

Statistika

Vanja Wagner

0. Opis kolegija

Kontakti

Vanja Wagner (predavanja četvrtkom 16-18h 003) — vanja.wagner@math.hr,
<https://web.math.pmf.unizg.hr/~wagner>

Petra Lazić (vježbe srijedom 17-19h Pr2) — petra.lazic@math.hr

Krunoslav Ivanović (vježbe četvrtkom 18-20h Pr1) — krunoslav.ivanovic@student.math.hr

Kontakti

Vanja Wagner (predavanja četvrtkom 16-18h 003) — vanja.wagner@math.hr,
<https://web.math.pmf.unizg.hr/~wagner>

Petra Lazić (vježbe srijedom 17-19h Pr2) — petra.lazic@math.hr

Krunoslav Ivanović (vježbe četvrtkom 18-20h Pr1) — krunoslav.ivanovic@student.math.hr

Polaganje

- Pišu se dva kolokvija (prvi u terminu izvanrednih jesenskih rokova, drugi na prvom zimskom roku) — jedan od kolokvija može se ispravljati na drugom zimskom roku.
- Alternativno, kolegij se može položiti preko pismenog ispita (koji obuhvaća cijelo gradivo). Uvjet za izlazak na pismeni ispit je minimalno 20 bodova (od 100) na oba kolokvija.

Kontakti

Vanja Wagner (predavanja četvrtkom 16-18h 003) — vanja.wagner@math.hr,
<https://web.math.pmf.unizg.hr/~wagner>

Petra Lazić (vježbe srijedom 17-19h Pr2) — petra.lazic@math.hr

Krunoslav Ivanović (vježbe četvrtkom 18-20h Pr1) — krunoslav.ivanovic@student.math.hr

Polaganje

- Pišu se dva kolokvija (prvi u terminu izvanrednih jesenskih rokova, drugi na prvom zimskom roku) — jedan od kolokvija može se ispravljati na drugom zimskom roku.
- Alternativno, kolegij se može položiti preko pismenog ispita (koji obuhvaća cijelo gradivo). Uvjet za izlazak na pismeni ispit je minimalno 20 bodova (od 100) na oba kolokvija.

Glavna literatura:



Michael C. Whitlock, Dolph Schluter, The Analysis of Biological Data, Macmillan Learning (2020)

B. Basrak: Statistika s primjenama u R-u - skripta.

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika
- 3 Osnove vjerojatnosti

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika
- 3 Osnove vjerojatnosti
- 4 Procjene raznih parametara

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika
- 3 Osnove vjerojatnosti
- 4 Procjene raznih parametara
- 5 Testiranje statističkih hipoteza

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika
- 3 Osnove vjerojatnosti
- 4 Procjene raznih parametara
- 5 Testiranje statističkih hipoteza
- 6 Korelacija i jednostavna linearna regresija

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *statistical inference*).

Sadržaj kolegija

- 1 Osnovni pojmovi i prikupljanje podataka
- 2 Deskriptivna statistika
- 3 Osnove vjerojatnosti
- 4 Procjene raznih parametara
- 5 Testiranje statističkih hipoteza
- 6 Korelacija i jednostavna linearna regresija
- 7 Neparametarski testovi

1. Osnovni pojmovi i prikupljanje podataka

Osnovni pojmovi.

Osnovni pojmovi.

- **Varijabla** je karakteristika (svojstvo, atribut, obilježje) koja može poprimiti različite vrijednosti. Primjeri varijable su: visina (osobe), težina (osobe), indeks tjelesne težine (BMI), vrijeme, broj sadnica u rasadniku, ocjena na ispitu, vrsta životinje.

Osnovni pojmovi.

- **Varijabla** je karakteristika (svojstvo, atribut, obilježje) koja može poprimiti različite vrijednosti. Primjeri varijable su: visina (osobe), težina (osobe), indeks tjelesne težine (BMI), vrijeme, broj sadnica u rasadniku, ocjena na ispitu, vrsta životinje.
- **Podatak** je izmjerena vrijednost varijable. U statistici obično krećemo od skupa podataka (*engl.* data set). Npr. za danu varijablu *visina osobe* izmjerimo visine (u cm) 10 slučajno odabranih osoba i dobijemo skup podataka:

175, 166, 181, 159, 188, 176, 171, 183, 179, 162.

Osnovni pojmovi.

- **Varijabla** je karakteristika (svojstvo, atribut, obilježje) koja može poprimiti različite vrijednosti. Primjeri varijable su: visina (osobe), težina (osobe), indeks tjelesne težine (BMI), vrijeme, broj sadnica u rasadniku, ocjena na ispitu, vrsta životinje.
- **Podatak** je izmjerena vrijednost varijable. U statistici obično krećemo od skupa podataka (*engl.* data set). Npr. za danu varijablu *visina osobe* izmjerimo visine (u cm) 10 slučajno odabranih osoba i dobijemo skup podataka:

175, 166, 181, 159, 188, 176, 171, 183, 179, 162.

- **Populacija** je skup jedinki koje imaju neko zajedničko svojstvo a koje su predmet proučavanja. Definicija populacije ovisi o istraživanju. Tako na primjer za varijablu *visina osobe* promatrana populacija može biti
 - svi ljudi na svijetu;
 - svi studenti Biološkog odsjeka PMF-a.

Osnovni pojmovi.

- **Varijabla** je karakteristika (svojstvo, atribut, obilježje) koja može poprimiti različite vrijednosti. Primjeri varijable su: visina (osobe), težina (osobe), indeks tjelesne težine (BMI), vrijeme, broj sadnica u rasadniku, ocjena na ispitu, vrsta životinje.
- **Podatak** je izmjerena vrijednost varijable. U statistici obično krećemo od skupa podataka (*engl.* data set). Npr. za danu varijablu *visina osobe* izmjerimo visine (u cm) 10 slučajno odabranih osoba i dobijemo skup podataka:

175, 166, 181, 159, 188, 176, 171, 183, 179, 162.

- **Populacija** je skup jedinki koje imaju neko zajedničko svojstvo a koje su predmet proučavanja. Definicija populacije ovisi o istraživanju. Tako na primjer za varijablu *visina osobe* promatrana populacija može biti
 - svi ljudi na svijetu;
 - svi studenti Biološkog odsjeka PMF-a.
- **Uzorak** je (bilo koji) podskup populacije (grupa jedinki) na kojoj vršimo opažanja. Da bi bio statistički valjan mora dobro *reprezentirati* populaciju.

Klasifikacija varijabli i podataka

Osnovna klasifikacija varijabli i podataka se svodi na dva tipa: **kvalitativne** i **kvantitativne**.

Klasifikacija varijabli i podataka

Osnovna klasifikacija varijabli i podataka se svodi na dva tipa: **kvalitativne** i **kvantitativne**.

- **Kvalitativne ili kategorijske varijable** — vrijednosti varijable mogu se jednoznačno podijeliti u kategorije. Na primjer na populaciji *svi studenti na PMF-BO* varijablu *županija prebivališta* možemo podijeliti u točno 22 kategorije (21 županija RH + inozemno prebivalište).

Klasifikacija varijabli i podataka

Osnovna klasifikacija varijabli i podataka se svodi na dva tipa: **kvalitativne** i **kvantitativne**.

- **Kvalitativne ili kategorijske varijable** — vrijednosti varijable mogu se jednoznačno podijeliti u kategorije. Na primjer na populaciji *svi studenti na PMF-BO* varijablu *županija prebivališta* možemo podijeliti u točno 22 kategorije (21 županija RH + inozemno prebivalište).
- **Kvantitativne ili numeričke varijable** — vrijednost varijable je rezultat mjerenja na numeričkoj skali. Primjeri numeričkih varijabli su:
 - broj stabala na parceli
 - kapacitet rasadnika
 - visina stabla
 - duljina potoka.

Klasifikacija varijabli i podataka

Osnovna klasifikacija varijabli i podataka se svodi na dva tipa: **kvalitativne i kvantitativne**.

- **Kvalitativne ili kategorijske varijable** — vrijednosti varijable mogu se jednoznačno podijeliti u kategorije. Na primjer na populaciji *svi studenti na PMF-BO* varijablu *županija prebivališta* možemo podijeliti u točno 22 kategorije (21 županija RH + inozemno prebivalište).
- **Kvantitativne ili numeričke varijable** — vrijednost varijable je rezultat mjerenja na numeričkoj skali. Primjeri numeričkih varijabli su:
 - broj stabala na parceli
 - kapacitet rasadnika
 - visina stabla
 - duljina potoka.

Važno je primijetiti da nije svaka varijabla koja poprima brojčane vrijednosti kvantitativna. Na primjer, za vrijednosti kategorijske varijable *stupanj zadovoljstva uslugom* u nekom istraživanju možemo uzeti: jako nezadovoljan, nezadovoljan, zadovoljan i jako zadovoljan. No uočimo da smo isto tako *stupanj zadovoljstva uslugom* mogli ocijeniti i brojčanim vrijednostima od 1 do 4, pri čemu se tip varijable ne bi promijenio (razina zadovoljstva je i dalje podijeljena u 4 kategorije).

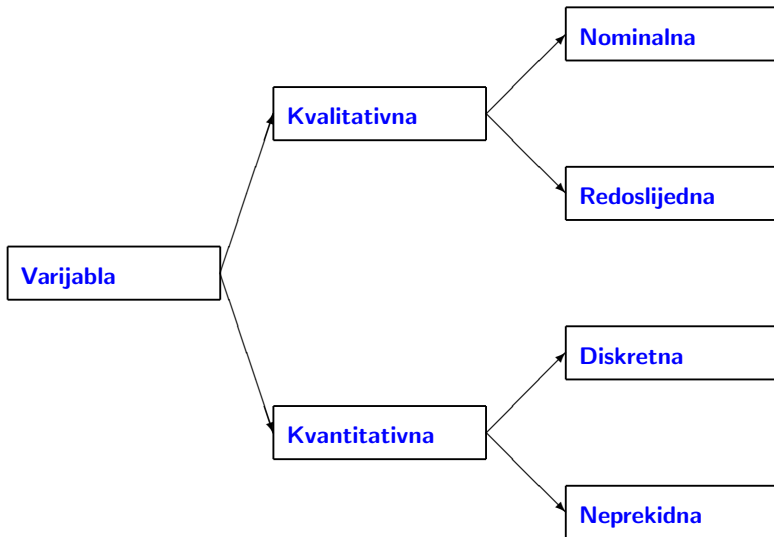
Daljnje podjele varijabli

Kategorijske varijable možemo dodatno podijeliti u dvije skupine - ovisno o tome postoji li među kategorijama jasan uređaj:

- **nominalne varijable** - ne postoji uređaj između kategorija (primjer je varijabla *županija prebivališta*)
- **ordinalne (redosljedne, uređajne) varijable** - postoji uređaj između kategorija (primjer je varijabla *razina zadovoljstva uslugom*)

Kvantitativne varijable dijelimo u dvije dodatne kategorije, ovisno o tipu numeričkih vrijednosti koje pooptimaju:

- **diskretne varijable** - skup vrijednosti je konačan ili prebrojiv (npr. varijabla *broj šestica u bacanja 3 igraće kocke* ili *broj bacanja kocke do pojave prve šestice*)
- **neprekidne varijable** - poprimaju sve realne vrijednosti unutar nekog intervala (npr. *visina stabla*, *duljina potoka*, *vrijeme*)



Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji) — kvantitativna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji) — kvantitativna — diskretna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji) — kvantitativna — diskretna
- plasman na utrci (populacija: sudionici Zagrebačkog maratona)

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji) — kvantitativna — diskretna
- plasman na utrci (populacija: sudionici Zagrebačkog maratona) — kvalitativna

Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI) — kategorijska — nominalna
- dob (populacija: studenti na PMF-BO) — kvantitativna — diskretna
- spol (populacija: mačke registrirane u vet. ordinaciji) — kategorijska — nominalna
- duljina repa (populacija: domaće mačke u RH) — kvantitativna — neprekidna
- ocjena (populacija: studenti na Statistici 24./p1/slike/25.) — kategorijska — ordinalna
- broj ptica u gnijezdu (populacija: rode u Baranji) — kvantitativna — diskretna
- plasman na utrci (populacija: sudionici Zagrebačkog maratona) — kvalitativna — ordinalna

Česte greške pri dizajnu pokusa

Iako se korisnici statističkih metoda često pribojavaju grešaka u statističkoj analizi, mnoge greške zapravo nastaju već u procesu dizajniranja pokusa, odnosno prikupljanja podataka. Iako u sklopu ovog kolegija nećemo govoriti o ovom problemu vrlo detaljno, osvrnut ćemo se ukratko na česte tipove grešaka.

Česte greške pri dizajnu pokusa

Iako se korisnici statističkih metoda često pribojavaju grešaka u statističkoj analizi, mnoge greške zapravo nastaju već u procesu dizajniranja pokusa, odnosno prikupljanja podataka. Iako u sklopu ovog kolegija nećemo govoriti o ovom problemu vrlo detaljno, osvrnut ćemo se ukratko na česte tipove grešaka.

Promotriti ćemo tri primjera čestih grešaka u dizajnu:

- Nereprezentativni uzorak;
- Kontrolna skupina nije odgovarajuća;
- Nejasno postavljene hipoteze.

Reprezentativnost uzorka

Primjer. Procjena prosjeka ocjena na studiju

Promotrimo na primjer populaciju *svi studenti Biološkog odsjeka PMF-a*. Za analizu nekog svojstva te populacije (npr. prosjek ocjena svih položenih kolegija) kao uzorak možemo uzeti npr.

- sve studente 2. godine PMF-BO;
- slučajno odabranih 100 studenata s PMF-BO.

Koji je od ova dva uzorka reprezentabilniji za promatranu populaciju?

Reprezentativnost uzorka

Primjer. Procjena prosjeka ocjena na studiju

Promotrimo na primjer populaciju *svi studenti Biološkog odsjeka PMF-a*. Za analizu nekog svojstva te populacije (npr. prosjek ocjena svih položenih kolegija) kao uzorak možemo uzeti npr.

- sve studente 2. godine PMF-BO;
- slučajno odabranih 100 studenata s PMF-BO.

Koji je od ova dva uzorka reprezentabilniji za promatranu populaciju?

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike

Populacija koja nas interesira mogu biti svi studenti prve godine Sveučilišta u Zagrebu. Planiramo uzorak veličine $n = 200$ studenata. Možemo prikupiti više podataka o studentima, tako što ćemo uključiti npr. varijable spol, dob, regiju iz koje dolaze, uspjeh na maturalskom ispitu iz matematike, fakultet na kojem studiraju, itd. Ključne varijable koju promatramo je postignuti broj bodova na ispitu i vrstu obroka koji im je poslužen prije ispita. Kako odabrati uzorak?

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.).

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu?

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu? → *randomized control trials*.

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu? → *randomized control trials*.

Ima li smisla da studenti znaju u kojoj su grupi?

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu? → *randomized control trials*.

Ima li smisla da studenti znaju u kojoj su grupi? → *placebo tretman*.

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu? → *randomized control trials*.

Ima li smisla da studenti znaju u kojoj su grupi? → *placebo tretman*.

Ima li smisla da ispravljajući znaju u kojoj je student grupi?

Odabir kontrolne skupina

Recimo da našim istraživanjem želimo procijeniti efekte nekog tretmana. Jedna mogućnost je da usporedimo promatranu varijablu kod jedinki u uzorku prije i poslije tretmana. No često takav pristup može dovesti do pogrešnih zaključaka, obzirom da ulogu u promjeni vrijednosti varijable mogu imati brojni drugi čimbenici (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U takvim situacijama je dobra praksa odabrati **kontrolnu grupu** (u medicini tipično placebo grupu) koja je po svemu usporediva s grupom pod tretmanom. Takav pristup je značajno unaprijedio posebno medicinska istraživanja i danas čine određeni istraživački standard. Iz istraživanja u kojem nemaju kontrolne skupine ili je ona neodgovarajuća, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 100 studenata koji dođu na testiranje staviti u testnu, a preostalih 100 u kontrolnu skupinu? → *randomized control trials*.

Ima li smisla da studenti znaju u kojoj su grupi? → *placebo tretman*.

Ima li smisla da ispravljajući znaju u kojoj je student grupi? → *double blind*.

Nejasno postavljene hipoteze i postavljanje hipoteza *a posteriori*

Istaživači bi mogli obaviti pokus i mjerenja korektno te prikupiti veliki broj podataka, ali postaviti svoje hipoteze tek nakon inicijalne statističke analize. Tako bi npr. mogli otkriti da je obrok bogat šećerom imao pozitivan efekt na rezultate testa, no samo za studentice i to one koje dolaze iz kontinentalnih dijelova zemlje.

Nejasno postavljene hipoteze i postavljanje hipoteza *a posteriori*

Istaživači bi mogli obaviti pokus i mjerenja korektno te prikupiti veliki broj podataka, ali postaviti svoje hipoteze tek nakon inicijalne statističke analize. Tako bi npr. mogli otkriti da je obrok bogat šećerom imao pozitivan efekt na rezultate testa, no samo za studentice i to one koje dolaze iz kontinentalnih dijelova zemlje.

Ovakav pristup istraživanju može biti zanimljiv, no ponovljeni pokus će rezultate rijetko replicirati. Ovakav način analize je karakterističan u području *data mininga*.

Nejasno postavljene hipoteze i postavljanje hipoteza *a posteriori*

Istaživači bi mogli obaviti pokus i mjerenja korektno te prikupiti veliki broj podataka, ali postaviti svoje hipoteze tek nakon inicijalne statističke analize. Tako bi npr. mogli otkriti da je obrok bogat šećerom imao pozitivan efekt na rezultate testa, no samo za studentice i to one koje dolaze iz kontinentalnih dijelova zemlje.

Ovakav pristup istraživanju može biti zanimljiv, no ponovljeni pokus će rezultate rijetko replicirati. Ovakav način analize je karakterističan u području *data mininga*. Može poslužiti za inicijalnu analizu i postavljanje hipoteza koje bi trebalo potvrditi na novim nezavisno prikupljenim podacima.

1. Deskriptivna statistika

Deskriptivna statistika se odnosi na prikupljanje, organizaciju, sažimanje i prikaz podataka. U kasnijim poglavljima bavit ćemo se i **inferencijalnom statistikom (statističkim zaključivanjem)** koja se sastoji od generalizacije zaključivanja s uzorka na populaciju, testiranja hipoteza, određivanja veza između varijabli i predviđanja.

Deskriptivna statistika se odnosi na prikupljanje, organizaciju, sažimanje i prikaz podataka. U kasnijim poglavljima bavit ćemo se i **inferencijalnom statistikom (statističkim zaključivanjem)** koja se sastoji od generalizacije zaključivanja s uzorka na populaciju, testiranja hipoteza, određivanja veza između varijabli i predviđanja.

Primjer

Od 97 studenata upisanih na kolegij Statistika, njih 82 je došlo na prvo predavanje. Dakle, (skoro) 85% uščenata je došlo na prvo predavanje iz Statistike. Uočimo:

- varijabla: pohađanje prvog predavanja iz Statistike;
- populacija: svi studenti upisani na kolegij Statistika u ak.g. 24./p1/slike/25.;
- za svakog je studenta poznato je li došao na prvo predavanje ili ne.

Deskriptivna statistika se odnosi na prikupljanje, organizaciju, sažimanje i prikaz podataka. U kasnijim poglavljima bavit ćemo se i **inferencijalnom statistikom (statističkim zaključivanjem)** koja se sastoji od generalizacije zaključivanja s uzorka na populaciju, testiranja hipoteza, određivanja veza između varijabli i predviđanja.

Primjer

Od 97 studenata upisanih na kolegij Statistika, njih 82 je došlo na prvo predavanje. Dakle, (skoro) 85% uščenata je došlo na prvo predavanje iz Statistike. Uočimo:

- varijabla: pohađanje prvog predavanja iz Statistike;
- populacija: svi studenti upisani na kolegij Statistika u ak.g. 24./p1/slike/25.;
- za svakog je studenta poznato je li došao na prvo predavanje ili ne.

No, da smo imali samo evidenciju dolaska za studente prijediplomskog studija Biologija (ti podaci bi nam činili uzorak koji je strogo manji od cijele populacije), morali bi naći mehanizam kako ocijeniti posjećenost prvog predavanja za cijelu populaciju temeljem podataka na manjem uzorku — inferencijalna statistika. Razmislite koji su potencijalni problemi u ovoj proceduri?

Prikaz podataka

Primjer:

Na zahtjev vlasnika farme, radnik je bilježio spol 397 zečeva. Sljedeći dan je rezultat prosljedio svome nadređenom:

M M Ž Ž M Ž M M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M M M Ž M Ž M Ž M Ž M Ž Ž
M Ž M Ž M Ž Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž Ž Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M
Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
M Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž

Ovakvi sirovi podaci dobiveni kao originalni rezultati mjerenja često su jako neprekledni. Čak i ako mjerimo ordinalnu ili kvantitativnu varijablu i mjerenja reprezentiramo uređenim nizom podataka (podaci posloženi po veličini, odnosno redu) - i dalje možemo imati nepreglednu situaciju.

Uočimo da nam je u prethodnom primjeru spolova zečeva za uvid u podatke dovoljno prebrojati zečeve svakog spola - u nizu podataka se M pojavljuje 172 puta, a Ž 225 puta. Stoga nam je za prikaz ovih podataka dovoljna informacija:

$$f_1 = 172, \quad f_2 = 225.$$

Uočimo da nam je u prethodnom primjeru spolova zečeva za uvid u podatke dovoljno prebrojati zečeve svakog spola - u nizu podataka se M pojavljuje 172 puta, a Ž 225 puta. Stoga nam je za prikaz ovih podataka dovoljna informacija:

$$f_1 = 172, \quad f_2 = 225.$$

→ **Frekvencija** (apsolutna frekvencija) je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Uočimo da nam je u prethodnom primjeru spolova zečeva za uvid u podatke dovoljno prebrojati zečeve svakog spola - u nizu podataka se M pojavljuje 172 puta, a Ž 225 puta. Stoga nam je za prikaz ovih podataka dovoljna informacija:

$$f_1 = 172, \quad f_2 = 225.$$

→ **Frekvencija** (apsolutna frekvencija) je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Alternativno, podatke smo mogli reprezentirati i udjelima M, odnosno Ž podataka u nizu podataka - 0.4332 (43.32%) podataka je M, a 0.5668 (56.68%) podataka je Ž:

$$r_1 = 0.4332, \quad r_2 = 0.5668.$$

Uočimo da nam je u prethodnom primjeru spolova zečeva za uvid u podatke dovoljno prebrojati zečeve svakog spola - u nizu podataka se M pojavljuje 172 puta, a Ž 225 puta. Stoga nam je za prikaz ovih podataka dovoljna informacija:

$$f_1 = 172, \quad f_2 = 225.$$

→ **Frekvencija** (apsolutna frekvencija) je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Alternativno, podatke smo mogli reprezentirati i udjelima M, odnosno Ž podataka u nizu podataka - 0.4332 (43.32%) podataka je M, a 0.5668 (56.68%) podataka je Ž:

$$r_1 = 0.4332, \quad r_2 = 0.5668.$$

→ **Relativna frekvencija** je omjer broja pojavljivanja određene vrijednosti varijable (u skupu podataka) i ukupne veličine podataka (**opseg podataka**).

Uočimo da nam je u prethodnom primjeru spolova zečeva za uvid u podatke dovoljno prebrojati zečeve svakog spola - u nizu podataka se M pojavljuje 172 puta, a Ž 225 puta. Stoga nam je za prikaz ovih podataka dovoljna informacija:

$$f_1 = 172, \quad f_2 = 225.$$

→ **Frekvencija** (apsolutna frekvencija) je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Alternativno, podatke smo mogli reprezentirati i udjelima M, odnosno Ž podataka u nizu podataka - 0.4332 (43.32%) podataka je M, a 0.5668 (56.68%) podataka je Ž:

$$r_1 = 0.4332, \quad r_2 = 0.5668.$$

→ **Relativna frekvencija** je omjer broja pojavljivanja određene vrijednosti varijable (u skupu podataka) i ukupne veličine podataka (**opseg podataka**).

Ove podatke možemo jednostavno prikazati i tablično

Spol	frekvencija	relativna frekvencija
M	172	0.4332
Ž	225	0.5668
Σ	397	1

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

- n - veličina skupa podataka (opseg podataka)

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

- n - veličina skupa podataka (opseg podataka)
- k različitih vrijednosti u podacima označimo s a_1, \dots, a_k

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

- n - veličina skupa podataka (opseg podataka)
- k različitih vrijednosti u podacima označimo s a_1, \dots, a_k
- f_i - frekvencija i -te vrijednosti (a_i) u skupu podataka (uočimo $n = f_1 + \dots + f_n$)

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

- n - veličina skupa podataka (opseg podataka)
- k različitih vrijednosti u podacima označimo s a_1, \dots, a_k
- f_i - frekvencija i -te vrijednosti (a_i) u skupu podataka (uočimo $n = f_1 + \dots + f_n$)
- r_i - relativna frekvencija i -te vrijednosti (a_i) u skupu podataka (uočimo $r_i = \frac{f_i}{n}$)

Frekvencijska tablica

Imamo n mjerenja obilježja koje poprima k različitih vrijednosti (u prethodnom primjeru $n = 397$, $k = 2$, a vrijednosti su $a_1 = M$, $a_2 = \check{Z}$).

- n - veličina skupa podataka (opseg podataka)
- k različitih vrijednosti u podacima označimo s a_1, \dots, a_k
- f_i - frekvencija i -te vrijednosti (a_i) u skupu podataka (uočimo $n = f_1 + \dots + f_n$)
- r_i - relativna frekvencija i -te vrijednosti (a_i) u skupu podataka (uočimo $r_i = \frac{f_i}{n}$)
- ako postoji uređaj na podacima ($a_1 < a_2 < \dots < a_k$) ima smisla promatrati i **relativne kumulativne frekvencije** - F_i je relativna kumulativna frekvencija podataka koji su manji ili jednaki od i -te vrijednosti (a_i), odnosno broj jedinki s vrijednošću obilježja manjim ili jednakim od i -te vrijednosti obilježja (uočimo $F_i = r_1 + \dots + r_i$).

Frenkvencijska tablica - opći oblik:

Obilježje	frekvencija	relativna frekvencija	relativne kumulativne frekvencije
a_1	f_1	r_1	F_1
a_2	f_2	r_2	F_2
.	.	.	.
.	.	.	.
.	.	.	.
a_k	f_k	r_k	$F_k (= 1)$
Σ	n	1	

Primjer.

Promotrimo frekvencijsku tablicu kojom prikazujemo dob 20 ispitanika:

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00
Σ	20	1	-	-

Primjer. Promotrimo bodove na kratkom testu u jednoj generaciji od 200 studenata.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

1 Koliko je studenata dobilo 5 ili 6 bodova?

Primjer. Promotrimo bodove na kratkom testu u jednoj generaciji od 200 studenata.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

1 Koliko je studenata dobilo 5 ili 6 bodova? $15 + 33 = 142 - 94 = 48$

Primjer. Promotrimo bodove na kratkom testu u jednoj generaciji od 200 studenata.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

- 1 Koliko je studenata dobilo 5 ili 6 bodova? $15 + 33 = 142 - 94 = 48$
- 2 Za izlazak na kolokvij potrebno je ostvariti minimalno 5 bodova na kratkom testu. Koliki je postotak studenata koji su ostvarili izlazak na kolokvij?

Primjer. Promotrimo bodove na kratkom testu u jednoj generaciji od 200 studenata.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

- 1 Koliko je studenata dobilo 5 ili 6 bodova? $15 + 33 = 142 - 94 = 48$
- 2 Za izlazak na kolokvij potrebno je ostvariti minimalno 5 bodova na kratkom testu. Koliki je postotak studenata koji su ostvarili izlazak na kolokvij? $100\% - 47\% = 53\%$

Grafički prikaz podataka

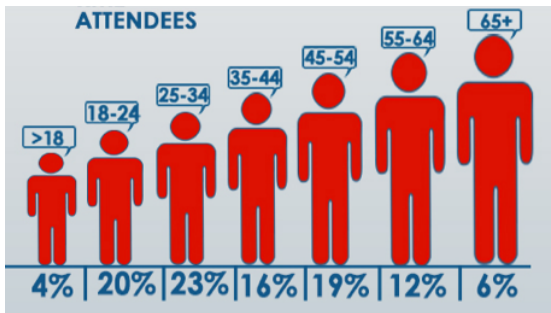
Neki od standardnih grafičkih prikaza podataka vrlo su nam dobro poznati iz svakodnevnog života (novine, portali, vijesti i sl.). Odlikuje ih jednostavnost i preglednost, iako općenito ne možemo očekivati preveliku preciznost takvog prikaza podataka.

Grafički prikaz podataka

Neki od standardnih grafičkih prikaza podataka vrlo su nam dobro poznati iz svakodnevnog života (novine, portali, vijesti i sl.). Odlikuje ih jednostavnost i preglednost, iako općenito ne možemo očekivati preveliku preciznost takvog prikaza podataka. Neki od primjera grafikona su:

- površinski grafikoni
 - grafikon stupaca (stupčasti dijagram, histogram)
 - grafikon krugova i polukrugova (tzv. *pie chart*)
- linijski grafikoni
- kartogrami
- slikovni grafikoni (piktograf,..)

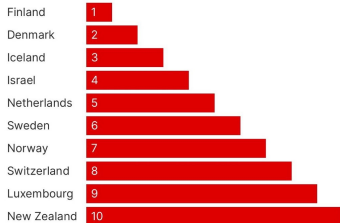
Primjer: Neprecizni i nepregledni grafički prikazi.



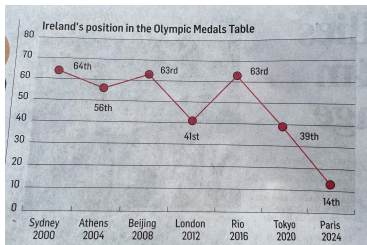
Primjer: Neprecizni i nepregledni grafički prikazi.

Top 10 Happiest Countries in 2023

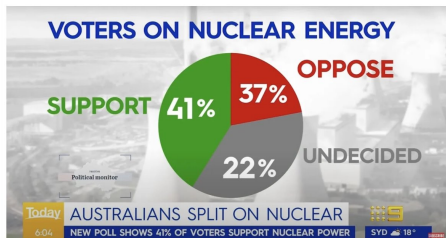
This chart shows the top 10 happiest countries according to the 2023 World Happiness Report.



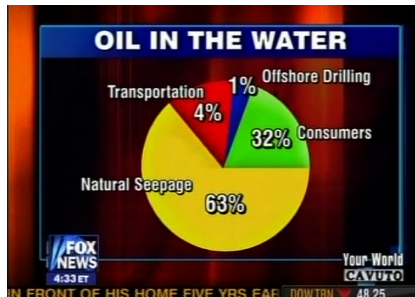
Kilde: [World Happiness Report](#)



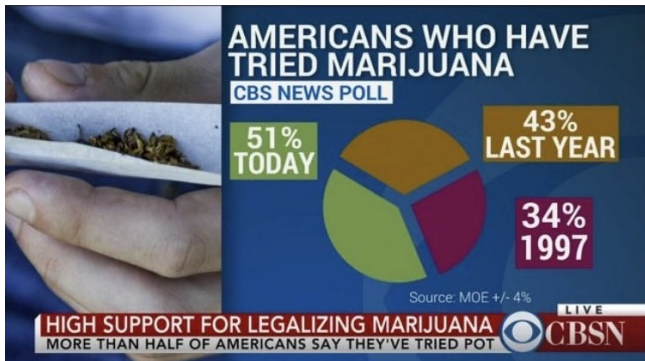
Primjer: Neprecizni i nepregledni grafički prikazi.



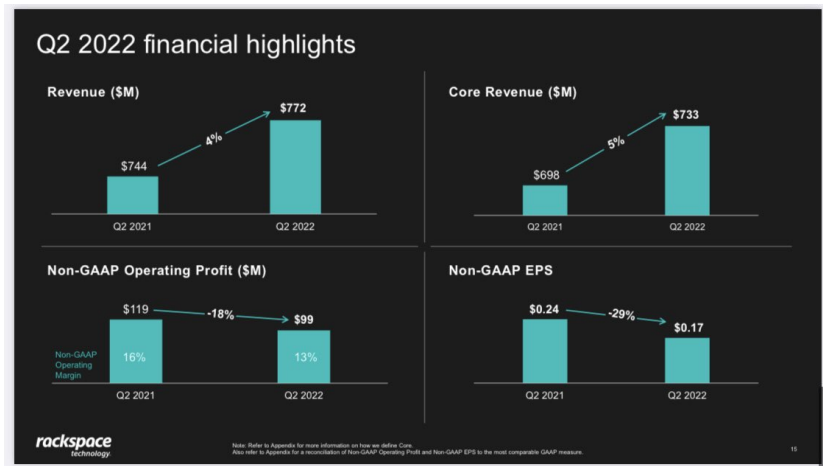
(IMAGE: MELANIE BRACEWELL/TWITTER)



Primjer: Neprecizni i nepregledni grafički prikazi.



Primjer: Neprecizni i nepregledni grafički prikazi.



Stupčasti dijagram

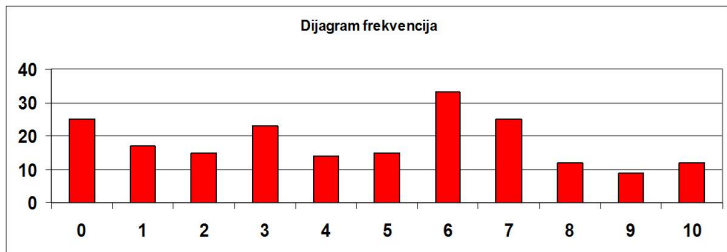
Primjer. Podsjetimo se tablice frekvencija bodova na kratkom testu.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

Stupčasti dijagram

Primjer. Podsjetimo se tablice frekvencija bodova na kratkom testu.

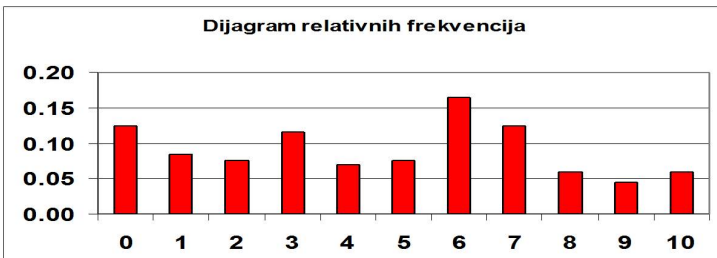
Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	



Stupčasti dijagram

Primjer. Podsjetimo se tablice frekvencija bodova na kratkom testu.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	



Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.

Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.
- **Histogram** je pogodan za prikaz relativnih frekvencija numeričkih kvantitativnih varijabli.

Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.
- **Histogram** je pogodan za prikaz relativnih frekvencija numeričkih kvantitativnih varijabli.
- Kod stupčastog dijagrama visina svakog stupca jednaka je (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.

Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.
- **Histogram** je pogodan za prikaz relativnih frekvencija numeričkih kvantitativnih varijabli.
- Kod stupčastog dijagrama visina svakog stupca jednaka je (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.
- Ako uzmemo stupac širine 1, tada je i *površina* (visina \times širina) svakog stupca stupčastom dijagramu jednaka (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.

Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.
- **Histogram** je pogodan za prikaz relativnih frekvencija numeričkih kvantitativnih varijabli.
- Kod stupčastog dijagrama visina svakog stupca jednaka je (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.
- Ako uzmemo stupac širine 1, tada je i *površina* (visina \times širina) svakog stupca stupčastom dijagramu jednaka (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.
- Ako su podaci uređeni kategorijski ili kvantitativni i sadrže veliki broj različitih vrijednosti, ponekad je zbog preglednosti podatke pogodno *grupirati* u skupine, tako da dobijemo manji broj *razreda*.

Grafikon stupaca

- **Stupčasti dijagram** je pogodan za prikaz (relativnih) frekvencija kategorijskih varijabli ili diskretnih kvantitativnih varijabli.
- **Histogram** je pogodan za prikaz relativnih frekvencija numeričkih kvantitativnih varijabli.
- Kod stupčastog dijagrama visina svakog stupca jednaka je (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.
- Ako uzmemo stupac širine 1, tada je i *površina* (visina x širina) svakog stupca stupčastom dijagramu jednaka (relativnoj) frekvenciji dane vrijednosti u promatranom skupu podataka.
- Ako su podaci uređeni kategorijski ili kvantitativni i sadrže veliki broj različitih vrijednosti, ponekad je zbog preglednosti podatke pogodno *grupirati* u skupine, tako da dobijemo manji broj *razreda*.
- Kod izrade histograma neprekidne numeričke podatke prvo grupiramo u razrede. U histogramu je **površina svakog stupca** jednaka relativnoj frekvenciji pojedinog razreda.

Grupiranje podataka (razredi)

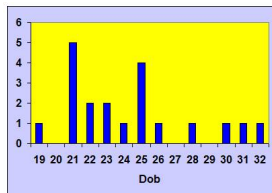
Primjer. Podsjetimo se podataka za (kvantitativnu) varijablu DOB danima tablicom frekvencija:

Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1

Grupiranje podataka (razredi)

Primjer. Podsjetimo se podataka za (kvantitativnu) varijablu DOB danima tablicom frekvencija:

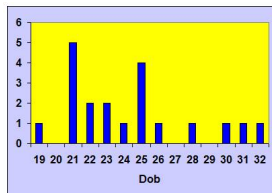
Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1



Grupiranje podataka (razredi)

Primjer. Podsjetimo se podataka za (kvantitativnu) varijablu DOB danima tablicom frekvencija:

Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1



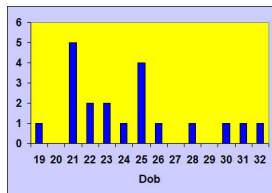
Ukoliko je broj mogućih vrijednosti velik, podaci se grupiraju u razrede. **Razredi** su intervali koji:

- potpuno pokrivaju cijeli skup mogućih vrijednosti obilježja (**načelo iscrpnosti** - svaki podatak se nalazi u nekom razredu)
- se ne preklapaju (**načelo isključivosti** - svaki podatak se nalazi samo u jednom razredu)

Grupiranje podataka (razredi)

Primjer. Podsjetimo se podataka za (kvantitativnu) varijablu DOB danima tablicom frekvencija:

Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1



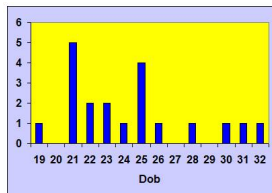
Razredi za varijablu DOB:

Razred	Frekvencija razreda
18 – 20	1
21 – 23	9
24 – 26	6
27 – 29	1
30 – 32	3

Grupiranje podataka (razredi)

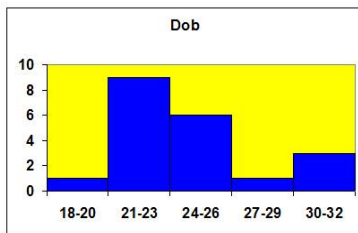
Primjer. Podsjetimo se podataka za (kvantitativnu) varijablu DOB danima tablicom frekvencija:

Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1



Razredi za varijablu DOB:

Razred	Frekvencija razreda
18 – 20	1
21 – 23	9
24 – 26	6
27 – 29	1
30 – 32	3



Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

- Razrede formiramo tako da odredimo **granica razreda** - najmanja i najveća moguća vrijednost koja pripada pojedinom razredu - L_i donja granica i -tog razreda, U_i gornja granica i -tog razreda.

Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

- Razrede formiramo tako da odredimo **granica razreda** - najmanja i najveća moguća vrijednost koja pripada pojedinom razredu - L_i donja granica i -tog razreda, U_i gornja granica i -tog razreda.
- **Širina** i -tog razreda jednaka je $L_{i+1} - L_i$ (često želimo da je širina svih razreda ista)

Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

- Razrede formiramo tako da odredimo **granica razreda** - najmanja i najveća moguća vrijednost koja pripada pojedinom razredu - L_i donja granica i -tog razreda, U_i gornja granica i -tog razreda.
- **Širina** i -tog razreda jednaka je $L_{i+1} - L_i$ (često želimo da je širina svih razreda ista)
- **Sredina** razreda je aritmetička sredina donje i gornje granice razreda, $X_i = \frac{L_i + U_i}{2}$.
- Granice razreda **se ne smiju preklapati** kod diskretnog obilježja.

Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

- Razrede formiramo tako da odredimo **granica razreda** - najmanja i najveća moguća vrijednost koja pripada pojedinom razredu - L_i donja granica i -tog razreda, U_i gornja granica i -tog razreda.
- **Širina** i -tog razreda jednaka je $L_{i+1} - L_i$ (često želimo da je širina svih razreda ista)
- **Sredina** razreda je aritmetička sredina donje i gornje granice razreda, $X_i = \frac{L_i + U_i}{2}$.
- Granice razreda **se ne smiju preklapati** kod diskretnog obilježja.
- Granice razreda **se preklapaju** kod neprekidnog numeričkog obilježja (gornja granica razreda jednaka je donjoj granici sljedećeg razreda $U_i = L_{i+1}$). Pritom se, ukoliko je podatak jednak granici razreda, dogovorno pridružuje jednom razredu (zaokruživanje na dolje ili gore).

Formiranje razreda ovisi o mnogim značajkama podataka i što želimo iz njih izvući.

- Razrede formiramo tako da odredimo **granica razreda** - najmanja i najveća moguća vrijednost koja pripada pojedinom razredu - L_i donja granica i -tog razreda, U_i gornja granica i -tog razreda.
- **Širina** i -tog razreda jednaka je $L_{i+1} - L_i$ (često želimo da je širina svih razreda ista)
- **Sredina** razreda je aritmetička sredina donje i gornje granice razreda, $X_i = \frac{L_i + U_i}{2}$.
- Granice razreda **se ne smiju preklapati** kod diskretnog obilježja.
- Granice razreda **se preklapaju** kod neprekidnog numeričkog obilježja (gornja granica razreda jednaka je donjoj granici sljedećeg razreda $U_i = L_{i+1}$). Pritom se, ukoliko je podatak jednak granici razreda, dogovorno pridružuje jednom razredu (zaokruživanje na dolje ili gore).

Razred	Donja granica	Gornja granica	Širina razreda	Sredina razreda
18 – 20	18	20	3	19
21 – 23	21	23	3	22
24 – 26	24	26	3	25
27 – 29	27	29	3	28
30 – 32	30	32	3	31

Neke posebitosti razreda:

- Ukoliko je prvi ili zadnji razred bez gornje ili donje granice, kažemo da je razred **otvoren**. Češći je primjer kada zadnji razred nema gornje granice.

Neke posebitosti razreda:

- Ukoliko je prvi ili zadnji razred bez gornje ili donje granice, kažemo da je razred **otvoren**. Češći je primjer kada zadnji razred nema gornje granice. Npr. u sljedećem primjeru dani su mjesečni prihodi, gdje u zadnji razred spadaju sve vrijednosti veće od 3.000 € i gornja granica razreda nije definirana:

Razred
0 – 600 €
600 – 1.000 €
1.000 – 2.000 €
2.000 – 3.000 €
više od 3.000 €

Neke posebitosti razreda:

- Ukoliko je prvi ili zadnji razred bez gornje ili donje granice, kažemo da je razred **otvoren**. Češći je primjer kada zadnji razred nema gornje granice. Npr. u sljedećem primjeru dani su mjesečni prihodi, gdje u zadnji razred spadaju sve vrijednosti veće od 3.000 € i gornja granica razreda nije definirana:

Razred
0 – 600 €
600 – 1.000 €
1.000 – 2.000 €
2.000 – 3.000 €
više od 3.000 €

- Ukoliko širine razreda nisu jednake, nekada je potrebno promotriti **korigirane frekvencije** - omjer frekvencije i širine razreda (ili neke druge prikladne veličine proporcionalne širini razreda).

Neke posebitosti razreda:

- Ukoliko je prvi ili zadnji razred bez gornje ili donje granice, kažemo da je razred **otvoren**. Češći je primjer kada zadnji razred nema gornje granice. Npr. u sljedećem primjeru dani su mjesečni prihodi, gdje u zadnji razred spadaju sve vrijednosti veće od 3.000 € i gornja granica razreda nije definirana:

Razred
0 – 600 €
600 – 1.000 €
1.000 – 2.000 €
2.000 – 3.000 €
više od 3.000 €

- Ukoliko širine razreda nisu jednake, nekada je potrebno promotriti **korigirane frekvencije** - omjer frekvencije i širine razreda (ili neke druge prikladne veličine proporcionalne širini razreda).

OPREZ

Za mali skup vrijednosti obilježja (k mali) vrijednosti se ne grupiraju.

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

- 1 Odredimo najmanju i najveću vrijednost (nazovimo ih x_{\min} i x_{\max}), te specificiramo željeni broj razreda k .

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

- 1 Odredimo najmanju i najveću vrijednost (nazovimo ih x_{\min} i x_{\max}), te specificiramo željeni broj razreda k .
- 2 Odredimo razrede I_1, \dots, I_k , odnosno disjunktne intervale oblika $I_j = [L_{j1}, L_j)$ za neke granice razreda L_0, L_1, \dots, L_k . Pri tome vodimo računa da svaki podatak ulazi u točno jedan razred ($L_0 \leq x_{\min} \leq x_{\max} \leq L_k$).

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

- 1 Odredimo najmanju i najveću vrijednost (nazovimo ih x_{\min} i x_{\max}), te specificiramo željeni broj razreda k .
- 2 Odredimo razrede I_1, \dots, I_k , odnosno disjunktne intervale oblika $I_j = [L_{j1}, L_j)$ za neke granice razreda L_0, L_1, \dots, L_k . Pri tome vodimo računa da svaki podatak ulazi u točno jedan razred ($L_0 \leq x_{\min} \leq x_{\max} \leq L_k$).
- 3 Za svaki razred izračunamo njegovu apsolutnu i relativnu frekvenciju (f_j i r_j) jednostavno brojeći podatke koji u njega ulaze.

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

- 1 Odredimo najmanju i najveću vrijednost (nazovimo ih x_{\min} i x_{\max}), te specificiramo željeni broj razreda k .
- 2 Odredimo razrede I_1, \dots, I_k , odnosno disjunktne intervale oblika $I_j = [L_{j1}, L_j)$ za neke granice razreda L_0, L_1, \dots, L_k . Pri tome vodimo računa da svaki podatak ulazi u točno jedan razred ($L_0 \leq x_{\min} \leq x_{\max} \leq L_k$).
- 3 Za svaki razred izračunamo njegovu apsolutnu i relativnu frekvenciju (f_j i r_j) jednostavno brojeći podatke koji u njega ulaze.

razred	granice	f_j	r_j
I_1	L_0, L_1	f_1	r_1
I_2	L_2, L_2	f_2	r_2
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
I_k	L_{k-1}, L_k	f_k	r_k

Izrada histograma

Procedura za izradu histograma neprekidnih numeričkih podataka x_1, x_2, \dots, x_n je sljedeća:

- 1 Odredimo najmanju i najveću vrijednost (nazovimo ih x_{\min} i x_{\max}), te specificiramo željeni broj razreda k .
- 2 Odredimo razrede I_1, \dots, I_k , odnosno disjunktne intervale oblika $I_j = [L_{j1}, L_j)$ za neke granice razreda L_0, L_1, \dots, L_k . Pri tome vodimo računa da svaki podatak ulazi u točno jedan razred ($L_0 \leq x_{\min} \leq x_{\max} \leq L_k$).
- 3 Za svaki razred izračunamo njegovu apsolutnu i relativnu frekvenciju (f_j i r_j) jednostavno brojeći podatke koji u njega ulaze.

razred	granice	f_j	r_j
I_1	L_0, L_1	f_1	r_1
I_2	L_2, L_2	f_2	r_2
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
I_k	L_{k-1}, L_k	f_k	r_k

- 4 Nacrtamo graf na kojem iznad pojedinog razreda povučemo liniju na visini koja odgovara relativnoj frekvenciji tog razreda podijeljenoj s duljinom razreda. Dakle, iznad razreda I_j crtamo stupac visine

$$c_j = \frac{r_j}{L_j - L_{j-1}}.$$

Histogram - napomene

Vrlo često koristimo razrede jednake širine. Tada je dovoljno odrediti L_0 i L_k , a ostale granice izračunamo tako da svi razredi budu širine $(L_k - L_0)/k$. Za takve razrede crtamo nekad i histogram apsolutnih frekvencija jednostavno povlačeći liniju na visini apsolutne frekvencije pojedinog razreda.

Histogram - napomene

Vrlo često koristimo razrede jednake širine. Tada je dovoljno odrediti L_0 i L_k , a ostale granice izračunamo tako da svi razredi budu širine $(L_k - L_0)/k$. Za takve razrede crtamo nekad i histogram apsolutnih frekvencija jednostavno povlačeći liniju na visini apsolutne frekvencije pojedinog razreda.

Konačni izgled dijagrama kod raznih statističkih alata u praksi ovisi o granicama razreda, ali i o tretmanu podataka koji leže na granici dva razreda. Statistički programi obično sami određuju broj i granice razreda (ostavljajući mogućnost korisniku da to izmijeni).

Histogram - napomene

Vrlo često koristimo razrede jednake širine. Tada je dovoljno odrediti L_0 i L_k , a ostale granice izračunamo tako da svi razredi budu širine $(L_k - L_0)/k$. Za takve razrede crtamo nekad i histogram apsolutnih frekvencija jednostavno povlačeći liniju na visini apsolutne frekvencije pojedinog razreda.

Konačni izgled dijagrama kod raznih statističkih alata u praksi ovisi o granicama razreda, ali i o tretmanu podataka koji leže na granici dva razreda. Statistički programi obično sami određuju broj i granice razreda (ostavljajući mogućnost korisniku da to izmijeni).

Postoje brojna pravila za određivanje pogodnog broja razreda - većina se svodi na to da se uzme neka funkcija odgovarajućeg korijena ili logaritma broja podataka. Excel koristi Scottovo pravilo:

$$\text{širina razreda} \approx \frac{3.5\sigma}{\sqrt[3]{n}}$$

(gdje je σ standardna devijacija podataka).

Stupčasti dijagram i histogram - dodaci

Uz stupčaste grafikone mogu se radi jasnijeg prikaza dodatno promotriti i:

- **Poligon frekvencija** - graf koji prolazi točkama (\bar{x}_i, f_i) , odnosno (\bar{x}_i, r_i) (\bar{x}_i je sredina razreda, f_i je frekvencija razreda, r_i je relativna frekvencija razreda)

Stupčasti dijagram i histogram - dodaci

Uz stupčaste grafikone mogu se radi jasnijeg prikaza dodatno promotriti i:

- **Poligon frekvencija** - graf koji prolazi točkama (\bar{x}_i, f_i) , odnosno (\bar{x}_i, r_i) (\bar{x}_i je sredina razreda, f_i je frekvencija razreda, r_i je relativna frekvencija razreda)
- **Poligon kumulativnih frekvencija (ogiva)** - graf koji prolazi točkama (\bar{x}_i, F_i) (\bar{x}_i je sredina razreda, F_i je kumulativna relativna frekvencija razreda)

Primjer.

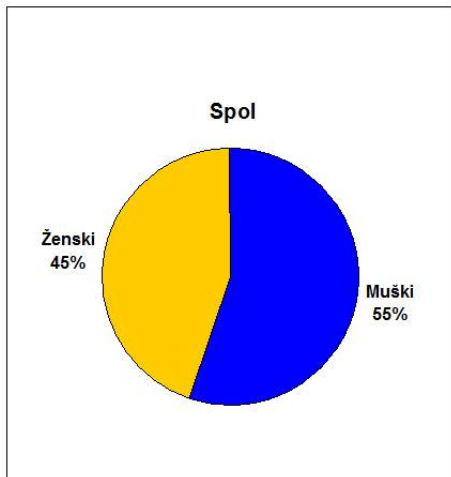
Izmjerena je visina učenika jednog razreda. Dobivene vrijednosti su (izražene u cm):

143	156	156	163	167
142	171	170	169	164
138	158	160	162	164
173	157	158	159	160
138	172	166	166	159
120	125	165	136	168

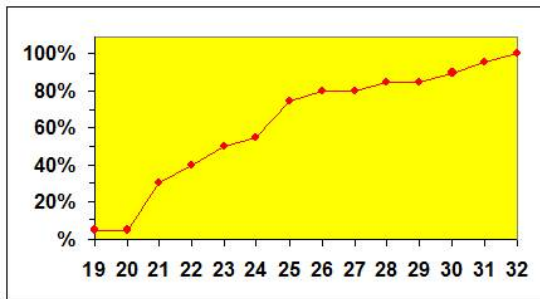
Grupirajte podatke u 6 razreda i odredite distribuciju frekvencija razreda. Distribuciju frekvencija razreda prikažite tablično (prikažite apsolutne i relativne frekvencije te kumulativne i relativne kumulativne frekvencije, odredite sredinu razreda) i grafički koristeći histogram, poligon frekvencija i ogivu.

Ostali grafički prikazi

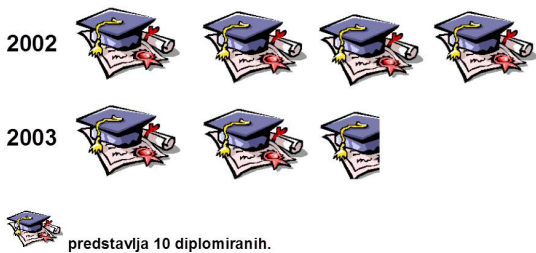
Kružni dijagram (pita, torta). Kružni dijagram se često koristi za prikaz relativnih frekvencija zbog svoje intuitivnosti i lakog razumijevanja, no e kako je teže uočiti razlike u relativnoj frekvenciji kategorija (u odnosu na stupčasti dijagram), rjeđe se upotrebljava u znanstvenom istraživanju.



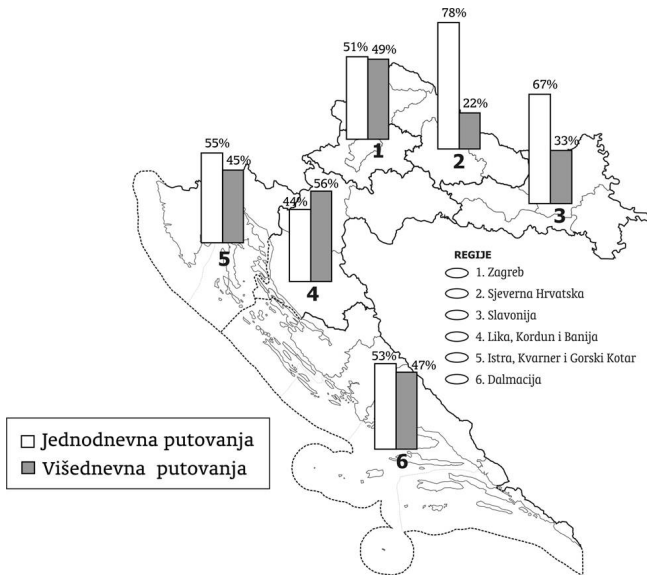
Linijski grafikon kumulativnih relativnih frekvencija



Piktograf.



Kartogram.



Sažeti opis podataka

Cilj nam je opisati naš (često jako) velik skup podataka pomoću nekoliko brojeva koji nam daju preglednu i jasnu informaciju o našim podacima. Pri sažetom opisu podataka koristimo sljedeće mjere:

Sažeti opis podataka

Cilj nam je opisati naš (često jako) velik skup podataka pomoću nekoliko brojeva koji nam daju preglednu i jasnu informaciju o našim podacima. Pri sažetom opisu podataka koristimo sljedeće mjere:

- **Mjere srednje vrijednosti (centralne tendencije)**
- **Mjere raspršenja (varijacije)**
- **Mjere asimetrije**
- **Mjere položaja**

Mjere srednje vrijednosti

- cilj je odrediti broj oko kojeg se grupiraju podaci
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća, najčešći broj mladunaca u leglu

Mjere srednje vrijednosti

- cilj je odrediti broj oko kojeg se grupiraju podaci
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća, najčešći broj mladunaca u leglu

Mjere srednje vrijednosti često dijelimo na dva tipa

- **Potpune** - računaju se na temelju svih podataka

Mjere srednje vrijednosti

- cilj je odrediti broj oko kojeg se grupiraju podaci
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća, najčešći broj mladunaca u leglu

Mjere srednje vrijednosti često dijelimo na dva tipa

- **Potpune** - računaju se na temelju svih podataka
→ aritmetička, geometrijska, harmonijska sredina

Mjere srednje vrijednosti

- cilj je odrediti broj oko kojeg se grupiraju podaci
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća, najčešći broj mladunaca u leglu

Mjere srednje vrijednosti često dijelimo na dva tipa

- **Potpune** - računaju se na temelju svih podataka
→ aritmetička, geometrijska, harmonijska sredina
- **Položajne** - određene rangom (položajem) podataka

Mjere srednje vrijednosti

- cilj je odrediti broj oko kojeg se grupiraju podaci
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća, najčešći broj mladunaca u leglu

Mjere srednje vrijednosti često dijelimo na dva tipa

- **Potpune** - računaju se na temelju svih podataka
→ aritmetička, geometrijska, harmonijska sredina
- **Položajne** - određene rangom (položajem) podataka
→ mod, medijan

Srednja vrijednost

Primjer. Prosječna plaća u tromjesečju: 1500 € , 1200 € , 1800 € .

$$\text{Prosječna plaća} = \frac{1500 + 1200 + 1800}{3} = \frac{4500}{3} = 1500.$$

Srednja vrijednost

Primjer. Prosječna plaća u tromjesečju: 1500 € , 1200 € , 1800 € .

$$\text{Prosječna plaća} = \frac{1500 + 1200 + 1800}{3} = \frac{4500}{3} = 1500.$$

Primjer. Prosječna ocjena: 3, 3, 4, 2, 5, 4.

$$\text{Prosječna ocjena} = \frac{3 + 3 + 4 + 2 + 5 + 4}{6} = \frac{21}{6} = 3.5.$$

Srednja vrijednost

Primjer. Prosječna plaća u tromjesečju: 1500 € , 1200 € , 1800 € .

$$\text{Prosječna plaća} = \frac{1500 + 1200 + 1800}{3} = \frac{4500}{3} = 1500.$$

Primjer. Prosječna ocjena: 3, 3, 4, 2, 5, 4.

$$\text{Prosječna ocjena} = \frac{3 + 3 + 4 + 2 + 5 + 4}{6} = \frac{21}{6} = 3.5.$$

Srednja vrijednost

Za konačan skup podataka x_1, x_2, \dots, x_N , **srednja vrijednost** (\bar{x} ili \bar{x}_N) je aritmetička sredina podataka:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Srednja vrijednost

Primjer. Prosječna plaća u tromjesečju: 1500 € , 1200 € , 1800 € .

$$\text{Prosječna plaća} = \frac{1500 + 1200 + 1800}{3} = \frac{4500}{3} = 1500.$$

Primjer. Prosječna ocjena: 3, 3, 4, 2, 5, 4.

$$\text{Prosječna ocjena} = \frac{3 + 3 + 4 + 2 + 5 + 4}{6} = \frac{21}{6} = 3.5.$$

Srednja vrijednost

Za konačan skup podataka x_1, x_2, \dots, x_N , **srednja vrijednost** (\bar{x} ili \bar{x}_N) je aritmetička sredina podataka:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Σ - oznaka za sumiranje. Još se koristi i $\sum_i x_i$, $\sum x_i$, $\sum x$.

Primjer. Gnijezda lastavica Sljedeći podaci o broju mladunaca lastavica dobiveni su u sklopu istraživanja u King's County, Nova Scotia, Canada u razdobljima podizanja mladih u rasponu od 1996 do 1997. godine. Podaci su dani sljedećom frekvencijskom tablicom:

broj mladunaca u gnijezdu	2	3	4	5	6	7	Σ
frekvencija	1	10	7	26	11	2	57

Koliki je prosječni broj mladunaca po gnijezdu?

Primjer. Gnijezda lastavica Sljedeći podaci o broju mladunaca lastavica dobiveni su u sklopu istraživanja u King's County, Nova Scotia, Canada u razdobljima podizanja mladih u rasponu od 1996 do 1997. godine. Podaci su dani sljedećom frekvencijskom tablicom:

broj mladunaca u gnijezdu	2	3	4	5	6	7	Σ
frekvencija	1	10	7	26	11	2	57

Koliki je prosječni broj mladunaca po gnijezdu?

Uočimo da je prosjek

$$\bar{x} = \frac{2 \cdot 1 + 3 \cdot 10 + 4 \cdot 7 + 5 \cdot 26 + 6 \cdot 11 + 7 \cdot 2}{1 + 10 + 7 + 26 + 11 + 2} = 4.77$$

mladunaca po gnijezdu. Bitno je primijetiti da iako imamo cjelobrojnu varijablu (broj jedinki), prosjek može biti *realan* broj.

Primjer. Gnijezda lastavica Sljedeći podaci o broju mladunaca lastavica dobiveni su u sklopu istraživanja u King's County, Nova Scotia, Canada u razdobljima podizanja mladih u rasponu od 1996 do 1997. godine. Podaci su dani sljedećom frekvencijskom tablicom:

broj mladunaca u gnijezdu	2	3	4	5	6	7	Σ
frekvencija	1	10	7	26	11	2	57

Koliki je prosječni broj mladunaca po gnijezdu?

Uočimo da je prosjek

$$\bar{x} = \frac{2 \cdot 1 + 3 \cdot 10 + 4 \cdot 7 + 5 \cdot 26 + 6 \cdot 11 + 7 \cdot 2}{1 + 10 + 7 + 26 + 11 + 2} = 4.77$$

mladunaca po gnijezdu. Bitno je primijetiti da iako imamo cjelobrojnu varijablu (broj jedinki), prosjek može biti *realan* broj.

Zapišimo sada ovu proceduru u općenitoj formi.

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	\dots	a_k	Σ
frekvencija	f_1	f_2	f_3	\dots	f_k	n

gdje su

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	\dots	a_k	Σ
frekvencija	f_1	f_2	f_3	\dots	f_k	n

gdje su

- $a_1, a_2, a_3, \dots, a_k$ različite vrijednosti promatranog obilježja (ukupno $k \in \mathbf{N}$ različitih vrijednosti);

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	\dots	a_k	Σ
frekvencija	f_1	f_2	f_3	\dots	f_k	n

gdje su

- $a_1, a_2, a_3, \dots, a_k$ različite vrijednosti promatranog obilježja (ukupno $k \in \mathbf{N}$ različitih vrijednosti);
- $f_1, f_2, f_3, \dots, f_k$ pripadne frekvencije (podsjetimo se da je ukupna duljina podataka $n = f_1 + f_2 + f_3 + \dots + f_k$).

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	\dots	a_k	\sum
frekvencija	f_1	f_2	f_3	\dots	f_k	n

gdje su

- $a_1, a_2, a_3, \dots, a_k$ različite vrijednosti promatranog obilježja (ukupno $k \in \mathbf{N}$ različitih vrijednosti);
- $f_1, f_2, f_3, \dots, f_k$ pripadne frekvencije (podsjetimo se da je ukupna duljina podataka $n = f_1 + f_2 + f_3 + \dots + f_k$).

Tada je

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n}.$$

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	\dots	a_k	\sum
frekvencija	f_1	f_2	f_3	\dots	f_k	n

gdje su

- $a_1, a_2, a_3, \dots, a_k$ različite vrijednosti promatranog obilježja (ukupno $k \in \mathbf{N}$ različitih vrijednosti);
- $f_1, f_2, f_3, \dots, f_k$ pripadne frekvencije (podsjetimo se da je ukupna duljina podataka $n = f_1 + f_2 + f_3 + \dots + f_k$).

Tada je

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n}.$$

Računanje srednje vrijednosti iz relativnih frekvencija

S druge strane, da su nam umjesto frekvencija f_i dane relativne frekvencije r_i srednju vrijednost bi mogli izračunati preko formule

$$\bar{x} = \left(\frac{\sum_{i=1}^k f_i x_i}{n} = \sum_{i=1}^k \frac{f_i}{n} x_i \right) = \sum_{i=1}^k r_i x_i.$$

Nedostatci

Primjetite da je aritmetička sredina vrlo osjetljiva na ekstremne podatke u uzorku. Ako se neki poznati međunarodni sportaš odluči uz karijeru studirati i upiše predmet Statistika, prosječni standard studenata u njegovoj grupi, izražen npr. kao aritmetička sredina iznosa koje studenti godišnje mogu potrošiti, bi izuzetno skočio. Dakako za ostale studente se pri tom ništa značajno nije promijenilo.

Nedostatci

Primjetite da je aritmetička sredina vrlo osjetljiva na ekstremne podatke u uzorku. Ako se neki poznati međunarodni sportaš odluči uz karijeru studirati i upiše predmet Statistika, prosječni standard studenata u njegovoj grupi, izražen npr. kao aritmetička sredina iznosa koje studenti godišnje mogu potrošiti, bi izuzetno skočio. Dakako za ostale studente se pri tom ništa značajno nije promijenilo.

Ponekad je poželjna mjera za sredinu podataka koja je manje osjetljiva na ekstremne vrijednosti u uzorku odn. mjera koju možemo smatrati *robustnijom*. Ako uredimo podatke x_1, x_2, \dots, x_n iz uzorka po veličini dobit ćemo uzlazni niz

$$x_{\min} = x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} = x_{\max}.$$

Ove brojeve uobičajeno zovemo **uredajnim statistikama**.

Nedostatci

Primjetite da je aritmetička sredina vrlo osjetljiva na ekstremne podatke u uzorku. Ako se neki poznati međunarodni sportaš odluči uz karijeru studirati i upiše predmet Statistika, prosječni standard studenata u njegovoj grupi, izražen npr. kao aritmetička sredina iznosa koje studenti godišnje mogu potrošiti, bi izuzetno skočio. Dakako za ostale studente se pri tom ništa značajno nije promijenilo.

Ponekad je poželjna mjera za sredinu podataka koja je manje osjetljiva na ekstremne vrijednosti u uzorku odn. mjera koju možemo smatrati *robustnijom*. Ako uredimo podatke x_1, x_2, \dots, x_n iz uzorka po veličini dobit ćemo uzlazni niz

$$x_{\min} = x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} = x_{\max}.$$

Ove brojeve uobičajeno zovemo **uredajnim statistikama**.

Promotrimo sada neke mjere srednje vrijednosti kojima izbjegavamo ovaj problem.

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana.

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana. Podatke poredamo prema veličini, od najmanjeg do najvećeg i odaberemo centralno pozicioniranu vrijednost.

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana. Podatke poredamo prema veličini, od najmanjeg do najvećeg i odaberemo centralno pozicioniranu vrijednost.

Uočimo da gornja definicija ne garantira da takav jedinstveni broj postoji, npr. kada je uzorak parne duljine: 1, 1, 2, 3, 3, 5.

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana. Podatke poredamo prema veličini, od najmanjeg do najvećeg i odaberemo centralno pozicioniranu vrijednost.

Uočimo da gornja definicija ne garantira da takav jedinstveni broj postoji, npr. kada je uzorak parne duljine: 1, 1, 2, 3, 3, 5.

Stoga medijan formalno definiramo na sljedeći način:

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana. Podatke poredamo prema veličini, od najmanjeg do najvećeg i odaberemo centralno pozicioniranu vrijednost.

Uočimo da gornja definicija ne garantira da takav jedinstveni broj postoji, npr. kada je uzorak parne duljine: 1, 1, 2, 3, 3, 5.

Stoga medijan formalno definiramo na sljedeći način:

- Ako je broj podataka neparan - medijan je središnji član niza podataka

$$m = x_{\left(\frac{n+1}{2}\right)}.$$

Medijan

Medijan m je načelno vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana. Podatke poredamo prema veličini, od najmanjeg do najvećeg i odaberemo centralno pozicioniranu vrijednost.

Uočimo da gornja definicija ne garantira da takav jedinstveni broj postoji, npr. kada je uzorak parne duljine: 1, 1, 2, 3, 3, 5.

Stoga medijan formalno definiramo na sljedeći način:

- Ako je broj podataka neparan - medijan je središnji član niza podataka

$$m = x_{(\frac{n+1}{2})}.$$

- Ako je broj podataka paran - medijan je aritmetička sredina dvaju središnjih članova niza podataka

$$m = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	10	11	11	11	17	podaci	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	10	11	11	11	17	podaci	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

$m = 9$

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	10	11	11	11	17	podaci	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

$m = 9$

Primjer. Odredite medijan za podatke

4, 2, 3, 5, 2, 1, 3, 4, 2, 1

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	10	11	11	11	17	podaci	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

$m = 9$

Primjer. Odredite medijan za podatke

4, 2, 3, 5, 2, 1, 3, 4, 2, 1

Podatke poredamo po veličini:

1	1	2	2	2	3	3	4	4	5	podaci
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	redni broj

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	9	10	11	11	11	17	podaci
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

$m = 9$

Primjer. Odredite medijan za podatke

4, 2, 3, 5, 2, 1, 3, 4, 2, 1

Podatke poredamo po veličini:

1	1	2	2	2	3	3	4	4	5	podaci
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	redni broj

$$m = \frac{2+3}{2} = 2.5$$

Određivanje medijana iz distribucije frekvencija.

Ukoliko je zadana distribucija frekvencija, medijan određujemo iz kumulativnih frekvencija.

Odreditmo medijan iz sljedećeg primjera.

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

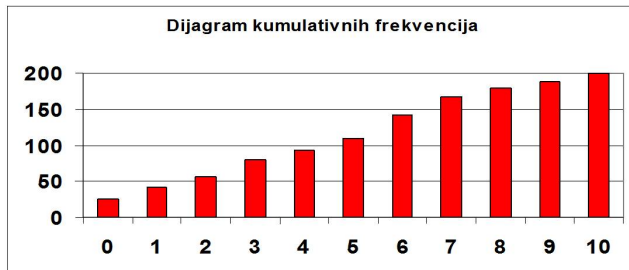
Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

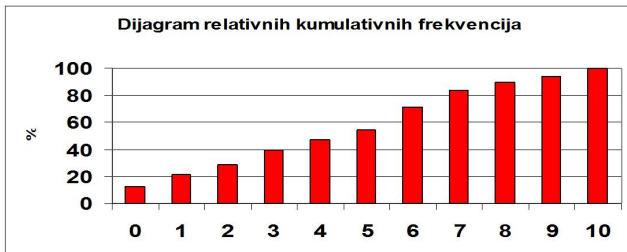
- 10 (50%) podataka je manje ili jednako od 23
- 10 (50%) podataka je veće ili jednako od 24

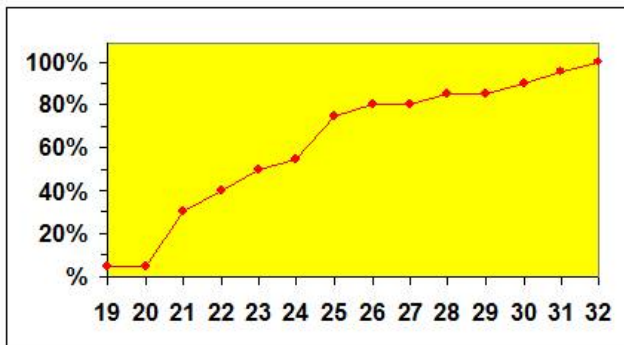
Znači, 23 i 24 su na središnjem mjestu u nizu podataka.

$$m = \frac{23 + 24}{2} = 23.5$$

Primjer. Odredite medijan iz sljedećih grafova kumulativne frekvencije:







Mod

Mod je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost. Uočimo da za razliku od srednje vrijednosti i medijana, varijabla ne mora biti numerička da bi mod bio dobro definiran (ovisi samo o frekvencijama).

Mod

Mod je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost. Uočimo da za razliku od srednje vrijednosti i medijana, varijabla ne mora biti numerička da bi mod bio dobro definiran (ovisi samo o frekvencijama).

Primjer. Odredimo mod za podatke
ACCABACABCAABBCBAACCBAABCCBCAA.

Mod

Mod je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost. Uočimo da za razliku od srednje vrijednosti i medijana, varijabla ne mora biti numerička da bi mod bio dobro definiran (ovisi samo o frekvencijama).

Primjer. Odredimo mod za podatke

ACCABACABCAABBCBAACCBAABCCBCAA.

Podatke prvo presložimo tako da bismo lakše odredili frekvencije pojedinih vrijednosti:

AAAAAAAAAAAA BBBBBBBB CCCCCCCCCC

Mod

Mod je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost. Uočimo da za razliku od srednje vrijednosti i medijana, varijabla ne mora biti numerička da bi mod bio dobro definiran (ovisi samo o frekvencijama).

Primjer. Odredimo mod za podatke

ACCABACABCAABBCBAACCBAABCCBCAA.

Podatke prvo presložimo tako da bismo lakše odredili frekvencije pojedinih vrijednosti:

AAAAAAAAAAAA BBBB BBBB CCCCCCCCCC

Mod je A jer se pojavljuje najviše puta (12).

Mod

Mod je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost. Uočimo da za razliku od srednje vrijednosti i medijana, varijabla ne mora biti numerička da bi mod bio dobro definiran (ovisi samo o frekvencijama).

Primjer. Odredimo mod za podatke
ACCABACABCAABBCBAACCBAABCCBCAA.

Podatke prvo presložimo tako da bismo lakše odredili frekvencije pojedinih vrijednosti:
AAAAAAAAAAAA BBBBBBBBBB CCCCCCCCCC

Mod je A jer se pojavljuje najviše puta (12).

Obilježje može imati i više modova. Npr. u nizu

1 1 2 2 2 3 4 4 5 5 5 6 7 7

mod je 2 i 5 (dva moda).

U ovom slučaju govorimo o **bimodalnoj** distribuciji frekvencija.

ZADATAK.

Izračunajte srednju vrijednost, medijan i mod za sljedeći niz podataka:

2	5	5	6	6
8	9	9	11	11
12	13	16	19	20
25	26	26	29	30
30	32	33	34	35
39	39	40	42	43