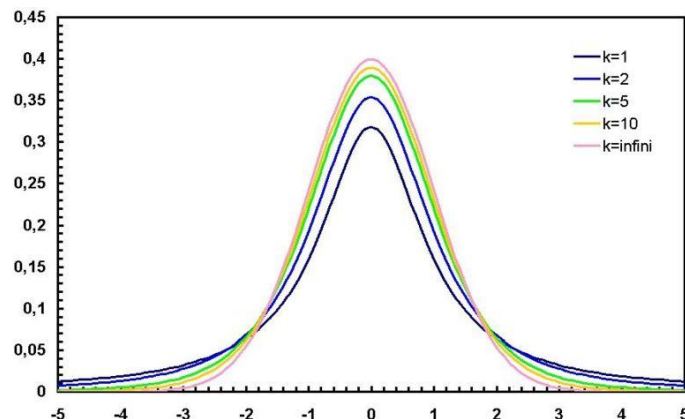


- Statistiku možemo izraziti kao t-vrijednost ili kao z-vrijednost (koristite t-vrijednost kada je veličina uzorka mala ili je standardna devijacija populacije nepoznata)
 - Standardna devijacija statistike
 - Standardna greška statistike

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$
$$SE_{\bar{x}} = s / \sqrt{n}$$

Studentova t-distribucija

- Obitelj sličnih distribucija vjerojatnosti.
- Specifična t-distribucija ovisi o stupnjevima slobode, ν .
- Stupnjevi slobode – broj neovisnih opažanja u uzorku podataka koji su dostupni za procjenu parametra populacije iz koje je taj uzorak izvučen
- Srednja vrijednost t-distribucije je nula.
- Kako ν raste, t-distribucija postaje normalnija.



$$f_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

$$-\infty \leq t \leq +\infty$$

T-test na jednom uzorku

- ▶ Pretvorimo statistiku u t-vrijednost:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad \Pr \left[-t_{df, \alpha/2} \leq \frac{\bar{x} - \mu}{s / \sqrt{n}} \leq t_{df, \alpha/2} \right] = 1 - \alpha$$

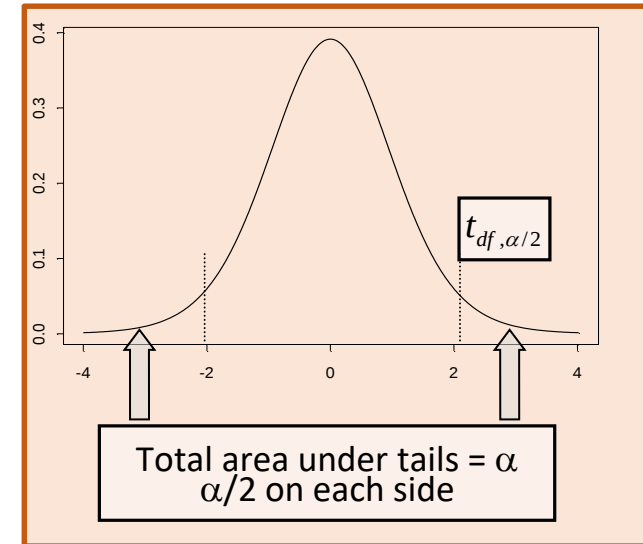
$$\Pr \left[-t_{df, \alpha/2}(s / \sqrt{n}) \leq \bar{x} - \mu \leq t_{df, \alpha/2}(s / \sqrt{n}) \right] = 1 - \alpha$$

$$\Pr \left[t_{df, \alpha/2}(s / \sqrt{n}) \geq -\bar{x} + \mu \geq -t_{df, \alpha/2}(s / \sqrt{n}) \right] = 1 - \alpha$$

$$\Pr \left[\bar{x} - t_{df, \alpha/2}(s / \sqrt{n}) \leq \mu \leq \bar{x} + t_{df, \alpha/2}(s / \sqrt{n}) \right] = 1 - \alpha$$

$$(1 - \alpha)\% \text{ C.I.: } \bar{x} \pm t_{df, \alpha/2} SE$$

$$SE_{\bar{x}} = s / \sqrt{n}$$



Kako konstruirati interval pouzdanosti

- Identificirajte statistiku uzorka.
- Odaberite razinu pouzdanosti.
- Pronađite marginu greške.
 - Margina greške = kritična vrijednost * standardna devijacija statistike
 - Margina greške = kritična vrijednost * standardna greška statistike
- Odredite interval pouzdanosti. Nesigurnost je označena razinom pouzdanosti.

Interval pouzdanosti = statistika uzorka $\pm$ margina greške

Margina greške

= kritična vrijednost * standardna devijacija (ili standardna greška) statistike

Primjer

- Želimo procijeniti prosječnu težinu odraslih muškaraca u Zagrebu. Odaberemo nasumičan uzorak 1,000 muškaraca iz populacije od 1,000,000 muškaraca i izvažemo ih. Izmjerili smo da je prosječna težina našeg uzorka 92 kg, a standardna devijacija uzorka je 14kg. Izračunajte 95% interval pouzdanosti.

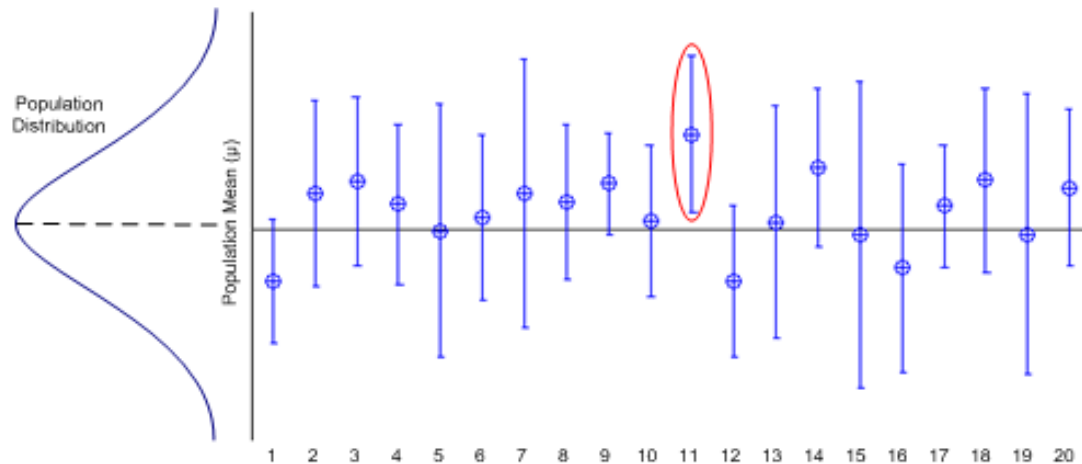
$$\alpha = 0.05; \bar{x} = 92; s = 14; n = 1000; df=999$$

df – degrees of freedom (stupnjevi slobode)

$$95\% \text{ interval pouzdanosti: } 92 \pm t_{999,0.025}(14/\sqrt{1000}) = 92 \pm 0.87$$

Interpretacija intervala pouzdanosti

- Kad bi se uzeli ponovljeni uzorci i izračunao interval pouzdanosti od 95% za svaki uzorak, 95% intervala sadržavalo bi srednju vrijednost populacije.



- Standardna devijacija – koliko su raspršena mjerenja
- Standardna greška – preciznost procjene određene mjere
- Interval pouzdanosti – preciznost metode uzorkovanja