Uvod u deskriptivnu i inferencijalnu statistiku

Doc. dr. sc. Irena Burić
Odjel za psihologiju
Sveučilište u Zadru

e-mail: inekic@unizd.hr

NASTAVNI MATERIJALI PRIPREMLJENI SU U SKLADU S ODLUKOM POVJERENSTVA ZA IZDAVAČKU DJELATNOST SVEUČILIŠTA U ZADRU

STUDIJ: Preddiplomski studij psihologije
SEMESTAR: Zimski
NAZIV KOLEGIJA: **Uvod u deskriptivnu i inferencijalnu statistiku**

- **Cilj kolegija** je upoznavanje s osnovnim pojmovima deskriptivne statistike (grafičko i tabelarno prikazivanje rezultata, mjere centralne tendencije, mjere raspršenja, distribucija rezultata te usporedba i položaj pojedinog rezultata u grupi), uvodnim sadržajima iz domene inferencijalne statistike (model normalne distribucije, z-vrijednosti, procjena populacijskih parametara na temelju vrijednosti dobivenih na uzorku, provjeravanje razlika među aritmetičkim sredinama, hi-kvadrat, određivanje korelacije među varijablama) te s pojmom statističke snage istraživanja. Kroz ovaj kolegij, studenti bi trebali moći definirati, razlikovati i razumjeti logiku osnovnih pojmova i metoda deskriptivne i inferencijalne statistike.

OČEKIVANI ISHODI UČENJA

- Klasificirati, opisati i prikazati rezultate psihologijskog mjerenja.
- Definirati, razlikovati i razumjeti logiku osnovnih pojmova i metoda deskriptivne i inferencijalne statistike.
- Izračunati različite mjere središnjih vrijednosti i raspršenja rezultata, odrediti granice pouzdanosti, izračunati t-test, hi-kvadrat test te Personov koeficijent korelacije.
- Odabrati primjerenu statističku metodu za obradu određenih rezultata.
- Dobivene rezultate statističkih analiza adekvatno interpretirati.

SADRŽAJ KOLEGIJA

1. Uloga statističkih metoda u psihologiji i istraživačkom procesu
2. Varijable u psihologiji i skale mjerenja
3. Tabelarno i grafičko prikazivanje rezultata
4. Struktura rezultata psihologijskog mjerenja i model normalne distribucije
5. Mjere centralne tendencije
6. Mjere varijabilnosti
7. Položaj pojedinog rezultata u grupi
8. Zaključivanje u statistici
9. Procjena populacijskih parametara i granice pouzdanosti
10. Testiranje hipoteza i pojam statističke značajnosti
11. Provjeravanje razlika među aritmetičkim sredinama
12. Korelacija
13. Hi-kvadrat
14. Statistička snaga istraživanja i veličina učinka

LITERATURA

OBVEZNA:

- Petz, B., Kolesarić, V. i Ivanec, D. (2012). *Petzova statistika: Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap.

DOPUNSKA:

- Aron, A., Coups, E. J., Aron, E. N. (2013). *Statistics for psychology*. Pearson.
- Howell, D. C. (2002). *Statistical methods for psychology*. Duxbury: Wadsworth Group.
- Kolesarić, V. i Petz, B. (2003). *Statistički rječnik*. Jastrebarsko: Naklada Slap.

1. Uloga statističkih metoda u psihologiji i istraživačkom procesu

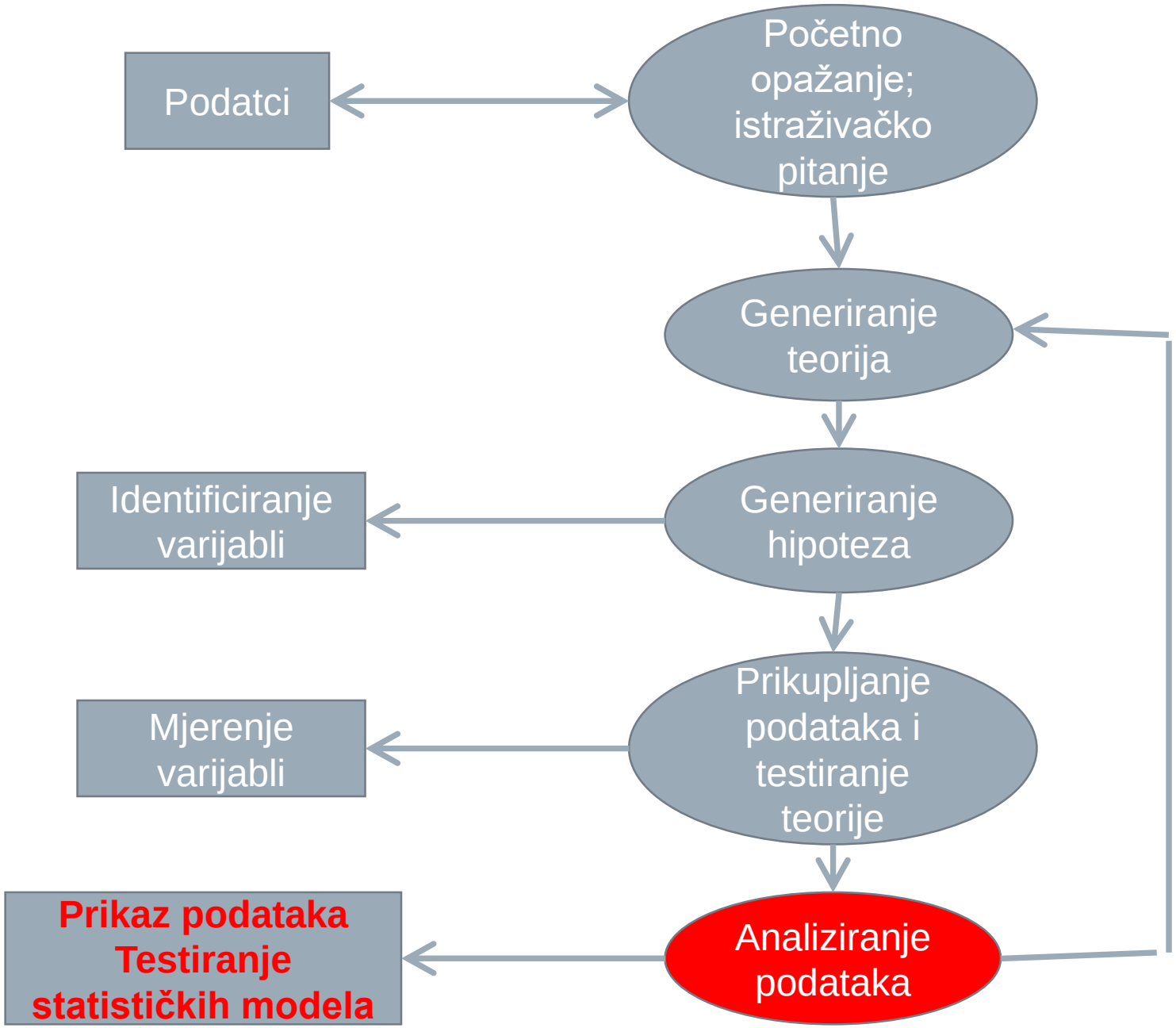
Zašto kolegij statistike na studiju psihologije?

- psihologija je društvena ***znanost*** koja proučava „zašto se ljudi ponašaju na način kako se ponašaju”
- tko su znanstvenici?
  - znatiželjni ljudi koji nastoje pronaći odgovore na zanimljiva i važna pitanja koristeći se znanstvenim metodama
  - znanstvenici su ljudi koji provode ***znanstvena istraživanja***

Zašto kolegij statistike na studiju psihologije?

- Istraživanja u psihologiji često rezultiraju podatcima koji su **kvantitativne** prirode, odnosno izraženi su *brojevima*
- Za razliku od kvantitativnih, podatci mogu biti i **kvalitativni** – riječi, fotografije i sl.
- Pomoću statističkih postupaka *brojčani podatci* obrađujuse kako bi istraživači dobili odgovor na svoja istraživačka pitanja – **STATISTIČKA OBRADA PODATAKA**
- Primjer istraživačkog pitanja: Koliko iznosi visina brucošica psihologije u Hrvatskoj?

Shematski prikaz istraživačkog procesa



Statistička analiza

- Statistički postupci dijele se u dvije široke skupine:
  - **deskriptivni** (M, SD, izgled distribucije…)
  - **inferencijalni** (t-test, korelacija, hi-kvadrat...)

Deskriptivna statistika

- opisuje činjenice dobivene opažanjem ili mjerenjem neke pojave
- primjeri:
  - prosječna tjelesna petogodišnjih dječaka
  - varijacije u inteligenciji
  - povezanost između motivacije i školskog uspjeha
  - oblik distribucije nesreća na radu
  - itd.

Inferencijalna statistika

- statistički postupci koji nam omogućuju **testiranje istraživačkih hipoteza**, odnosno **zaključivanje**
- iz podataka dobivenih na pojedinačnim slučajevima (tj. **uzorku**) zaključujemo o zakonitostima koji vrijede za čitavu **populaciju –** iz pojedinačnog slučaja zaključujemo općenito
- 100 %-tna sigurnost u statističkim zaključivanjima je praktički nemoguća (**ali je moguće postići sigurnost koja je vrlo blizu 100 %!)**
- npr. spolne razlike u verbalnim i spacijalnim sposobnostima

Zašto je psiholozima važna statistika?

1. za razumijevanje rezultata istraživanja drugih istraživača
   - praćenje znanstvene i stručne literature
   - npr. korelacija između samopoštovanja i školskog uspjeha iznosi r=0.50; to znači da su samopoštovanje i školski uspjeh umjereno i pozitivno povezani – učenici koji imaju bolji školski uspjeh ujedno imaju i veće samopoštovanje i obrnuto
2. za vlastiti znanstveni rad
   - planiranje istraživanja i obrada dobivenih rezultata (praktikumi, završni i diplomski rad)
3. za razvijanje **analitičkog i kritičkog mišljenja**
   - kritička evaluacija informacija i logička analiza rezultata psihologijskih istraživanja

Zašto ljudi ne vole statistiku?

(Pogrešna?) uvjerenja o statistici:

- statistika je suhoparna i dosadna
- statistički simboli djeluju čudno i nepoznato (χ^2, $\sum$, σ)
- statistiku je nemoguće savladati bez znanja matematike (to je točno, ali samo za ona elementarna znanja matematike)
- statistika laže
  - „*There are three kinds of lies: lies, damn lies and statistics*” (Benjamin Disraeli, britanski premijer, 19. stoljeće)
  - *„Figures don’t lie, but liers figure”.* (Mark Twain?)

Laže li statistika?

- Primjer. Neko istraživanje pokazalo je da studenti pušači imaju slabije akademsko postignuće. Kako ćemo interpretirati ovaj nalaz?
  - pušenje uzrokuje slabije akademsko postignuće?
  - moguća alternativna objašnjenja: a. slabije postignuće dovodi do češćeg posezanja za cigaretom; b. studenti koji više izlaze, više i puše te imaju manje vremena za učenje

Je li statistika bauk za studente psihologije?

- statistika ≠ matematika
- glavni se statistički postupci, principi i način mišljenja mogu usvojiti potpuno ***logičkim putem***, a od matematike je potrebno znati samo 4 osnovne operacije: *zbrajanje, oduzimanje, množenje, dijeljenje i vađenje drugog korijena* (a i za to postoje kalkulatori i računala ☺)
- matematika je potrebna samo profesionalnim statističarima koji izrađuju statističke metode i formule, a psiholozima je dovoljno znati ***razumjeti ih, ispravno ih primijeniti i interpretirati dobivene rezultate!***

„Statistički način mišljenja”

- = prilaziti istraživanju pojedinih problema na način koji osigurava dobivanje što pouzdanijih podataka: počevši od definiranja problema, definiranja metode istraživanja, preko biranja uzorka na kojem će se obavljati mjerenje, pa sve do odabira metode obrade podataka
- da bi se to moglo, potrebno je **razumjeti logiku statističkog mišljenja i zaključivanja**, a to se uči na ovom kolegiju
- bez statističkog mišljenja nema ni znanstvenog mišljenja

Računala i statistika

- SPSS, SAS, STATISTICA – programi za statističku obradu podataka
- omogućuju brze izračune velikog broja podataka i ostavljaju više vremena za promišljanje o prikladnosti pojedinih analiza, značenju i interpretaciji rezultata
- ali, računalo ne razumije ono što računa
  - **„Garbage in, garbage out!”**

Kako biti uspješan na ovom kolegiju?

1. **Obratiti pažnju na koncepte i principe.** Ovaj kolegij je više „logički” nego „matematički” – kada se vježbaju zadatci, treba razmišljati *zašto* se provode određeni koraci umjesto da se samo pokušava mehanički memorirati postupak.
2. **Redovito pohađati nastavu.** Gradivo kolegija se logički nadovezuje jedno na drugo – ako se ne savlada ono što se učilo ranije, kasnije lekcije djelovat će nerazumljivo i besmisleno.
3. **Učiti u paru ili grupi.** Često je najbolji način da se postignete što veće razumijevanje statistike upravo objašnjavanje gradiva drugim kolegama.
4. **Pravovremeno doći na konzultacije.**
4. **Ne plašiti se ovog kolegija!** (*proročanstvo koje samo sebe ispunjava*?)

2. Varijable u psihologiji i skale mjerenja

Što je konstrukt?

- psihički fenomen, pojava, ponašanje koje je predmet interesa ili istraživanja u psihologiji
- Primjeri konstrukata u psihologiji: inteligencija, motivacija, senzorno pamćenje, stres, depresivnost, pažnja, prosocijalno ponašanje itd.

Što je konstrukt?

- kako možemo *mjeriti* navedene psihološke konstrukte?
- što je karakteristično za navedene primjere psiholoških konstrukata?
  - većina konstrukata u psihologiji je na neki način „nevidljiva” ili „prikrivena”, tj. nije ih moguće **direktno** mjeriti

Mjerenje psiholoških konstrukata

- međutim, psiholozi su se domislili načinima kako mjeriti ove „nevidljive” konstrukte
- Primjeri:
  - inteligenciju mjerimo testom inteligencije
  - depresivnost mjerimo skalom depresivnosti
  - prosocijalno ponašanje mjerimo opažanjem nečijeg ponašanja u određenim okolnostima

Mjerenje

- MJERENJE = pridjeljivanje brojeva objektima, pojavama itd., u skladu s pravilima (Kolesarić i Petz, 2003); proces prikupljanja podataka u psihologijskim istraživanjima
- ***mjeriti*** (tj. ispitivati) možemo određene osobine ili fenomene na pojedincu ili skupini pojedinaca
- podatci dobiveni mjerenjem mogu biti različite vrste što određuje način njihova opisivanja i analiziranja

Vježba

- Procijenite na skali od 0 do 10 razinu stresa kojoj ste bili izloženi u periodu polaganja državne mature
  - pri tome brojevi imaju sljedeće značenje: 0 = uopće nisam bio/la pod stresom, a 10 = totalno sam bio/la pod stresom
- Razmislite...
  - Što smo u ovom primjeru mjerili?
  - Kojom metodom?
  - Na koji još način bismo mogli izmjeriti razinu stresa?

Varijable, vrijednosti i rezultati

- U prethodnom primjeru razina stresa jest KONSTRUKT ili VARIJABLA, koja može poprimiti VRIJEDNOST od 0 do 10, a vaše pojedinačne vrijednosti predstavljaju REZULTATE.

Varijable, vrijednosti i rezultati

- **Varijabla** – karakteristika koja je naš *predmet mjerenja, tj. konstrukt* koji želimo zahvatiti i koja može poprimiti različite vrijednosti
- **Vrijednost** – mogući *broj* ili *kategorija* koju može imati neki rezultat
- **Rezultat** – vrijednost koju pojedinac postiže u varijabli

  Pitanje: Zašto se razina stresa iz našeg primjera naziva „VARIJABLA”?

Što je varijabla?

- PRIMJER 1. mjerenje jednostavnog vremena reakcije na svjetlosni podražaj na **jednom** ispitaniku
  - kako bismo dobili što precizniju i točniju mjeru VR-a određenog ispitanika, napravili smo 50 mjerenja i dobili 50 brojčanih vrijednosti izraženih u ms
- PRIMJER 2. mjerenje inteligencije na **skupini** studenata
  - kako bismo dobili što precizniju i točniju mjeru stupnja razvijenosti inteligencije hrvatske studentske populacije, pomoću *RPM – Advanced* izmjerili smo inteligenciju stotini po slučaju odabranih studenata i dobili 100 brojčanih vrijednosti izraženih kao broj točno riješenih zadataka u testu

  - Pitanje: Razlikuju li se *međusobno* dobivene brojčane vrijednosti tj. rezultati mjerenja iz oba primjera?

Što je varijabla?

- odgovor na prethodno pitanje je naravno – DA
- rezultati dobiveni mjerenjem vremena reakcije *VARIRAJU* unutar pojedinca, a rezultati dobiveni mjerenjem inteligencije među pojedincima
- zato vrijeme reakcije i inteligencija, u ovim primjerima, imaju status VARIJABLI

Što je varijabla?

- primjeri varijabli u psihologiji: spol, socioekomomski status, vrijeme reakcije, inteligencija, anksioznost, neuroticizam itd.
- te pojave ili osobine mogu poprimiti različite veličine, tj. vrijednosti, *spontano* ili *pod utjecajem* nekih drugih čimbenika
- suprotno od „konstante”

Što je varijabla?

- varijabla = nešto što varira, što može poprimiti različite vrijednosti (oznake)
- osobine ili svojstva stvari, pojava ili osoba, koje je na temelju ***operacionalne definicije*** moguće opažati i/ili mjeriti i koje predstavljaju prihvatljivo prepoznatljiv, jasan i jednoznačan entitet u nekom području ljudskih aktivnosti
- **operacionalna definicija** – instrument odnosno sredstvo kojim nastojimo izmjeriti neki psihološki konstrukt ili varijablu

Latentne vs. manifestne varijable

- **latentne varijable** = psihološki konstrukti koji su „nevidljivi” ili „prikriveni” i koje nije moguće direktno mjeriti
- **manifestne varijable** = operacionalne definicije latentnih varijabli; „vidljivi” i „očiti” ***indikatori*** latentnih varijabli
- neku latentnu varijablu može predstavljati jedan ili više manifestnih indikatora

Latentna varijabla/ konstrukt	Operacionalna definicija	Manifestni indikatori
Inteligencija	Rezultat na Problemnom testu	Odgovori na pojedinim zadatcima u testu
Stres	Stupanj samoprocijenjenog intenziteta stresa na skali od 1 do 10	Broj koji je ispitanik zaokružio na skali
Vrijeme reakcije	Vrijeme izraženo u ms koje protekne od trenutka zadavanja nekog podražaja do reakcije	Vrijednosti višekratnog mjerenja vremena reakcije
Ekstraverzija	Rezultat na Goldbergovom IPIP upitniku ličnosti	Odgovori na pojedinim česticama skale

Što je varijabla?

- Varijable mogu biti:
  - numeričke (**kvantitativne**)
    - vrijednosti su brojčane (primjeri: položaj pacijenta NN na listi čekanja za CT; samoprocijenjena razina stresa kod brucoša; dužina stopala)
  - kategorijalne (**kvalitativne**, nominalne)
    - vrijednosti su nazivi ili kategorije (primjeri: spol – muški, ženski; psihijatrijska dijagnoza – depresija, PTSP, opsesivno-kompulzivni poremećaj)

Što je varijabla?

- **diskretne**
  - vrijednosti su jasno međusobno odijeljene (primjeri: spol; bračni status; broj posjeta zubaru u proteklih godinu dana; broj izostanaka s nastave na fakultetu u semestru)
- **kontinuirane**
  - u teoriji, postoji neograničen broj vrijednosti između neke dvije vrijednosti (primjeri: dob, visina, vrijeme, razina stresa, psihoticizam)

Mjerenje

- vrste varijabli usko su povezane s razinama mjerenja
- **mjerenje**: pridjeljivanje brojeva (ili slova, oznaka) objektima, pojavama, pojedincima itd. u skladu s određenim pravilima
- to pravilo može biti *bilo koje konzistentno* pravilo (osim pridjeljivanja po slučaju!)
- npr. mjerenje (ispitivanje) vremena reakcije, etničke pripadnosti, socioekonomskog statusa itd.

Što je mjerenje (u psihologiji)?

› za mjerenje su važne 4 značajke sustava brojeva:

1. svaki brojčani simbol ima svoj identitet, tj. on je sigurno različit od svakog drugog broja i reprezentira uvijek isto (npr. spol) – ***utvrđivanje identiteta***
2. svi brojevi koji nisu identični, manji su ili veći, pa se mogu poredati po veličini (npr. socio-ekonomski status) – ***utvrđivanje reda***
3. za razlike među brojevima vrijedi isto što i za same brojeve pa se može utvrditi red među tim razlikama (npr. neuroticizam) **–** ***utvrđivanje razlika***
4. sustav brojeva sadrži jedinstven broj tj. nulu koji pokazuje odsutnost bilo kakve pojave ili količine (npr. dužina) – ***utvrđivanje omjera***

Što je skala mjerenja?

› određena je svojstvima brojčanog sustava koja određuju **pravila** pridjeljivanja brojeva pojavama

› ta pravila uključuju:

1. utvrđivanje ***identiteta*** **–** NOMINALNA SKALA
2. utvrđivanje ***reda*** **–** ORDINALNA SKALA
3. utvrđivanje ***razlika*** – INTERVALNA SKALA
4. utvrđivanje ***omjera*** **–** OMJERNA SKALA

› neku skalu mjerenja definira mogućnost manipuliranja brojčanim vrijednostima koje će je ostaviti invarijantnom

NOMINALNA SKALA

- to zapravo – i nije skala!
- rezultate mjerenja ili opažanja svrstavamo u kvalitativne ili približno kvantitativne klase (npr. ispitanike svrstavamo u muškarce i žene)
- jedini računski postupak: brojenje slučajeva u svakoj kategoriji
- **brojevi**, kojima se često označavaju pojedine kategorije, nemaju nikakve kvantitativne vrijednosti, već služe jedino kao **„etiketa”** (npr. muškarci=1, žene=2) i mogu biti zamijenjeni i slovima ili nekim drugim simbolima (npr. muškarci=A, žene=B)
- pravilo: **utvrđivanje identiteta ili jednakosti** – ili je a=b ili je a ≠ b; ako je a=b, tada je b=a
- npr. muškarci=1, žene=2 ili žene=1, muškarci=2 → dodjeljivanje brojeva je proizvoljno – muškarci su i dalje muškarci, a žene su i dalje žene, tj. skala je i nakon zamjene ostala **invarijantna, tj. nepromijenjena**

ORDINALNA SKALA

- označavanje redoslijeda, tj. ranga elemenata koje opažamo ili mjerimo
- određivanje „**veće od**” ili „**manje od**” unutar neke klase pojava ili objekata koje se može izvoditi na temelju kvantitativnih razlika ili promjena unutar neke kvalitativno definirane klase ili na temelju čestine (frekvencije) kojom se nešto pojavljuje
- npr. kućni brojevi neke ulice, redoslijed trkača na cilju, školske ocjene
- rangovi se pravilno redaju od 1 na dalje, ali stvarne razlike između pojedinih elemenata su nam nepoznate (npr. poredak skijaša ili školske ocjene)

INTERVALNA SKALA

- određivanje jednakosti ili razlika
- skala s ekvidistantnim jedinicama: razlika između 6 i 7 cm jednaka je kao i razlika između 13 i 14 cm, a to je – 1 cm
- pravilo: (a-b)+(b-c)=a-c
- mora postojati korespondencija između razlika u pojavi koja je predmet mjerenja i razlika u skalnim brojevima
- stvarna **nulta točka ne postoji,** već je samo arbitrarno određena – zato kada zbrajamo razlike na ovoj skali, ne zbrajamo i njihove apsolutne vrijednosti
- transformacije koje ovu skalu ostavljaju invarijantnom moraju biti monotone i linearne, tj. njezine vrijednosti mogu se jedino zbrajati i množiti s konstantnom

OMJERNA SKALA

- najpoželjnija skala
- ne samo da ima ekvidistantne jedinice već posjeduje i **apsolutnu nultu vrijednost**
- jedino se na toj skali može reći da je nešto npr. „tri puta” veće od nečega drugog (npr. dijete od 30 kg tri je puta teže od djeteta od 10 kg)
- primjeri ove skale: težina, dužina, vrijeme, brzina
- kod ove skale dopušteni su svi statistički postupci

Razmislite...

- Na kojoj skali se izražava izmjerena temperatura?
  - °C
  - °F
  - kelvin

- ˚C i ˚F pripadaju *intervalnoj* skali. Zašto?
  - 0 je arbitrarno određena – 0˚C i 0 ˚F ne označavaju odsustvo temperature
  - može se ustvrditi da je razlika između 10 i 20 ˚C jednaka razlici između 40 i 50 ˚C, ali ne može se reći da je 20 ˚C dva puta toplije od 10 ˚C
  - dokaz: ako konvertiramo celzijuse u farenhajte (što je posve legitimno), čak ni ne dobivamo iste omjere u vrijednosti temperature (40 ˚F : 80 ˚F ≠ 4.4 ˚C : 26.7 ˚C)!
- kelvin pripada *omjernoj* skali – 0 je apsolutna nula jer je 0 k najniža moguća temperatura kod koje prestaje kretanje atoma (0 k=-273.15 ˚C)

Zašto su skale mjerenja uopće važne?

- skala na kojoj se nalaze rezultati mjerenja određuje razinu računskih operacija, a time i *vrstu statističkih postupaka* koje je opravdano koristiti prilikom njihova analiziranja
- nominalna i ordinalna skala dopuštaju nam samo mali i ograničen broj računskih operacija i statističkih postupaka (brojanje, <, >)
- omjerna skala najpoželjnija (ali u psihologiji rijetka!) jer su nam dopušteni svi računski postupci
- intervalna skala je *dovoljno* dobra

Zašto su skale mjerenja uopće važne?

- ne postoji jasan i uniforman dogovor o tome koji rezultati mjerenja pripadaju kojoj skali – to je često odluka pojedinog istraživača
- same brojeve možemo podvrgnuti bilo kakvim matematičkim operacijama, ali da bi ispravno interpretirali svoje rezultate, brojevi moraju biti na smislen način povezani s pojavama ili osobinama koje predstavljaju – *metodološko, a ne statističko pitanje*
- „The numbers do not remember where they came from” (Lord, 1953)

Razmislite...

- Kojoj skali pripadaju rezultati izraženi kao IQ?
  - **omjernoj** (IQ teoretski može biti nula; dakle, skala ima apsolutnu nultu vrijednost)

    ILI
  - **intervalnoj** (skala nema apsolutnu nulu, ali razlika između IQ=50 i IQ=60 te između IQ=110 i IQ=120 je ista, tj. iznosi 10 jedinica)?

    ILI
  - **ordinalnoj** (jer mi zapravo ne možemo znati jesu li ove razlike zapravo ekvidistantne i u samoj mjerenoj pojavi tj. „količini” inteligencije)?

- u psihologiji rijetko raspolažemo takvim rezultatima mjerenja koje možemo smjestiti na pravu intervalnu skalu
- no, iskustvo je pokazalo da čak i rezultate mjerenja koji se nalaze na „približno” intervalnim skalama možemo obrađivati i tretirati kao da se nalaze na „stvarno” intervalnim skalama

3. Tablično i grafičko prikazivanje rezultata

Opisivanje podataka

- provedbom istraživanja prikupljamo određene podatke koje je potom potrebno opisati – to se postiže grafičkim i tabličnim prikazima

Tablica frekvencija

- pokazuje koliko se puta određeni rezultat pojavio u nekoj skupini rezultata
- služi za prikaz rezultata na brojčanim i kategorijalnim varijablama
- omogućuje lakši i brži uvid u rezultate

Tablica frekvencija – NUMERIČKA varijabla

- **Primjer.** Rezultati 30 studenata dobiveni samoprocjenama razine stresa na skali od 0 do 10 (Aron i sur., 1995):
  - 8, 7, 4, 10, 8, 6, 8, 9, 9, 7, 3, 7, 6, 5, 0, 9, 10, 7, 7, 3, 6, 7, 5, 2, 1, 6, 7, 10, 8, 8
  - zadatak: prikazati ove rezultate pomoću *tablice frekvencija*

Tablica frekvencija – NUMERIČKA varijabla

Tablica 1. Studentske samoprocjene razine stresa (N=30)		
Procjena stresa	**Frekvencija**	**Postotak**
0	1	3.3
1	1	3.3
2	1	3.3
3	2	6.7
4	1	3.3
5	2	6.7
6	4	13.3
7	7	23.3
8	5	16.7
9	3	10.0
10	3	10.0

Tablica frekvencija – grupirani rezultati

- ako se u nekom skupu rezultata pojavljuje velik broj međusobno različitih rezultata, provodi se grupiranje rezultata u razrede (radi veće preglednosti)
- razred=ograničeni interval, tj. raspon rezultata
- npr. ako su se nekim mjerenjem dobili rezultati koji se kreću u rasponu od 1 do 25, rezultate možemo grupirati u sljedećih 5 razreda: 1-5, 6-10, 11-15, 16-20, 21-25
- **gornja i donja granica razreda**: najmanja odnosno najveća vrijednost koju neki rezultat može poprimiti unutar jednog razreda – npr. donja granica prvog razreda (iz prethodnog primjera) je 1, a gornja granica je 5.

Tablica grupiranih frekvencija – NUMERIČKA varijabla

Tablica 2. Studentske samoprocjene razine stresa (grupirane frekvencije)		
Razred	**Frekvencija**	**Postotak**
0-1	2	6.7
2-3	3	10.0
4-5	3	10.0
6-7	11	36.7
8-9	8	26.7
10-11	3	10.0

Tablica frekvencija – KATEGORIJALNA varijabla

- **Primjer:** Istraživači su od 208 studenata tražili da navedu najbliskiju osobu u svom životu: 33 studenta su odabrala člana obitelji, 76 prijatelja, 92 (romantičnog) partnera, a 7 ih je odabralo neku drugu osobu (Aron, Aron i Smollan, 1992).
- zadatak: prikazati ove rezultate pomoću *tablice frekvencija*

Tablica frekvencija – KATEGORIJALNA varijabla

Tablica 3: Najbliskija osoba u životu 208 studenata

Najbliskija osoba	Frekvencija	Postotak
Član obitelji	33	15.9
Prijatelj	76	36.5
(Romantični) partner	92	44.2
Drugi	7	3.4

Napomene za tabelarno prikazivanje rezultata

- svaka tablica mora imati kratak i jasan naslov koji se stavlja na vrh, tj. iznad tablice
- dodatna objašnjenja ili legende se stavljaju odmah ispod tablice
- stupci i redci moraju biti logički poredani da olakšaju usporedbu
- tablice moraju biti uredne i čitljive

Grafički prikazi

- omogućuju:
  - preglednost
  - brzo i lako razumijevanje rezultata
  - uspješniju komunikacija među stručnjacima
  - uočavanje posebnih i neočekivanih karakteristika rezultata (npr. aberantni rezultati, asimetrične distribucije itd.)
- rezultati se, ovisno o njihovoj vrsti, mogu grafički prikazati na različite načine

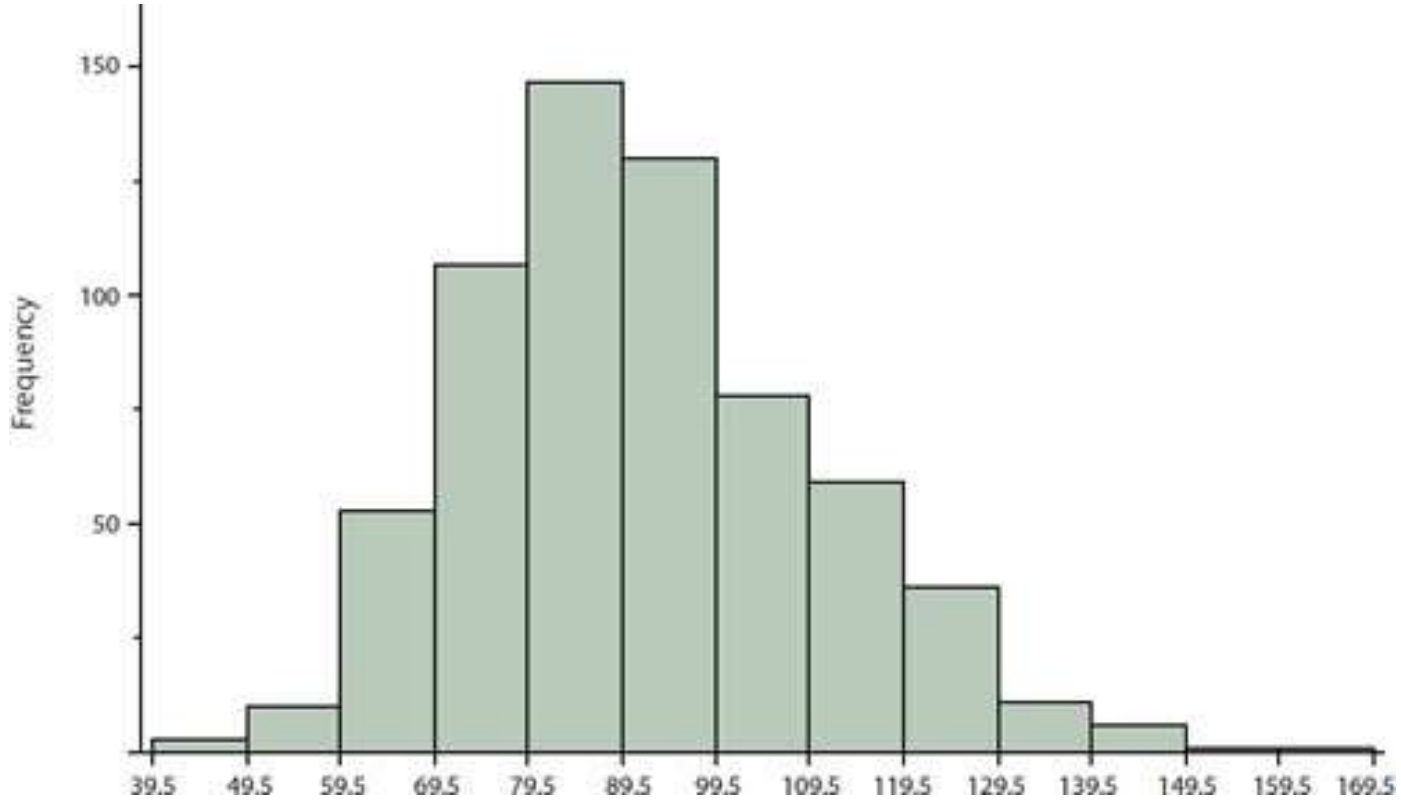
Grafički prikazi

- za rezultate izražene na ***intervalnim*** i ***omjernim*** skalama najčešće se koriste HISTOGRAM i POLIGON FREKVENCIJA
  - › prikaz *distribucije* rezultata u koordinatnom sustavu
- za rezultate izražene na ***nominalnim*** i ***ordinalnim*** skalama najčešće se koriste KRUŽNI DIJAGRAM te VERTIKALNI i HORIZONTALNI STUPCI

Histogram – prikaz podataka na intervalnoj ili omjernoj skali

- grafički prikaz u koordinatnom sustavu pomoću **stupaca**
- na apscisi se nalaze *vrijednosti rezultata*, a širinom svakog stupca označen je jedan rezultat (ili gornje granice rezultata ako su rezultati grupirani u razrede)
- na ordinati se nalazi *čestina* rezultata, a visina svakog stupca proporcionalna je čestini kojom se pojavio neki rezultat, odnosno površina stupca odgovara broju rezultata u razredu (ako su rezultati grupirani)
- obično su stupci spojeni, tj. između njih **ne** dolazi razmak

Primjer histograma



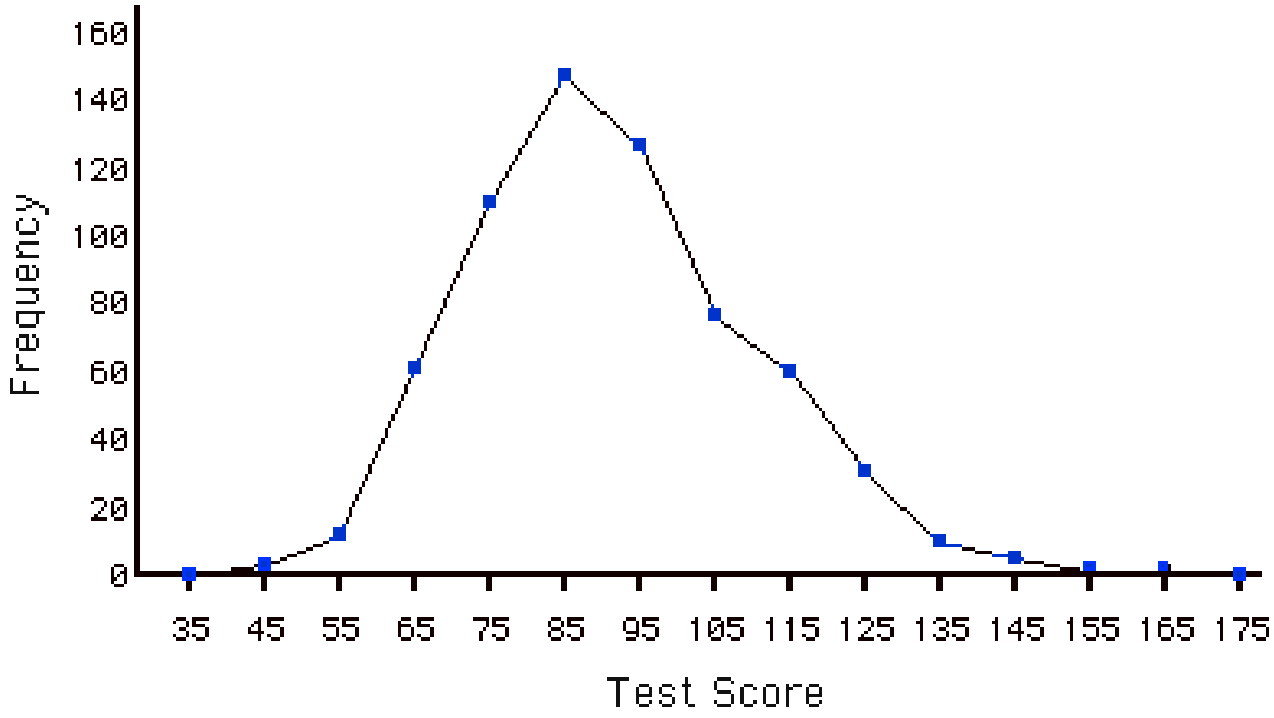
Slika 1. Histogram rezultata na ispitu iz psihologije

Preuzeto s: **http://onlinestatbook.com/2/graphing_distributions/histograms.html**

Poligon frekvencija – prikaz podataka na intervalnoj ili omjernoj skali

- na apscisu se unose vrijednosti rezultata (ako su rezultati grupirani u razrede, unosi se sredina razreda), a na ordinatu frekvencije, pa se spajanjem točaka u koordinatnom sustavu dobije crtež distribucije neke skupine rezultata
- cjelokupna površina krivulje odgovara ukupnoj frekvenciji, tj. ukupnom broju rezultata (N), ali samo ako je krivulja ***„uzemljena”,*** tj. lijevo i desno spuštena na apscisu

Primjer poligona frekvencija



Slika 2. Poligon frekvencija rezultata na ispitu iz psihologije

Preuzeto s: **http://onlinestatbook.com/2/graphing_distributions/freq_poly.html**

Poligon frekvencija vs. histogram

- poligon frekvencija *manje je precizan* način grafičkog prikazivanja od histograma
- kod histograma površina svakog stupca točno odgovara i frekvenciji rezultata unutar pojedinog razreda (ako su rezultati grupirani u razrede), dok kod poligona frekvencija samo cjelokupna površina krivulje odgovara ukupnoj frekvenciji, dakle ukupnom broju rezultata (N) – histogram ima prednost, osobito kada su rezultati grupirani u razrede

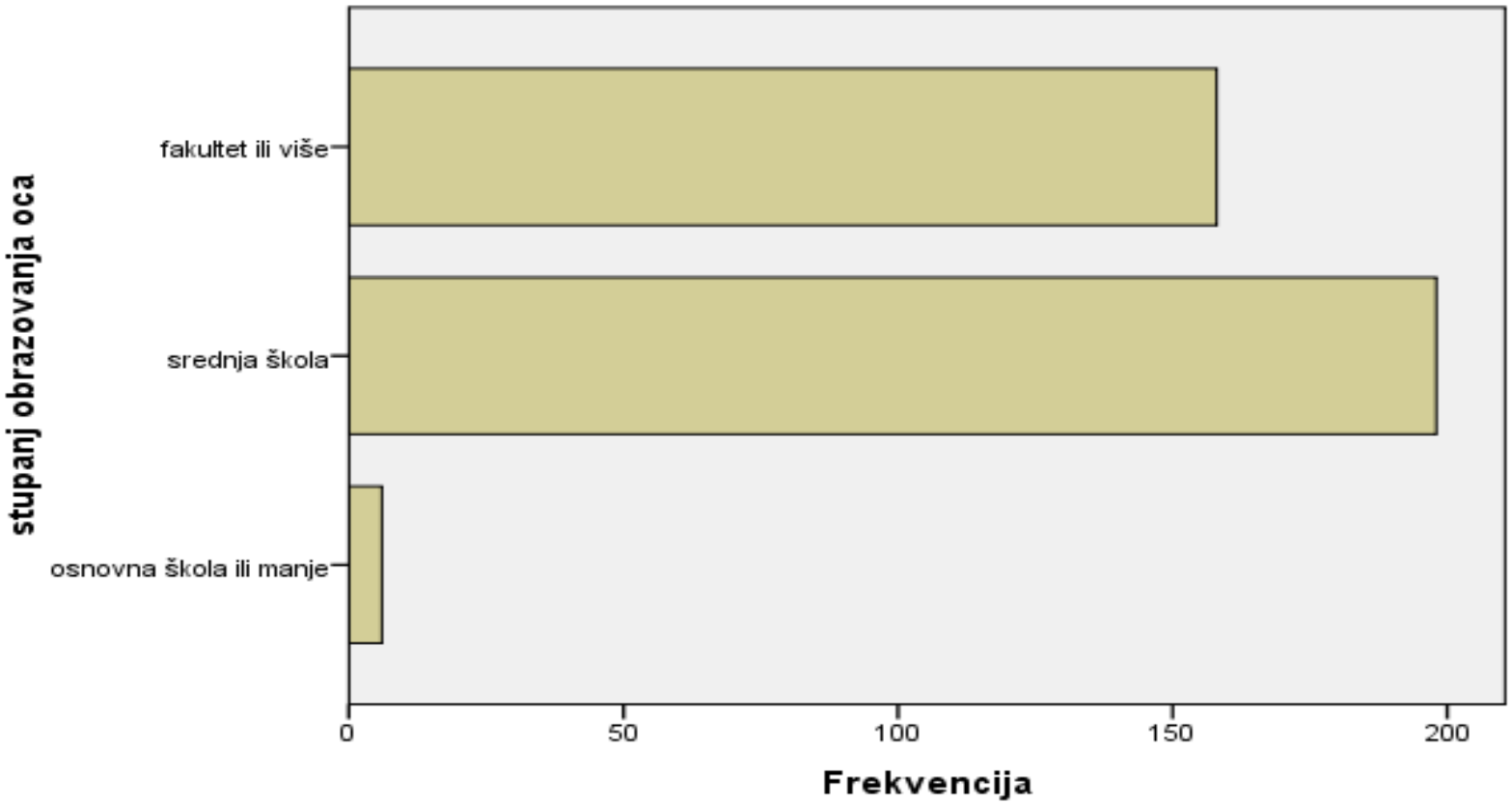
Poligon frekvencija vs. histogram

- poligon frekvencija se, unatoč manjoj preciznosti, češće koristi od histograma jer:
  - krivulja se više približava realnoj distribuciji rezultata od histograma
  - kod prikaza dvije ili više distribucija na jednoj slici, prikazivanje histogramima dalo bi posve nepreglednu sliku

Napomene za crtanje histograma ili poligona frekvencije

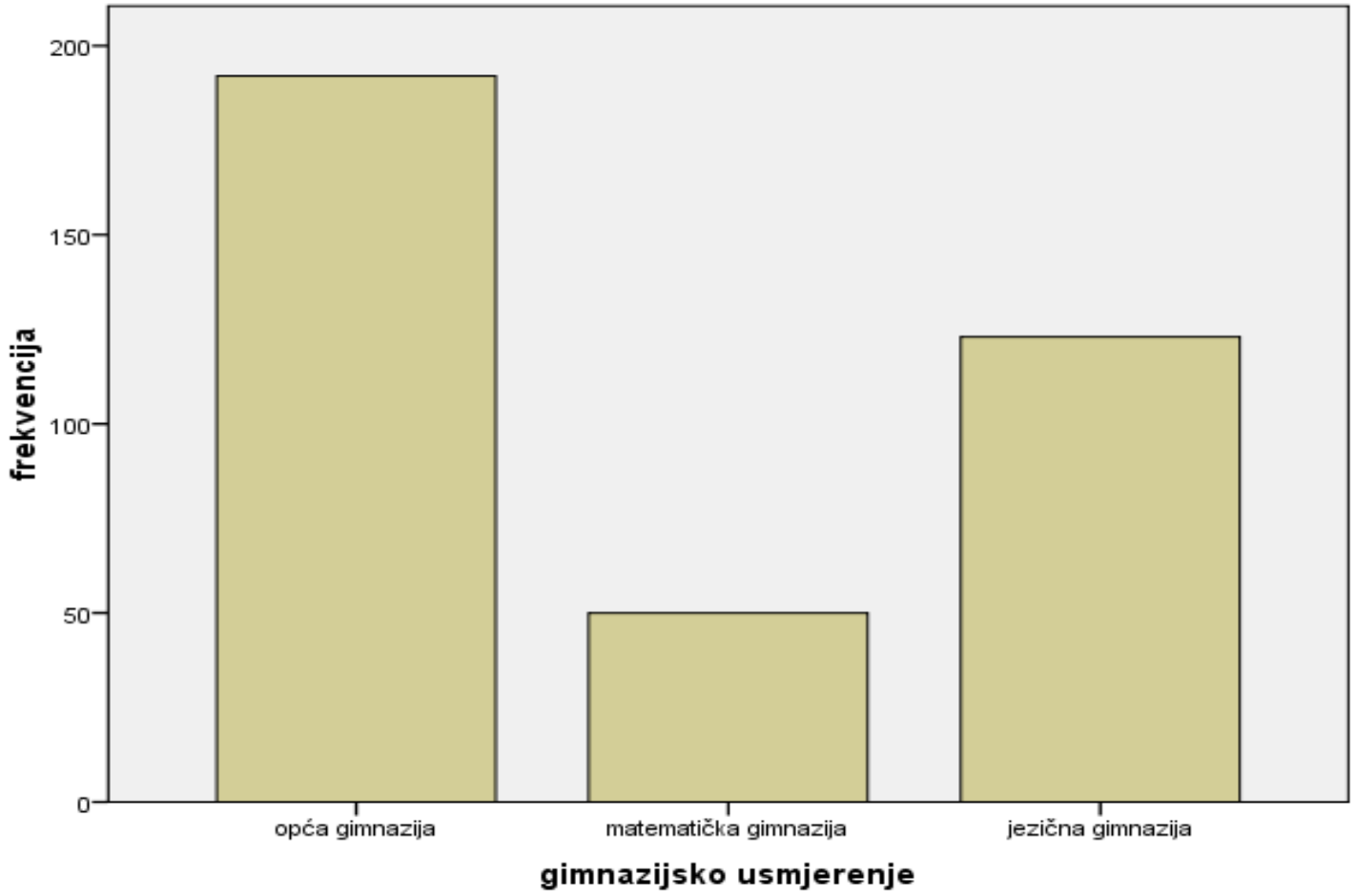
- na horizontalnu os – x (apscisu) unose se *vrijednosti* rezultata (ili sredine, odnosno gornje granice razreda)
- na vertikalnu os – y (ordinatu) unose se *frekvencije*
- **omjer** vertikalne i horizontalne osi otprilike treba iznositi ***2/3*** (y=2/3x)
- u ishodištu se nalazi nula
- apscisa i/ili ordinata (udaljenost od nule) *presijecaju* se kada je raspon rezultata velik
- potrebno je voditi računa o kompoziciji grafikona, tj. paziti da on bude na *sredini slike*
- X i Y osi trebale bi biti pravilno označene i imati *ekvidistantne jedinice*
- svaki grafički prikaz mora biti numeriran i imati jednostavan i **jasan naslov** koji se prikazuje ispod slike
- ako je potrebno, grafički prikaz sadrži i legendu koja se obično stavlja u gornji desni kvadrant slike

Horizontalni stupci – prikaz podataka na nominalnoj ili ordinalnoj skali



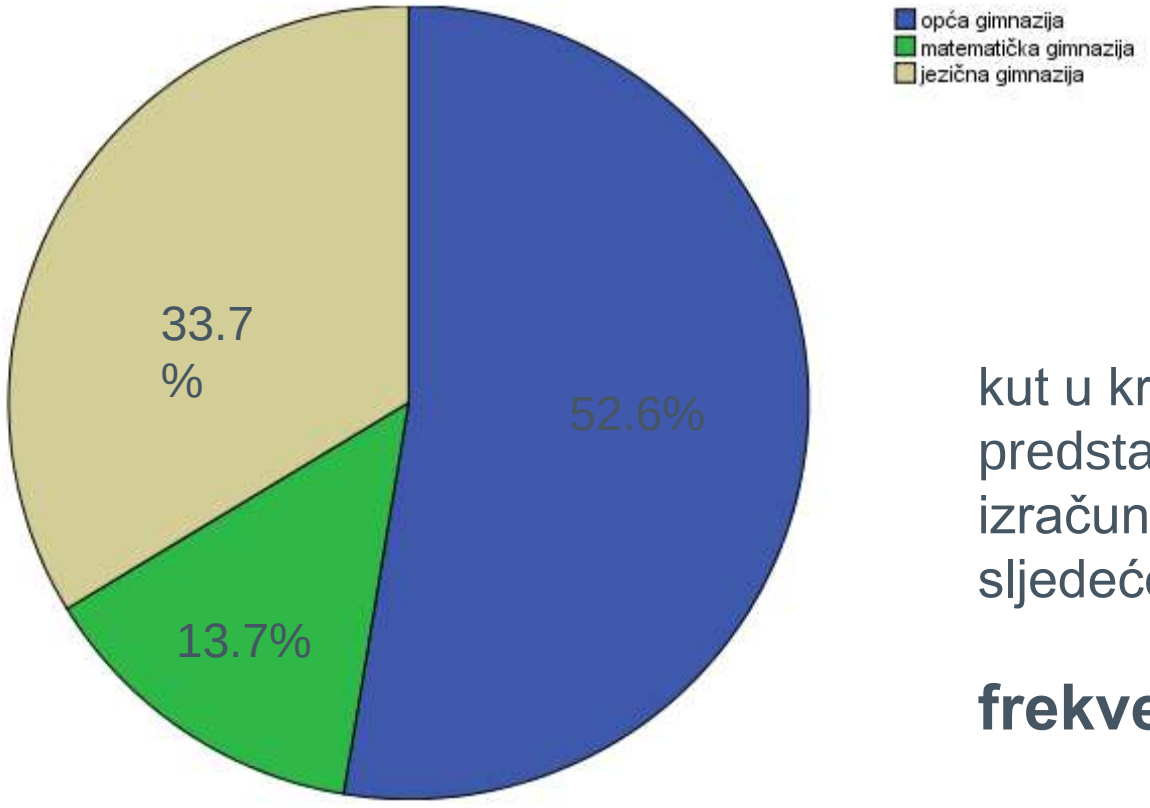
Slika 3. Grafički prikaz stupnja obrazovanja očeva šibenskih gimnazijalaca školske godine 2008./2009. (N=364) (Burić, 2010)

Vertikalni stupci – prikaz podataka na nominalnoj ili ordinalnoj skali



Slika 4. Grafički prikaz vrste usmjerenja šibenskih gimnazijalaca u školskoj godini 2008./2009. (N=365) (Burić, 2010)

Kružni dijagram – prikaz podataka izraženih na nominalnoj ili ordinalnoj skali



kut u kružnom dijagramu predstavlja postotak, a izračunava se prema sljedećoj formuli:

frekvencija·360/N

Slika 5. Kružni dijagram vrste gimnazijskog usmjerenja šibenskih gimnazijalaca u školskoj godini 2008./2009. (N=365) (Burić, 2010)

4. Struktura rezultata psihologijskog mjerenja i model normalne distribucije

Struktura rezultata psihologijskog mjerenja

- pretpostavka mjerenja:
  - postoji **prava vrijednost mjerenja** (tj. fenomen, proces ili osobina) koja je *latentne* prirode i koju je moguće zahvatiti određenom **operacionalnom definicijom, tj. manifestnim indikatorima**
  - primjer: *vrijeme reakcije* (predmet mjerenja) operacionalno definirano kao *vrijeme izraženo u ms koje protekne od trenutka zadavanja podražaja do trenutka ispitanikove reakcije* (operacionalizacija VR)

Razmislite...

- kako bismo izmjerili nečije *pravo* VR na svjetlosni podražaj?
- kako bismo izmjerili *pravu* visinu kvocijenta inteligencije studenata FER-a?
- da bismo mogli odrediti (*približno*) pravu vrijednost neke ispitivane pojave (npr. VR) za određenog ispitanika, potrebno je provesti *veći broj* mjerenja (npr. 50)
- da bismo odredili koja je (*približno*) prava vrijednost neke pojave (npr. inteligencije) u određenoj populaciji, potrebno je obaviti mjerenja na *većem broju* pripadnika te populacije (npr. N=500)
- zašto?

Struktura rezultata psihologijskog mjerenja

- zato što na svaki rezultat mjerenja djeluju:
  - **konstantni faktori** koji se odnose na sam *predmet mjerenja* (npr. pojedinčeva inteligencija)
  - **nesistematski varijabilni faktori** ili svi faktori koji pri mjerenju neke pojave utječu na veličinu dobivene vrijednosti, a njihovo djelovanje je *slučajno* = POGREŠKA
- svaki rezultat mjerenja (većine ispitivanih pojava ili osobina u psihologiji) ima sljedeću strukturu:

$$X = Xp + Xs$$

X = izmjerena vrijednost

Xp = prava vrijednosti mjerenja

Xs = nesistematski varijabilni faktori (pogreška)

Konstantni faktori

- odnose se na predmet mjerenja, tj. **pravu vrijednost mjerenja** (npr. inteligencija, VR, neuroticizam)
- nazivaju se **konstantnim** jer se pretpostavlja da je ono što se mjeri ili opaža relativno konstantno (barem za vrijeme mjerenja – u suprotnom, mjerenje ne bi imalo smisla!)
- dakle, *prava vrijednost mjerenja je konstantna*, tj. uvijek ista bez obzira na broj mjerenja (npr. prava vrijednost VR-a nekog ispitanika ista je u svih 50 pokušaja mjerenja ili prosječan IQ populacije je uvijek 100)

Pogreška

- odgovorna je za diskrepancu između vrijednosti koje smo mi dobili mjerenjem i prave vrijednosti mjerenja
- Primjer 1: Vaša prava težina iznosi 65 kg. Vaša vaga pokazuje da imate 62 kg. Koliko iznosi pogreška?
- Primjer 2: Prosječna visina odraslih muškaraca u Hrvatskoj iznosi 182 cm. Prosječna visina izračunata na uzorku od 200 muškaraca iz različitih dijelova Hrvatske iznosi 183.5 cm. Koliko iznosi pogreška?
- Pogreška iz primjera 1 zove se **POGREŠKA MJERENJA**, a pogreška iz primjera 2 zove se **POGREŠKA UZORKOVANJA**
- obje pogreške odraz su djelovanja ***slučajnih*** faktora

Pogreška = slučajni ili nesistematski varijabilni faktori (NSVF)

- ***po slučaju*** i međusobno nezavisno djeluju na veličinu mjerene pojave, skrećući mjereni rezultat čas na jednu, čas na drugu stranu
- iako znamo da postoje i da djeluju na rezultate, ne mogu se u potpunosti kontrolirati ni eliminirati
- odgovorni su za **raspršenje** (tj. međusobno razlikovanje) rezultata – što je pri mjerenju neke pojave više NSVF-a, to je veće raspršenje dobivenih rezultata, odnosno veća je pogreška
- što je pri nekom mjerenju moguće kontrolirati više tih faktora, rezultati mjerenja imaju manje raspršenje, tj. mjerenje je *točnije* (manja je pogreška) i manja je vjerojatnost odstupanja od prave vrijednosti

Pogreška mjerenja

- izvori NSVF-a:
  - nesistematske promjene u ispitaniku (ispitanicima) (npr. fluktuacije u pažnji i koncentraciji unutar jednog ispitanika)
  - faktori iz neposredne okoline u kojoj se odvija mjerenje ili opažanje, uključujući i nesistematske faktore eksperimentalne ili istraživačke procedure (npr. buka, razlike u osvjetljenju, doba dana u kojem se obavlja mjerenje, obilježja prostora itd.)
  - sam postupak mjerenja ili opažanja – nesustavne promjene u sredstvu mjerenja ili u mjeritelju (npr. fluktuacije u pažnji eksperimentatora)

Pogreška uzorkovanja

- izvori NSVF-a
  - djelovanje slučaja na slučaj prilikom odabira uzorka ispitanika (npr. ako u dva navrata po slučaju odabiremo 50 ispitanika iz neke populacije, realno je očekivati da će u drugom pokušaju biti odabrani (i) neki novi ispitanici)

- kod određivanja prave vrijednosti mjerenja na skupini ispitanika uglavnom su prisutne obje vrste pogreške – i pogreška mjerenja i pogreška uzorkovanja!

Veličina pogreške mjerenja

- Kako smanjiti pogrešku mjerenja?
- kroz osiguravanje adekvatnih **psihometrijskih karakteristika** mjernih instrumenata kojima se koristimo prilikom mjerenja različitih osobina/fenomena
  - VALJANOST – sposobnost mjernog instrumenta da mjeri ono što bi trebao mjeriti (konstruktna, sadržajna, kriterijska)
  - POUZDANOST – sposobnost mjernog instrumenta da u ponovljenim mjerenjima daje iste ili slične rezultate
- pouzdanost je nužan preduvjet valjanosti; međutim, pouzdanost sama po sebi ne jamči i valjanost!

Razmislite…

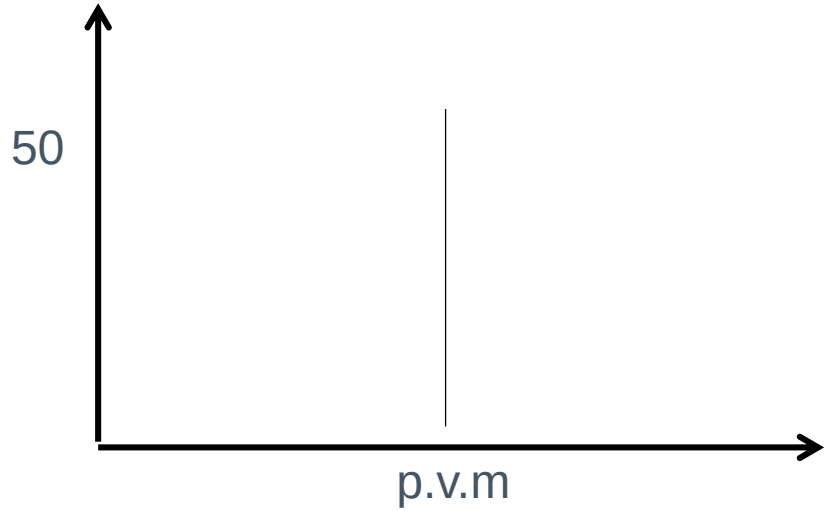
- Koliko je VR na svjetlosni podražaj mjereno pomoću specijaliziranog softvera *valjana* mjera nečije brzine reagiranja?
- Koliko je veličina EDR-a (elektrodermalna reakcija) *valjana* mjera ispitne anksioznosti?
- Koliko je rezultat na testu B-serije (verbalni test) *valjana* mjera opće inteligencije?

Pitanja:

1. Kada bi imali savršeno mjerenje u kojem bismo u potpunosti eliminirali djelovanje pogreške te bi na izmjereni rezultat djelovali samo konstantni faktori, koliko bi nam mjerenja bilo potrebno da utvrdimo pravu vrijednost mjerenja?
2. Kakvo, tj. koliko bi bilo raspršenje rezultata?
3. Kako bi izgledala distribucija rezultata dobivenih iz 50 pokušaja mjerenja (x – izmjerene vrijednosti, y – frekvencija)?

Odgovori:

1. dovoljno bi bilo samo jedno mjerenje
2. raspršenja ne bi bilo (nula)
3. distribucija bi izgledala ovako:

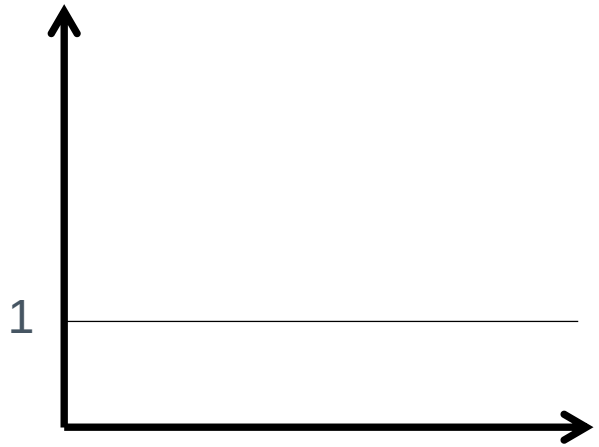


Još nekoliko pitanja:

1. Kada bi na rezultate mjerenja djelovali samo NSVF-i, tj. kada bi pri nekom mjerenju svaki put dobili drugi rezultat, kako bi izgledala distribucija?
2. Ako NSVF-i djeluju po zakonu slučaja, tj. skrećući mjereni rezultat čas na jednu, čas na drugu stranu u odnosu na pravu vrijednost mjerenja, koliko iznosi prosječna vrijednost odstupanja (tj. *razlika*) dobivenih rezultata od prave vrijednosti mjerenja (ako imamo dovoljno velik broj mjerenja)?

Odgovori:

1. distribucija bi izgledala ovako:

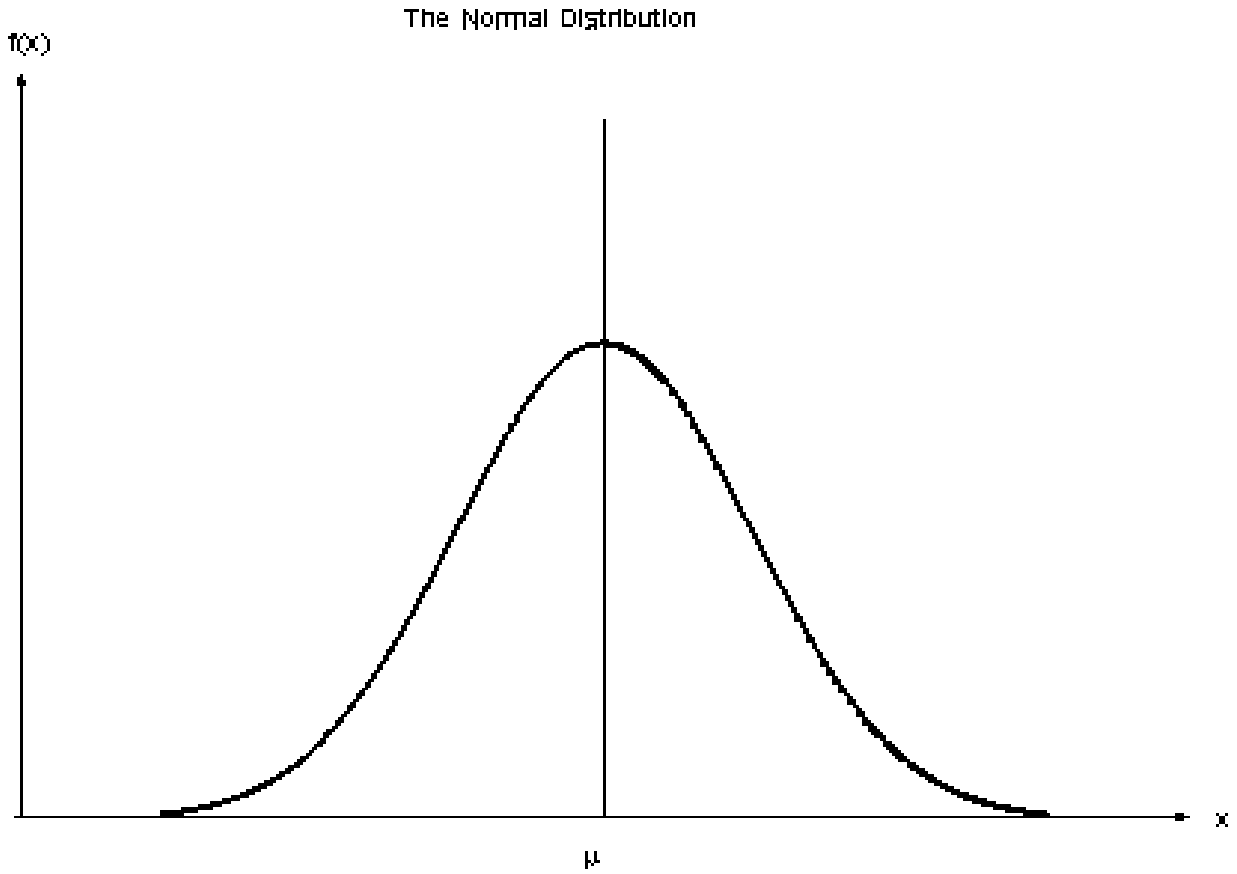


2. 0 (nula)!
   - budući da djeluju po slučaju, ako imamo dovoljno velik broj mjerenja, oni se *međusobno poništavaju*, pa je prosječna vrijednost odstupanja od prave vrijednosti mjerenja jednaka **nuli**
   - tako **X koji je jednak Xp + Xs** uz dovoljno velik broj mjerenja postaje **samo Xp**

I posljednje pitanje...

- A kako izgleda distribucija skupine rezultata ako na njih djeluju i konstantni i nesistematski varijabilni faktori, što je ujedno i najčešći scenarij psihologijskih mjerenja (x=izmjerene vrijednosti, y=frekvencija)?

Prikaz normalne distribucije



Preuzeto s: http://www.itl.nist.gov/div898/handbook/pmc/section5/pmc51.htm

Normalna distribucija

- dobila ime po astronomu Karlu Friedrichu Gaussu, iako ju je „otkrio” matematičar Abraham de Moivre
- naziva se još i zvonolika ili Gaussova krivulja
- tendencija **grupiranja rezultata** oko jedne **središnje vrijednosti**
- kod mjerenja različitih pojava ili osobina dobivamo najviše srednjih, prosječnih rezultata, a sve manje onih koji od tog prosjeka sve više odstupaju
- do toga dolazi zbog:
  - djelovanja konstantnih i nesistematskih varijabilnih faktora prilikom mjerenja neke pojave na jednom pojedincu
  - jer se neke osobine u populaciji normalno distribuiraju (npr. inteligencija)

Normalna distribucija

- simetrična
- zvonolika
- unimodalna
- granice: ± ∞
- apscisa predstavlja različite moguće vrijednosti X, a ordinata frekvenciju ili vjerojatnost pojavljivanja X

Uvjeti za dobivanje normalne distribucije – mjerenje neke osobine na jednom ispitaniku

- jasno i precizno definiran predmet mjerenja
- svaki rezultat ovisi o djelovanju konstantnih i nesistematskih varijabilnih faktora
- skala na kojoj se izražavaju rezultati mora omogućiti variranje rezultata na obje strane od prosječne vrijednosti
- velik broj mjerenja

Uvjeti za dobivanje normalne distribucije – mjerenje neke osobine na skupini ispitanika

1. ono što mjerimo, mora se i u prirodi distribuirati po normalnoj raspodjeli (npr. izuzetak je bilirubin u krvi)
2. skupina koju mjerimo mora biti heterogena (neselekcionirana) po osobini koju mjerimo, ali mora biti homogena po svim ostalim svojstvima koja mogu djelovati na ono što mjerimo
3. članovi uzorka moraju biti izabrani po slučaju
4. uzorak mora biti dovoljno velik
5. sva mjerenja trebamo provesti u jednakim vanjskim uvjetima

Zašto je normalna distribucija važna?

1. mnoge pojave u psihologiji se normalno distribuiraju i u populaciji (npr. visina) – kad bismo izmjerili neku pojavu kod svih članova populacije, dobili bismo normalnu distribuciju
2. ako se neka varijabla normalno distribuira, možemo koristiti različite statističke postupke koji nam omogućuju donošenje zaključaka (preciznih ili vjerojatnih) o vrijednostima te varijable u populaciji
3. većina (korisnih) statističkih postupaka počiva na pretpostavci o normalnoj distribuciji

Zašto je normalna distribucija važna?

4. normalna distribucija je temelj za razumijevanje glavnih *statističkih pojmova vjerojatnosti*
   - ukupna površina normalne distribucije iznosi 1.0 ili 100 % pa je moguće očitati sve moguće veličine površina između nekog rezultata i bilo kojeg drugog rezultata u toj distribuciji, odnosno odrediti *vjerojatnost pojavljivanja* bilo kojeg (raspona) rezultata
   - z – vrijednosti i tablice normalne distribucije
5. normalna distribucija je temelj za razumijevanje *inferencijalne* statistike!

Važno!

- Iskustva iz nastave statistike redovno potvrđuju da netko tko ne razumije smisao normalne raspodjele (i sve ono što iz toga proizlazi) ne može razumjeti ni mnoge druge statističke principe, a osobito ne one koji se izravno logički izvode iz normalne raspodjele.

Normalna distribucija

- normalna distribucija je matematička ili teorijska distribucija što znači da se raspodjela dobivenih rezultata više ili manje razlikuje od „savršene” normalne distribucije (iako joj nalikuje)
- odstupanja od normalne distribucije najčešće se određuju pomoću 2 parametra:
  - **asimetričnost** (engl. *skewness*)
  - **kurtičnost (spljoštenost)** (engl. *kurtosis*)

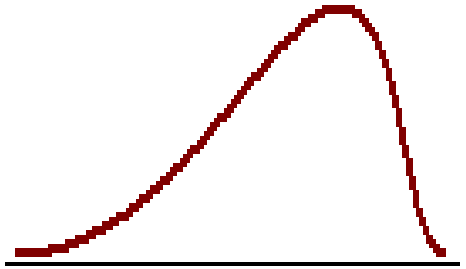
Asimetričnost

- najčešće odstupanje od normalne distribucije
- glavnina rezultata grupira se više prema lijevoj ili desnoj strani raspona dobivenih rezultata
- često se javlja zbog nemogućnosti variranja rezultata u jednu stranu
- npr. na skupini učenika 7. razreda nadarenih u matematici primijeni se test znanja koji se uobičajeno koristi za sve učenike sedmih razreda
- s obzirom da se može očekivati da će nadareni učenici znati riješiti većinu zadataka, njihovi rezultati će se grupirati oko viših, tj. maksimalnih vrijednosti
- posljedica je djelovanja SISTEMATSKIH FAKTORA (npr. umor, uvježbavanje, selekcionirana skupina ispitanika)

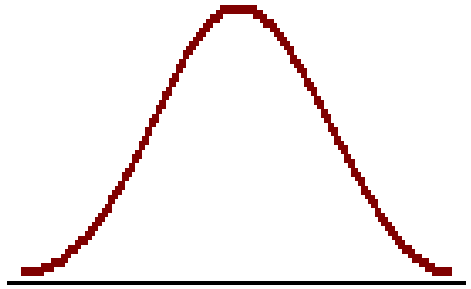
Asimetričnost

- distribucija može biti:
  - **pozitivno asimetrična**
    - većina rezultata koncentrira se na manjim vrijednostima
  - **negativno asimetrična**
    - većina rezultata koncentrira se na većim vrijednostima

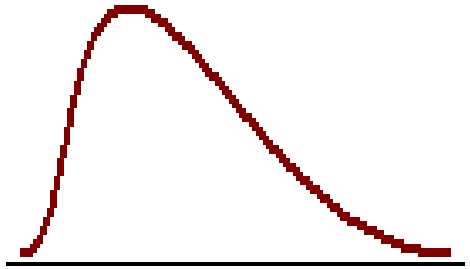
Asimetričnost: *negativno asimetrična*, *normalna* i *pozitivno asimetrična* distribucija



Negatively skewed distribution
or Skewed to the left
Skewness <0



Normal distribution
Symmetrical
Skewness = 0



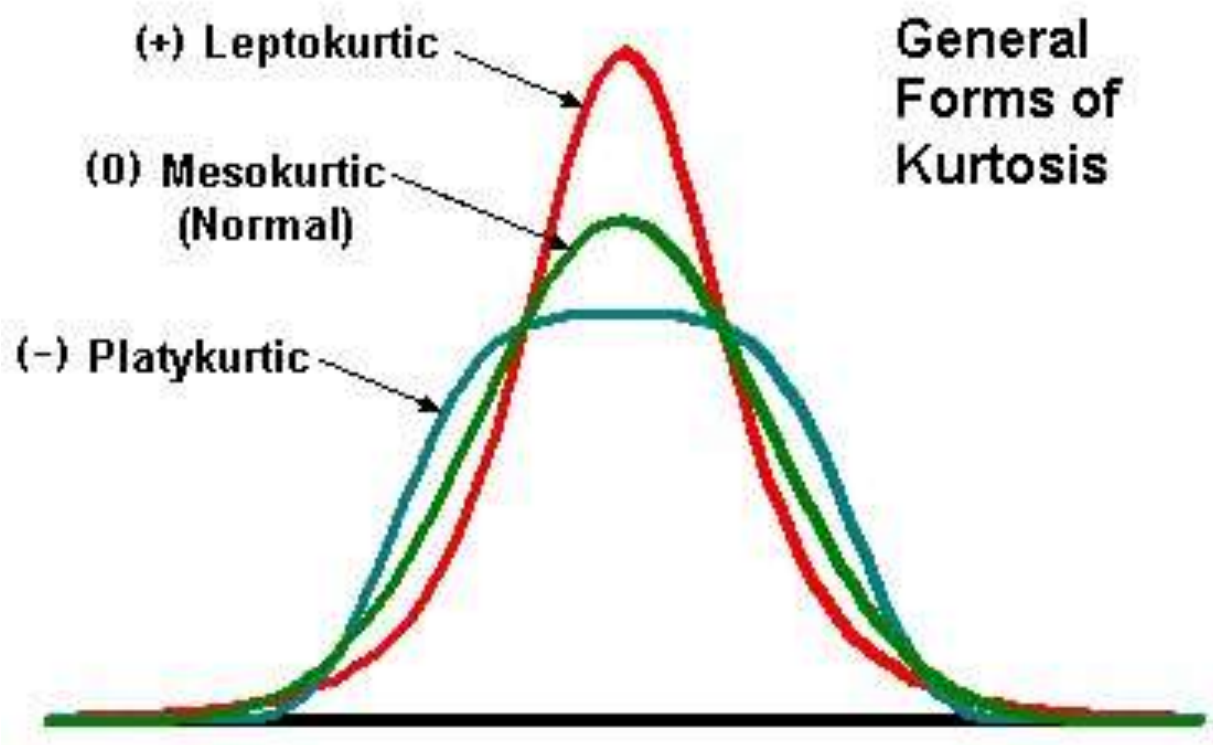
Positively skewed distribution
or Skewed to the right
Skewness > 0

https://tekmarathon.com/2015/11/13/importance-of-data-distribution-in-training-machine-learning-models/

Kurtičnost

- odnosi se na konveksnost krivulje normalne distribucije, tj. njezin zvonolik dio, odnosno na koncentraciju rezultata oko aritmetičke sredine
- 3 razine kurtičnosti:
  - **mezokurtičnost** – kod normalne distribucije
  - **leptokurtičnost** – kod zvonaste krivulje kod koje se rezultati izrazitije grupiraju oko središnje vrijednosti, odnosno raspršenje je manje, zbog čega krivulja izgleda “visoka i vitka”
  - **platikurtičnost** – kod zvonaste krivulje koja se očituje u slabijem grupiranju rezultata oko središnje vrijednosti, pa krivulja izgleda „spljoštena i široka”

Kurtičnost: *leptokurtična*, *mezokurtična* (normalna) i *platikurtična* distribucija



https://mvpprograms.com/help/mvpstats/distributions/SkewnessKurtosis

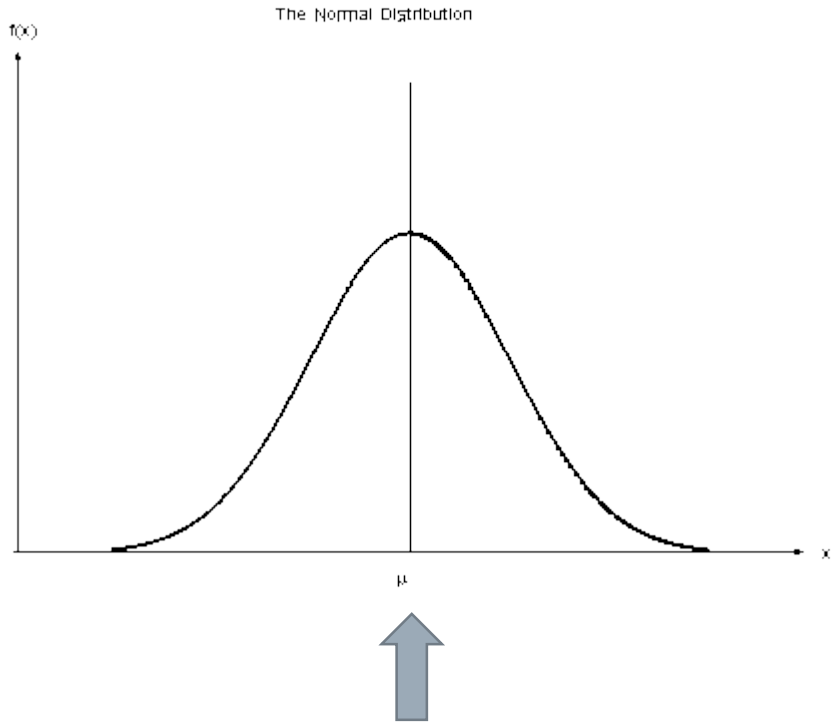
- odstupa li neka distribucija značajno od normalne (bilo u pogledu asimetričnosti ili u pogledu kurtičnosti), najjednostavnije se određuje pomoću **Kolmogorov-Smirnovljevog testa**
- ako se utvrdi da distribucija *statistički značajno* odstupa od normalne, moguće je provesti matematičke **transformacije podataka** (npr. vađenjem drugog korijena iz svakog rezultata) koje će dovesti do distribucije koja je bliža normalnoj (ili koristiti robustnije algoritme za procjenu parametara u složenijim statističkim postupcima)

5. Središnje vrijednosti

Što su mjere centralne tendencije?

- središnje vrijednosti
- brojčane vrijednosti koje „*reprezentiraju*” skupinu rezultata u slučajevima kad rezultati imaju tendenciju grupiranja oko neke središnje (ili neke druge!) vrijednosti

Normalna distribucija



Rezultati koji se normalno distribuiraju međusobno se razlikuju, ali imaju *tendenciju grupiranja oko središnje vrijednosti*.

određivanje mjera centralne tendencije → procjenjivanje ***prave vrijednosti mjerenja***

Mjere centralne tendencije

- Aritmetička sredina
- Centralna vrijednost ili medijan
- Dominantna vrijednost ili mod

Aritmetička sredina – M

› određuje se tako da se sve vrijednosti u nekom skupu rezultata zbroje i taj zbroj podijeli ukupnim brojem rezultata

$$M = \frac{\sum X}{N}$$

M – engl. *mean* – aritmetička sredina
∑ - suma (zbroj)
X – svaki pojedini rezultat
N – ukupan broj rezultata

Aritmetička sredina

- ***težište rezultata***
- Na njezinu vrijednost djeluje ***težina,*** tj. vrijednost svakog pojedinog rezultata, koja se očituje u *veličini odstupanja od M.*
- Ako je distribucija rezultata savršeno *normalna*, koliko će iznositi zbroj odstupanja svih rezultata od M, bilo s lijeve ili desne strane ($\sum(X - M)=?$)?
- 0
- Koliko će iznositi ova vrijednost ako distribucija nije normalna?
- Isto!

Aritmetička sredina

- algebarski zbroj odstupanja pojedinačnih rezultata od aritmetičke sredine jednak je nuli, a zbroj kvadrata tih odstupanja ***manji*** je od zbroja odstupanja od bilo koje druge vrijednosti u određenom skupu rezultata (bez obzira na oblik njihove distribucije)
- $\sum(X_i - M)=0$
- $\sum(X_i - M)^2 < \sum(Xi - Y)^2$ (Y – bilo koja druga vrijednost u skupu rezultata)

Aritmetička sredina

- upravo zbog toga što predstavlja težište rezultata, aritmetičku sredinu koja ima smisao najbolje reprezentativne vrijednosti nije opravdano računati ako distribucija rezultata nije barem *približno normalna* ili ako su prisutne *ekstremne vrijednosti* rezultata

Razmislite...

- Koja vrijednost najbolje reprezentira sljedeći niz rezultata: 2, 3, 4, 4, 5, 6, 18?
- Koliko iznosi M ovog niza rezultata?

- zbog prisutnosti ekstremne vrijednosti rezultata (x=18), aritmetička sredina iznosi 6, iako vrijednost 4 bolje reprezentira većinu rezultata

Centralna vrijednost ili medijan – C

- vrijednost koja se u nizu rezultata, poredanih po veličini, nalazi točno u sredini – broj rezultata većih i manjih od medijana je jednak
- položaj centralne vrijednosti, tj. **redno mjesto** se u nizu rezultata poredanih po (numeričkoj) veličini određuje kao:

$$R_C = \frac{N}{2} + 0.5$$ ili $$R_C = \frac{N+1}{2}$$

Rc – redno mjesto medijana
N – broj rezultata

Medijan

- ako je broj rezultata *neparan*, C je jedna ***stvarno dobivena*** vrijednost
  - Npr. 2, 3, 3, 4, **4**, 4, 5, 5, 6
  - N=9, Rc=9/2+0.5=**5**
  - 5. vrijednost u nizu iznosi ***C=4***

Medijan

- ako je broj rezultata *paran*, C je ***izvedena*** vrijednost – dvije susjedne vrijednosti koje se nalaze u sredini niza, zbroje se i podijele s 2
  - Npr. 2, 3, 4, 5, 6, 7
  - N=6, Rc=6/2+0.5=3.5
  - C je vrijednost koja se nalazi između 3. i 4. mjesta u nizu rezultata: C=(4+5)/2=**4.5**

Rc≠C!

Medijan

- na C ne utječe vrijednost pojedinog rezultata (kao što je slučaj kod aritmetičke sredine), već samo *broj* rezultata
- zato je C osobito pogodna kao mjera centralne tendencije za skupinu rezultata koja sadrži jednu ili dvije ekstremne vrijednosti

Dominantna vrijednost ili mod – D

- D je vrijednost koja među rezultatima dominira ***čestinom*** pojavljivanja
- određena je samo čestinom rezultata, a ne i njihovim ukupnim brojem ili pojedinačnim vrijednostima

Mod

- koristi se u situacijama kada:
  - distribucija značajno odstupa od normalne
  - je mjerenje izraženo na nominalnoj ili ordinalnoj skali (npr. stupanj obrazovanja zaposlenika neke tvrtke)
  - kad drugi načini određivanja središnje vrijednosti daju besmislene rezultate (npr. prosječan broj djece u hrvatskoj obitelji)
- upotrebljava se samo onda kad je *broj rezultata dovoljno velik* i kad samo jedna vrijednost dominira u nizu mjerenja

Rezime:

- M (aritmetičku sredinu) računamo:
  - kad su rezultati izraženi na intervalnoj i omjernoj ljestvici
  - kad se rezultati distribuiraju normalno
- C (medijan) određujemo:
  - kad je distribucija rezultata izrazito asimetrična (tj. kad značajno odstupa od normalne)
  - kad su među rezultatima prisutne ekstremne vrijednosti (ali u samo jednom smjeru!)
  - kad su rezultati izraženi na ordinalnoj, intervalnoj i omjernoj skali
- D (mod) određujemo:
  - kad druge mjere centralne tendencije daju besmislene rezultate
  - kad su rezultati izraženi na nominalnoj, ordinalnoj, intervalnoj i omjernoj skali

6. Mjere varijabilnosti

Raspršenje rezultata

- pri ponavljanom mjerenju iste pojave (npr. VR na jednom ispitaniku) ili jednokratnom mjerenju neke karakteristike na velikom broju slučajeva (npr. mjerenje visine dvogodišnje djece), dobivaju se rezultati koji su *međusobno različiti*, ali i slični jer se grupiraju oko neke središnje vrijednosti
- međusobno razlikovanje rezultata → ***raspršenje*** rezultata
- rezultat djelovanja *slučajnih* faktora

Raspršenje rezultata

- reprezentativnost središnje vrijednosti ovisi o raspršenju rezultata:
  - ako se vrijednosti dobivene mjerenjem gusto grupiraju oko jedne središnje vrijednosti, tj. ako je raspršenje MALO, ta središnja vrijednost dobro predstavlja rezultate
  - ako se vrijednosti dobivene mjerenjem međusobno jako razlikuju te ne pokazuju nikakvo grupiranje oko jedne središnje vrijednosti, odnosno raspršenje je VELIKO, takva središnja vrijednost je nikakav ili loš reprezentant rezultata

Raspršenje rezultata

- višekratno mjerenje jedne osobine na jednom ispitaniku
  - pod pretpostavkom da mjerena osobina predstavlja neku relativno konstantnu veličinu, raspršenje rezultata oko središnje vrijednosti ukazuje na ***točnost kojom je mjerenje izvršeno***
  - što je bolje kontrolirano djelovanje NSVF-a, rezultati mjerenja manje će se razlikovati, tj. raspršenje će biti manje
  - indeks raspršenja rezultata oko središnje vrijednosti; ukazuje na veličinu **POGREŠKE MJERENJA**

Raspršenje rezultata

- jednokratno mjerenje jedne osobine na skupini ispitanika
  - raspršenje rezultata ukazuje na ***reprezentativnost* središnje vrijednosti**
  - što je raspršenje rezultata veće, središnja vrijednost slabije predstavlja skupinu ispitanika kao cjelinu – veliko raspršenje pokazuje da središnja vrijednost odgovara relativno malom broju individualnih rezultata
  - indeks raspršenja rezultata koje ukazuje u kojoj mjeri ispitivana osobina varira među različitim pojedincima ukazuje na veličinu VARIJABILITETA ispitivane pojave u populaciji

Raspršenja rezultata

- postoje različite mjere, tj. INDEKSI RASPRŠENJA koje više ili manje uspješno ukazuju na veličinu raspršenja rezultata oko središnje vrijednosti
  - RASPON
  - STANDARDNA DEVIJACIJA & VARIJANCA
  - KOEFICIJENT VARIJABILNOSTI
  - POLUINTERKVARTILNO RASPRŠENJE

Raspon

- razlika između najvećeg i najmanjeg rezultata u nekom mjerenju
- najjednostavnija, ali i *najmanje točna* mjera raspršenja
- Zašto?
  - ako je pri mjerenju slučajno dobiven ekstremni rezultat, vrijednost totalnog raspona povećava se, iako se možda svi ostali rezultati homogeno grupiraju oko neke središnje vrijednosti
  - ekstremni rezultati koji određuju raspon najmanje su pouzdani, tj. promjenjivi su od skupa do skupa rezultata
  - ovisi o broju rezultata u skupu – što je uzorak mjerenja veći, raspon ima tendenciju porasta (jer se povećava vjerojatnost pojavljivanja ekstremnih rezultata)

Raspon

- upotreba totalnog raspona:
  - kao grubi pokazatelj varijabilnosti rezultata mjerenja
  - za utvrđivanje krajnjih granica unutar kojih variraju rezultati

  Primjer A. 5, 7, 7, 8, 8, 8, 9, 9, 10
  Primjer B. 5, 7, 7, 8, 8, 8, 9, 9, 20

  Koliko iznosi raspon rezultata u ova dva primjera?

- Rješenja:
  - primjer A. raspon=10-5=5
  - primjer B. raspon=20-5=15

Standardna devijacija

- najbolja mjera raspršenja rezultata
- statistička mjera koja pokazuje kako se „gusto” rezultati nekog mjerenja grupiraju oko aritmetičke sredine, tj. koliko se rezultati nekog mjerenja razlikuju od aritmetičke sredine kao referenične točke
- mjera prosječnog odstupanja rezultata od aritmetičke sredine
- koristi se jedino uz aritmetičku sredinu i ima svoj smisao samo ako je *distribucija rezultata barem približno normalna*
- Ako se koristi kao mjera raspršenja rezultata na uzorku, označava se sa **SD**, a ako se koristi kao mjera raspršenja rezultata u populaciji, označava se sa **σ**

Računanje standardne devijacije

- zasniva se na kvadratima razlika između aritmetičke sredine i pojedinačnih rezultata
- odstupanja svakog pojedinog rezultata od aritmetičke sredine se *kvadriraju* $(X_i - M) \rightarrow (X_i - M)^2$ iz dva razloga:
  - kako bi izbjegli predznake odstupanja
  - kako bi (većim) razlikama dali veću težinu
- zbroj kvadrata tih odstupanja $\sum(X_i - M)^2$ manji je od zbroja kvadrata odstupanja od bilo koje druge vrijednosti u određenom skupu rezultata (bez obzira na oblik distribucije rezultata)

Varijanca - σ^2

- ako sva odstupanja pojedinih rezultata od aritmetičke sredine kvadriramo, zbrojimo i podijelimo s brojem rezultata, dobivamo VARIJANCU
- dakle, varijanca je aritmetička sredina kvadriranih razlika između svakog pojedinog rezultata i aritmetičke sredine, tj. njihova prosječna vrijednost
- izravni pokazatelj varijabiliteta rezultata

Varijanca

› varijanca se određuje prema formuli:

$$\sigma^2 = \frac{\sum \left(X - M \right)^2}{N}$$

X=pojedini rezultat
M=aritmetička sredina
N=broj rezultata

Primjena varijance

- varijanca se primjenjuje pri nekim složenijim postupcima statističke analize (npr. faktorska analiza), ali ne i kao uobičajena mjera raspršenja nekog skupa rezultata (u deskriptivnoj statistici)
- ali, ako izvadimo *drugi korijen iz varijance*, dobivamo STANDARDNU DEVIJACIJU koja predstavlja standard za mjerenje varijabiliteta rezultata

Računanje standardne devijacije

$$\sigma = \sqrt{\frac{\sum (X - M)^2}{N}}$$

ili

$$SD = \sqrt{\frac{\sum (X - M)^2}{N - 1}}$$

Računanje standardne devijacije

- formula s N – 1 u nazivniku (SD)
  - koristi se u *inferencijalnoj statistici,* tj. kad na temelju karakteristika utvrđenih na uzorku želimo zaključiti o karakteristikama populacije
  - važnost **stupnjeva slobode**! (vrijednost prave M populacije nam oduzima 1 stupanj slobode)
- formula s N u nazivniku (σ)
  - koristi se u *deskriptivnoj statistici,* tj. kad nas zanimaju samo karakteristike neke skupine podataka
  - kada računamo standardnu devijaciju *populacije* (što je u praksi rijetko)
  - u inferencijalnoj statistici, ali samo ako imamo *velik broj* rezultata

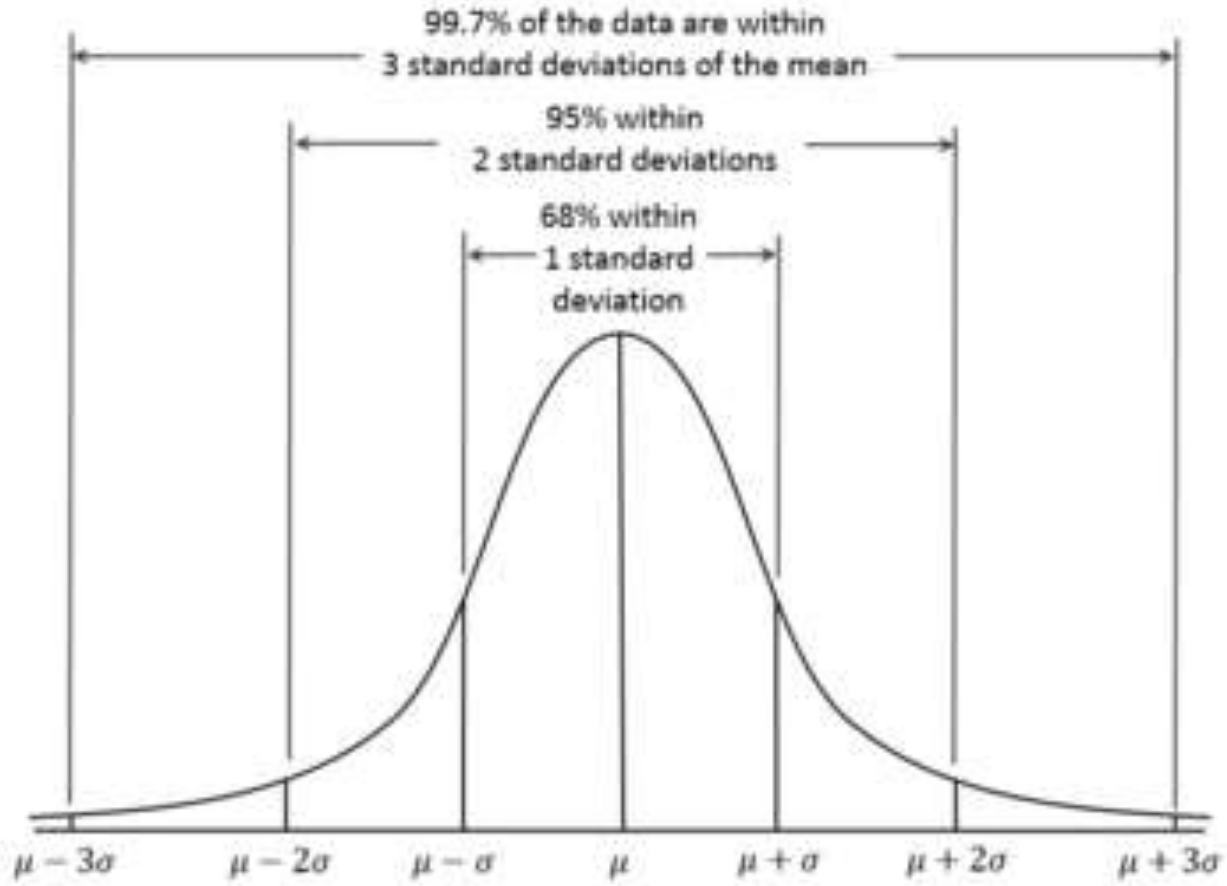
Prednosti standardne devijacije

- smanjivanje ili povećavanje veličine uzorka ne djeluje sistematski na standardnu devijaciju, već ona – uz manja ili veća odstupanja – ostaje približno jednaka
- standardna devijacija, uz aritmetičku sredinu, potpuno *definira normalnu distribuciju* – ako su nam poznate vrijednosti M i SD, možemo točno nacrtati krivulju normalne distribucije jer sve krivulje normalne distribucije imaju jednake karakteristike (npr. mjesto na kojem se krivulja iz konkavne pretvara u konveksnu („točka infleksije”) nalazi se na svim (normalnim) krivuljama točno iznad 1 SD)

Normalna distribucija i razumijevanje vjerojatnosti

- ako se rezultati grupiraju normalno te ako su nam poznate vrijednosti M i SD, možemo odrediti *vjerojatnost* pojavljivanja nekog (raspona) rezultata
- vjerojatnost (*p*) – odnos između ostvarenih slučajeva prema broju mogućih
- što je manja frekvencija rezultata određene vrijednosti u skupu, manja je i vjerojatnost pojavljivanja takvog rezultata

Normalna distribucija



U normalnoj distribuciji, 68 % rezultata nekog mjerenja nalazi se u intervalu M±1SD, 95 % rezultata u intervalu M±2SD, a 99.7 % rezultata u intervalu M±3SD.

Preuzeto s: https://en.wikipedia.org/wiki/Normal_distribution

Koeficijent varijabilnosti – V

- kad su nam poznati M i SD skupa rezultata, moguće ih je uspoređivati s nekim drugim rezultatima – npr. između nekoliko skupova rezultata utvrditi kod kojeg skupa aritmetička sredina najbolje predstavlja rezultate
- Primjer. U mjerenju A utvrđene su: M=100, SD=10; u mjerenju B utvrđene su: M=8, SD=2
  - U kojem mjerenju aritmetička sredina bolje reprezentira rezultate?

Koeficijent varijabilnosti

- odgovor na prethodno pitanje omogućuje nam izračun ***koeficijenta varijabilnosti (V)*** koji pokazuje koliki postotak vrijednosti aritmetičke sredine iznosi vrijednost standardne devijacije

$$V = \frac{SD \cdot 100}{M}$$

- korisna mjera kada želimo znati koja skupina ispitanika varira u većoj mjeri u nekoj osobini (npr. variraju li u školskom uspjehu više djevojčice ili dječaci) ili koja osobina (od više njih) varira u većoj mjeri na istoj skupini ispitanika (npr. variraju li ispitanici više u neuroticizmu ili ekstraverziji)

Koeficijent varijabilnosti

- Primjer: SD=10 uz aritmetičku sredinu M=100 je relativno manja od SD=2 uz M=8
  - V_A=10 %
  - V_B=25 %
- Može se zaključiti da aritmetička sredina bolje reprezentira rezultate mjerenja A.

Poluinterkvartilno raspršenje - Q

- polovica razlike između numeričke veličine rezultata koji se nalaze na granici 1. četvrtine (tj. kvartila) i treće četvrtine (3. kvartila) u nizu rezultata poredanih po veličini
- računa se uz centralnu vrijednost
- rijetko se koristi jer ne može poslužiti u složenijim (i korisnijim) statističkim analizama

Računanje poluinterkvartilnog raspršenja

$$Q = \frac{Q3 - Q1}{2} \qquad R_{Q1} = \frac{N}{4} \qquad R_{Q3} = \frac{N}{4} \cdot 3$$

Q – indeks poluinterkvartilnog raspršenja
Q1 – granična vrijednost 1. kvartila
Q3 – granična vrijednost 3. kvartila
R_{Q1} – redno mjesto vrijednosti 1. kvartila
R_{Q2} – redno mjesto vrijednosti 3. kvartila

7. Položaj pojedinog rezultata u grupi

Položaj pojedinog rezultata u grupi

- z – vrijednosti
- Skala centila
- Skala decila

Baždarenje

- postupak normiranja kojim se bruto rezultati nekog mjerenja pretvaraju u relativne mjerne jedinice pa se tako rezultatima dodjeljuje njihova usporedna vrijednost, tj. vrijednost koju imaju u odnosu na druge rezultate postignute u nekoj skupini rezultata ili na različitim apsolutnim skalnim vrijednostima
- pretvaranje bruto rezultata u skalne vrijednosti baždarnih skala omogućuje uvid u relativni položaj pojedinca s obzirom na njegov rezultat u skupini kojoj pripada
- npr. Neki ispitanik je na nekom testu postigao rezultat 50. Je li bio uspješan?
- na pretvaranju rezultata u z-vrijednosti zasniva se većina skala baždarenja

z-vrijednosti

- veličina površine nekog dijela krivulje normalne distribucije označava vjerojatnost pojavljivanja rezultata koji pripadaju tom dijelu ukupnog raspona rezultata
- *ukupna* površina krivulje normalne distribucije iznosi 1.0 ili 100 %
- pomoću z-vrijednosti i ***tablica površine ispod normalne krivulje*** možemo odrediti *vjerojatnost* pojavljivanja rezultata koji su jednaki i veći/manji od određene vrijednosti, kao i vjerojatnost pojavljivanja rezultata koji se kreću u određenom rasponu vrijednosti

z-vrijednosti

- Z-vrijednosti = standardne vrijednosti
- označuju položaj pojedinog rezultata u grupi rezultata koji imaju *normalnu distribuciju,* i to tako da se njegova vrijednost izrazi u *dijelovima standardne devijacije*
- ***odstupanje nekog rezultata X od pripadajuće M izraženo u terminima SD***
- kako je normalna definicija u potpunosti definirana aritmetičkom sredinom i standardnom devijacijom, za svaki rezultat, ako izračunamo na koji dio standardne devijacije pada, možemo potpuno točno ustanoviti koliki postotak rezultata je ispod, a koliki iznad njega

Veza između M, SD i z-vrijednosti

M±1SD=M±1z

M±2SD=M±2z

M±3SD=M±3z

z-vrijednosti

- ako aritmetička sredina neke normalne distribucije iznosi 100, a standardna devijacija 10, gdje pada rezultat 110?
- rezultat 110 pada na točno 1 standardnu devijaciju iznad prosjeka i ima *z – vrijednost jednaku 1*; 15.87 % rezultata ima veću vrijednost od 110

z-vrijednosti

- Primjer: Ako je M=85 i SD=8, gdje pada rezultat 73, tj. koliko ima rezultata koji su manji ili jednaki, odnosno veći od te vrijednosti?

z-vrijednosti

› odgovor na prethodno pitanje najlakše je dobiti određivanjem z-vrijednosti prema formuli:

$$Z = \frac{X - M}{SD}$$

X=rezultat
M=aritmetička sredina
SD=standardna devijacija

z-vrijednosti

- z-vrijednost rezultata 73 iznosi **z= - 1.5** ((73-85)/8=-1.5) koja se nalazi točno na polovici raspona rezultata između -2SD i -1SD; otprilike samo 7 % rezultata ima vrijednosti manju ili jednaku vrijednosti 73
- Razmislite...
  - Koliko % rezultata je veće od vrijednosti 73?

- **tablice z-vrijednosti** – pokazuju postotak slučajeva od M do pojedine z-vrijednosti (ili od pojedine z-vrijednosti do kraja distribucije) i visine ordinate (y) za zadane z-vrijednosti

z-vrijednosti

- upotreba i prednosti z-vrijednosti:
  - usporedba rezultata različitih mjerenja izraženih u različitim mjernim jedinicama kod jednog ispitanika, ali i na skupini ispitanika (npr. Je li neki učenik postigao bolji uspjeh na testu iz fizike ili povijesti; ili postižu li studenti psihologije u prosjeku veći rezultat na Ravenovim standardnim progresivnim matricama ili Cattellovom kulturalno nepristranom testu?)
  - određivanje prosječne ili skupne ocjene iz niza mjerenja čije vrijednosti jesu izražene u istim mjernim jedinicama, ali imaju različit varijabilitet

Skala decila

- ukupan broj rezultata *poredanih po veličini* dijelimo u 10 kategorija – u svaku kategoriju spada 10 % rezultata
- npr. u 1. decil spada 10 % najnižih rezultata, u 2. decil idućih 10 % rezultata, u 10. decil spada 10 % najviših rezultata
- ako neki rezultat X spada u 6. decil, to znači da u nekom skupu rezultata ima 50 % rezultata koji su manji od X i 40 % rezultata koji su veći od X, a X spada u preostalih 10 % rezultata
- rezultat koji se nalazi u određenom decilu gubi svoju individualnu vrijednost i izjednačuje se sa svim rezultatima unutar tog decila

Skala decila

- GORNJA GRANICA DECILA
  - zadnja vrijednost koja spada u određeni decil
  - za 5. decil jednaka je M (ako je distribucija normalna) ili C

  - Razmislite...
    - Koliko iznosi gornja granica 10. decila?

Skala decila

- gornja granica 10. decila ne određuje se jer završava u beskonačnosti (krakovi krivulje asimptotski se bliže apscisi), jednako kao što ni 1. decil nema granicu koja bi označavala njegov početak

Određivanje gornje granice decila

- iz **bruto rezultata**:
  - rezultati moraju biti *poredani po veličini*
  - očita se zadnja vrijednost rezultata podskupine koja obuhvaća određenih 10 % rezultata (npr. ako imamo 100 mjerenja, gornja granica prvog decila bit će jednaka 10. vrijednosti u nizu rezultata poredanih po veličini)
  - koji je problem ovakvog načina određivanja gornje granice decila?

Određivanje gornje granice decila

- pomoću **z-vrijednosti**
  - određuju se z-vrijednosti za rezultat X_1 od kojeg ima 90 % većih rezultata (gornja granica 1. decila), za rezultat X_2 od kojeg ima 80 % većih rezultata (gornja granica 2. decila), za rezultat X_3 od kojeg ima 70 % većih rezultata (gornja granica 3. decila) itd.
  - pomoću formule $\mathbf{X=z \cdot SD+M}$ odredi se vrijednost bruto rezultata za svaki decil
  - prikladna metoda ako je distribucija rezultata normalna

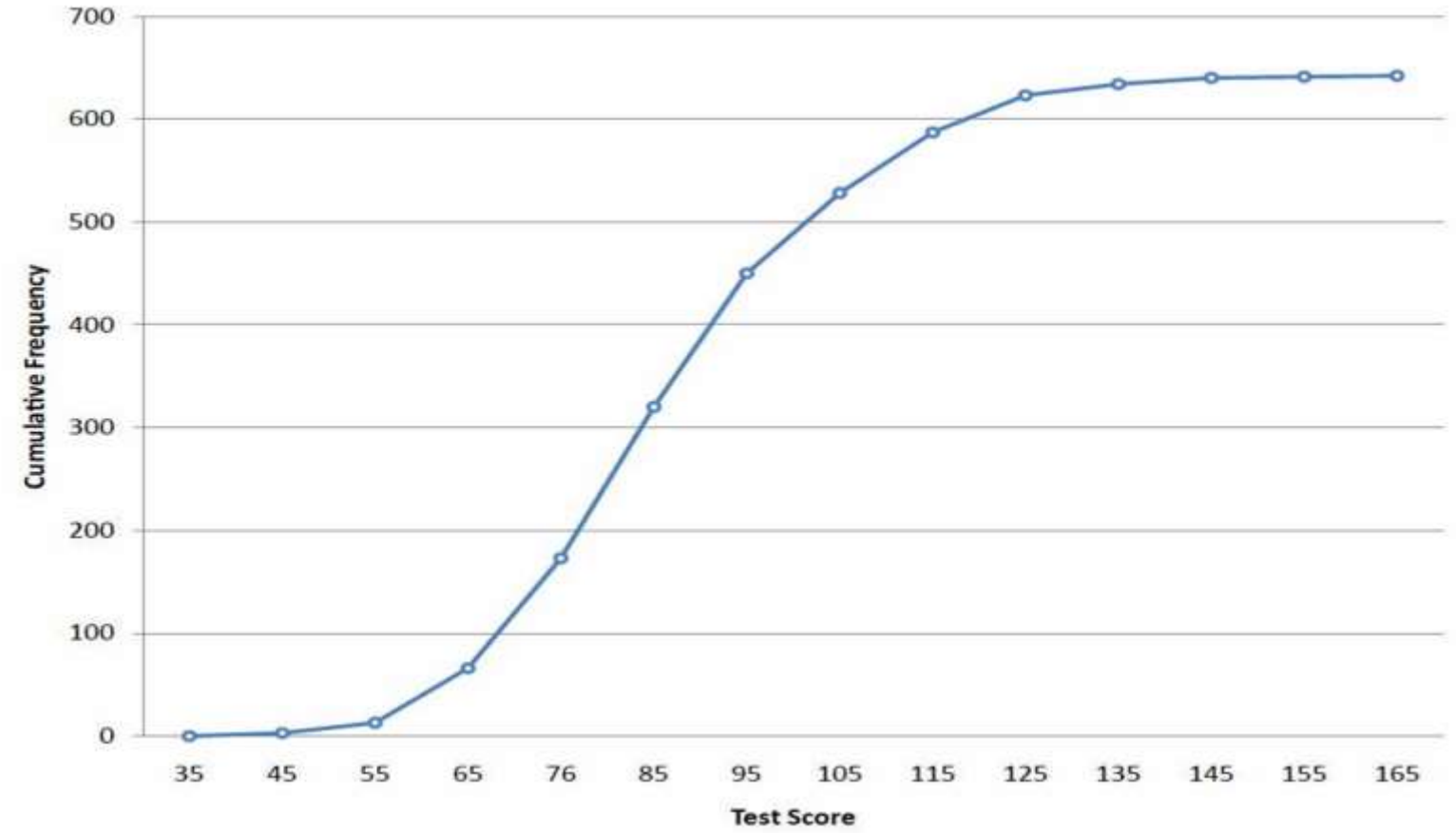
Određivanje gornje granice decila

- **grafički pomoću krivulje kumulativnih frekvencija**
  - na os X nanose se vrijednosti rezultata (ili prave gornje granice razreda), a na os Y relativne kumulativne frekvencije pojedinih rezultata (ili razreda) izražene u postotcima
  - od vrijednosti 10 % na osi Y povlači se pravac paralelan s apscisom do krivulje i zatim okomito do točke na osi X koja je jednaka gornjoj granici 1. decila
  - itd.

Krivulja kumulativnih frekvencija

- grafički prikaz distribucije rezultata u koordinatnom sustavu gdje se na apscisi nalaze dobiveni *rezultati* ili *prave gornje granice razreda*, a na ordinati *kumulativne frekvencije*
- naziva se još i **sigmoidna** krivulja (ali samo ako je distribucija normalna!)

Prikaz krivulje kumulativnih frekvencija



Preuzeto s: http://onlinestatbook.com/2/graphing_distributions/freq_poly.html

Skala decila

- prednosti:
  - ne zahtijeva normalnu distribuciju
- nedostatci:
  - neprecizna i gruba jer rezultate dijeli na samo 10 kategorija
  - dobro razlikuje rezultate samo oko aritmetičke sredine, dok u prvim i posljednjim decilima slabo razlikuje rezultate jer je varijabilitet rezultata unutar tih decila veći
  - decili nisu ekvidistantni

Skala centila

- pokazuje u kojem se postotku među svim rezultatima nalazi neki određeni rezultat
- 1. centil obuhvaća 1 % najnižih rezultata, 2. centil 1 % idućih rezultata itd., dok 100. centil obuhvaća 1 % najviših rezultata
- računski se centil nekog rezultata računa po formuli:

$$centil_x = \frac{R_x}{N} \cdot 100$$

X – rezultat
R_x- redno mjesto rezultata
N – ukupan broj rezultata

Skala centila

- određivanje centila:
  - rezultati se prvo poredaju po veličini
  - npr. ako je među 50 rezultata neki rezultat 45. po redu (računajući od najnižeg rezultata), on zauzima (45/50)100=90. centil

Skala centila

- prednosti:
  - korisna kad se rezultati ne distribuiraju normalno
- nedostatak:
  - centili ne predstavljaju ekvidistantne jedinice ako je distribucija rezultata normalna (1 % rezultata koje je smješteno blizu M (npr. 51. centil), smješteno je na mnogo užem rasponu nego 1 % rezultata koji se nalaze na ekstremu distribucije (npr. 2. centil)
  - Razmislite...
    - Kakav mora biti oblik distribucije rezultata da bi jedinice skale decila i centila bile ekvidistantne?

Skala centila

- ekvidistantne granične vrijednosti decila i centila dobivaju se jedino kod PRAVOKUTNE DISTRIBUCIJE
- zbog toga se ne smije izračunavati „prosjek” ili „ukupni centil” iz rezultata postignutih u više različitih mjerenja (kao kod z-vrijednosti)

8. Zaključivanje u statistici

Inferencijalna statistika

- bit inferencijalne statistike – ***zaključivanje*** o zakonitostima koje vrijede u *populaciji* na temelju podataka dobivenih na *uzorku*
- zasniva se na logici i zakonima NORMALNE DISTRIBUCIJE

Inferencijalna statistika

- osnovno pitanje: kolikoj se ***pogrešci*** izlažemo kada iz rezultata dobivenih na uzorku zaključujemo na populaciju, tj. vrijedi li i koliko vrijedi neka činjenica, nađena na uzorku, i za populaciju, ili je dobiveni rezultat samo posljedica slučajnog variranja među uzorcima (Kolesarić i Petz, 1999)

Inferencijalna statistika

- svodi se na testiranje statističkog modela o zakonitostima nekih pojava, odnosno na *usporedbu hipotetskog modela s dobivenim podacima*
- cilj je imati što točniji i precizniji model koji će što bolje objašnjavati pojave oko nas → statistički model mora što bolje *pristajati* uz podatke

Inferencijalna statistika

- čitava inferencijalna statistika može se svesti na temeljnu jednadžbu:

$$\text{ishod} = \text{model} + \text{pogreška}$$

- ishod = ono što smo opazili, podatci koje smo dobili
- model = statistički model koji smo pretpostavili i usporedili s dobivenim podatcima
- Pogreška = posljedica činjenice da smo podatke prikupili na ***uzorku***

Inferencijalna statistika

- primjeri statističkih modela:
  - t-test (utvrđivanje razlike u prosječnoj visini odraslih muškaraca ili žena u Hrvatskoj)
  - regresijska analiza (utvrđivanje proporcije objašnjene varijance školskog uspjeha adolescenata na temelju sociodemografskih varijabli)
  - analiza traga (utvrđivanje odnosa između osobina ličnosti, motivacije i školskog uspjeha)
  - itd.

Populacija

- „univerzum” ili osnovni skup
- svi članovi neke skupine s određenom karakteristikom koju mjerimo

- što predstavlja populaciju u sljedećim primjerima?
  - mjerenje težine prvašića u Hrvatskoj u školskoj godini 2015./2016.
  - mjerenje ispitne anksioznosti brucoša na zagrebačkom sveučilištu akademske godine 2015./2016.
  - mjerenje jednostavnog vremena reakcije na svjetlosni podražaj kod jednog ispitanika

Populacija

- ograničena populacija
  - svi slučajevi (statističke jedinice) u nekom konačnom skupu elemenata, tj. svi mogući članovi neke skupine s određenim karakteristikama (npr. kod mjerenja težine novorođenčadi u Hrvatskoj, populaciju čine *sva* novorođenčad u Hrvatskoj)
- neizmjerna ili neograničena populacija
  - neizmjeran broj slučajeva u nekom beskonačnom skupu elemenata; neizmjeran broj mjerenja (npr. kod mjerenja VR-a na svjetlosni podražaj, populaciju čini neizmjeran broj mjerenja VR-a)

Uzorak

- u većini slučajeva mjerenje se provodi na *uzorcima uzetima iz populacije.* Zašto?
  - ponekad je populaciju nemoguće izmjeriti jer je ona beskonačno velika (npr. mjerenje VR-a)
  - ponekad je mjerenje čitave populacije vrlo skupo i komplicirano (npr. mjerenje inteligencije svih osoba starije životne dobi)
  - ponekad bi mjerenje cijele populacije bilo besmisleno i štetno (ako se prilikom mjerenja uništava ono što se mjeri – testiranje Milka čokolade)

Uzorak

- dio populacije koji po svim, za mjerenje određene veličine važnim svojstvima, osim po broju elemenata, **reprezentira** populaciju
- ograničen broj članova neke populacije, izabran sa svrhom da što bolje predstavlja populaciju iz koje je izabran
- uvjeti da uzorak bude reprezentativan za populaciju:
  - mora biti dovoljno velik
  - mora biti **nepristran,** tj. ne smije biti pod utjecajem različitih sistematskih faktora
  - nepristrani uzorak → **slučajan uzorak**

Slučajan uzorak

- uzorak u kojem svaki član populacije ima jednaku šansu da bude izabran u uzorak, a izbor jednog člana ne smije ni na koji način utjecati na izbor bilo kojeg drugog člana
- to se postiže izvlačenjem uzorka *uz povrat* → nakon što se izvuče svaki pojedini rezultat, taj se rezultat vrati u populaciju; nakon toga se izvlači sljedeći rezultat pa se vraća u populaciju itd. do željene veličine uzorka
- ako je populacija vrlo velika, uzorak se *ne* mora birati uz povrat jer to neće narušiti njegovu reprezentativnost

Zašto je slučajan uzorak važan?

- slučajan uzorak omogućuje ***generalizaciju*** ili uopćavanje spoznaja dobivenih znanstvenim istraživanjem na čitavu populaciju

Slučajan uzorak

- **slučajan uzorak → normalna distribucija → logika statističkog zaključivanja**
- slučajan uzorak bira se po strogim principima
- 2 načina dobrog izbora slučajnog uzorka:
  - izvlačenje iz „bubnja” – svaki žeton ili kuglica u bubnju mora predstavljati jednu osobu, tj. jednog člana populacije
  - tablice slučajnih brojeva

Slučajan uzorak

- rezultati unutar uzorka variraju po slučaju, a veličina te slučajne varijacije ovisi o stanju populacije – što je veći **varijabilitet rezultata u populaciji**, to će biti veći i varijabilitet uzoraka uzetih iz te populacije
- ove *slučajne varijacije* predstavljaju **pogrešku**
- veličina ove pogreške određuje stupanj *pristajanja* našeg statističkog modela podatcima, odnosno određuje *možemo li, i uz koji stupanj sigurnosti (ili rizika),* ***generalizirati*** *podatke dobivene na uzorku na čitavu populaciju*

Populacijski parametri

- vrijednosti koje su prisutne u populaciji – **prava** aritmetička sredina (μ), **prava** standardna devijacija (σ), **prava** korelacija (r), **prava** proporcija (p) itd.
- iako nam je cilj utvrditi upravo ove vrijednosti, tj. populacijske parametre, u praksi je to gotovo nemoguće pa njihovu veličinu procjenjujemo na osnovu vrijednosti dobivenih na uzorku
- vrijednosti dobivene na uzorku (npr. M, SD, Q) su dakle samo (više ili manje točne) **procjene** populacijskih parametara – nazivaju se još i ***statistici***

Parametri populacije i njihova procjena

- Npr. mjerenje zaključne ocjene iz matematike na populaciji 850 učenika jedne OŠ
  - M i SD određene na svih 850 učenika nazivaju se PRAVA ARITMETIČKA SREDINA i PRAVA STANDARDNA DEVIJACIJA
  - M i SD izračunati na slučajno odabranom uzorku iz ove populacije (N=100) odnose se na PROCJENE populacijskih parametara

Osobine uzorka

- Primjer: 10000 rezultata mjerenja visine djece (N populacije iznosi 10000) napišemo na kuglice koje ubacimo u kutiju i potom vadimo 5000 (ili neizmjerno mnogo) slučajnih uzoraka veličine N=30
- potom izračunamo aritmetičku sredinu za sve uzorke
- Razmislite...
  - Kakva će biti distribucija ovih vrijednosti?
  - Kolika će biti aritmetička sredina u ovoj distribuciji?
  - Kakvo će biti raspršenje ove distribucije?

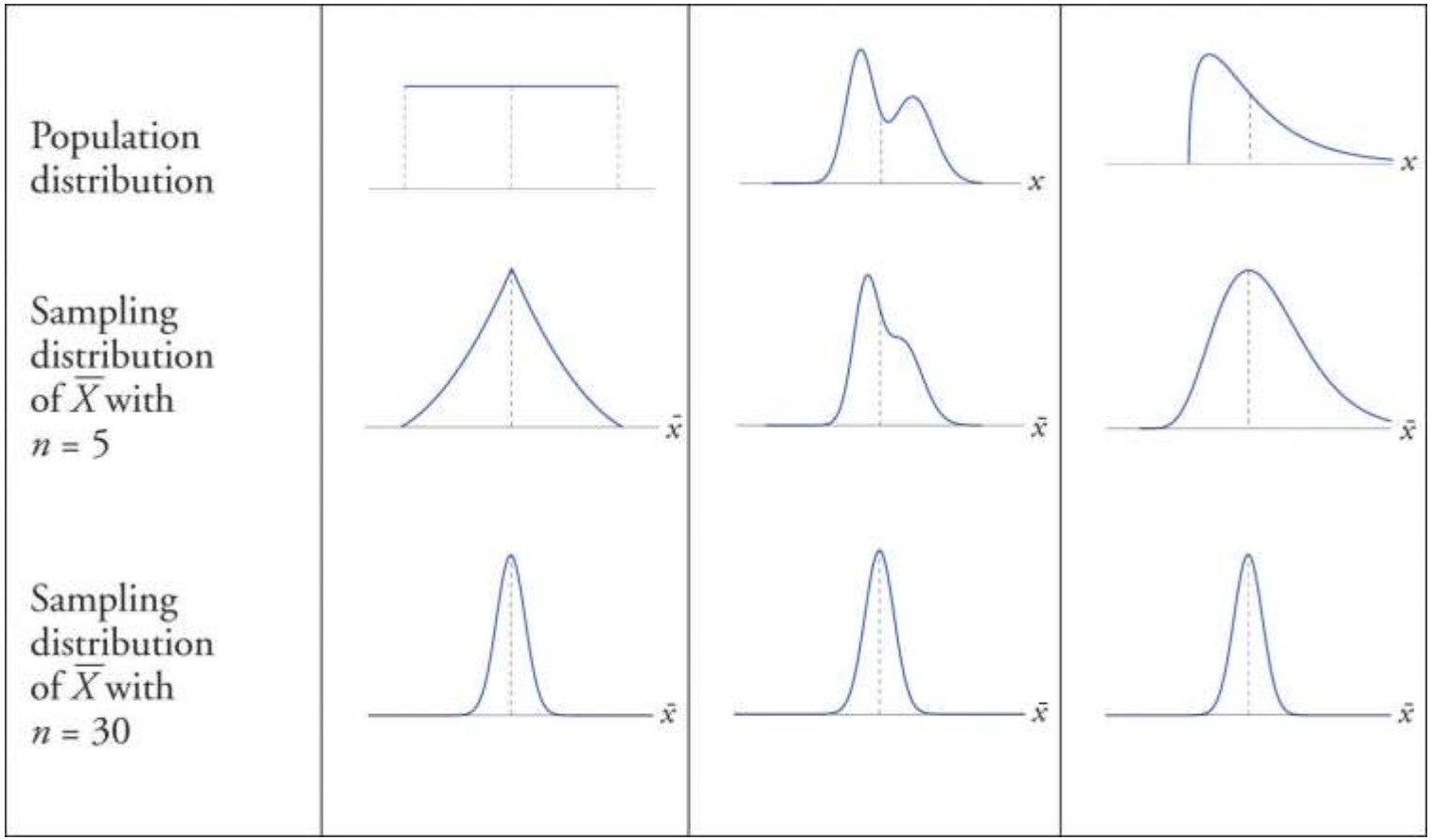
Teorem centralnih granica

1. Aritmetička sredina aritmetičkih sredina rezultata neizmjerno velikog broja uzoraka jednaka je aritmetičkoj sredini populacije ili pravoj aritmetičkoj sredini: $\mu_M=\mu$
2. Standardna devijacija distribucije aritmetičkih sredina rezultata neizmjerno velikog broja uzoraka manja je od standardne devijacije vrijednosti u populaciji
3. Distribucija aritmetičkih sredina rezultata neizmjerno velikog broja uzoraka je NORMALNA ako: **a)** su uzorci veći od 30 ili **b)** ako je distribucija iz koje su uzimani uzorci normalna

Teorem centralnih granica

- važan jer nam omogućuje donošenje nekih zaključaka i onda kada radimo s populacijama koje nisu normalno distribuirane

Teorem centralnih granica



Distribucija neizmjerno velikog broja aritmetičkih sredina uzoraka iste veličine bit će normalna ako je N>30, čak i ako se mjerena pojava u populaciji ne distribuira normalno.

Preuzeto s: https://stats.libretexts.org/Textbook_Maps/General_Statistics/Map%3A_Introductory_Statistics_(Shafer_and_Zhang)/06%3A_Sampling_Distributions/6.2%3A_The_Sampling_Distribution_of_the_Sample_Mean

9. Procjena populacijskih parametara i granice pouzdanosti

Standardna pogreška

- pogreška kojoj se izlažemo kada iz neke vrijednosti dobivene mjerenjem na uzorku zaključujemo o toj istoj vrijednosti (parametru) u populaciji (npr. o pravoj aritmetičkoj sredini, o pravoj korelaciji itd.)
- to je **standardna devijacija** distribucije pojedinih statističkih vrijednosti dobivenih na neizmjerno velikom broju uzoraka iste veličine
- logika izračunavanja standardnih pogrešaka za različite statističke vrijednosti uvijek je jednaka – tj. zanima nas kako izgleda distribucija tih vrijednosti ako bismo iz populacije uzimali neizmjerno velik broj uzoraka iste veličine

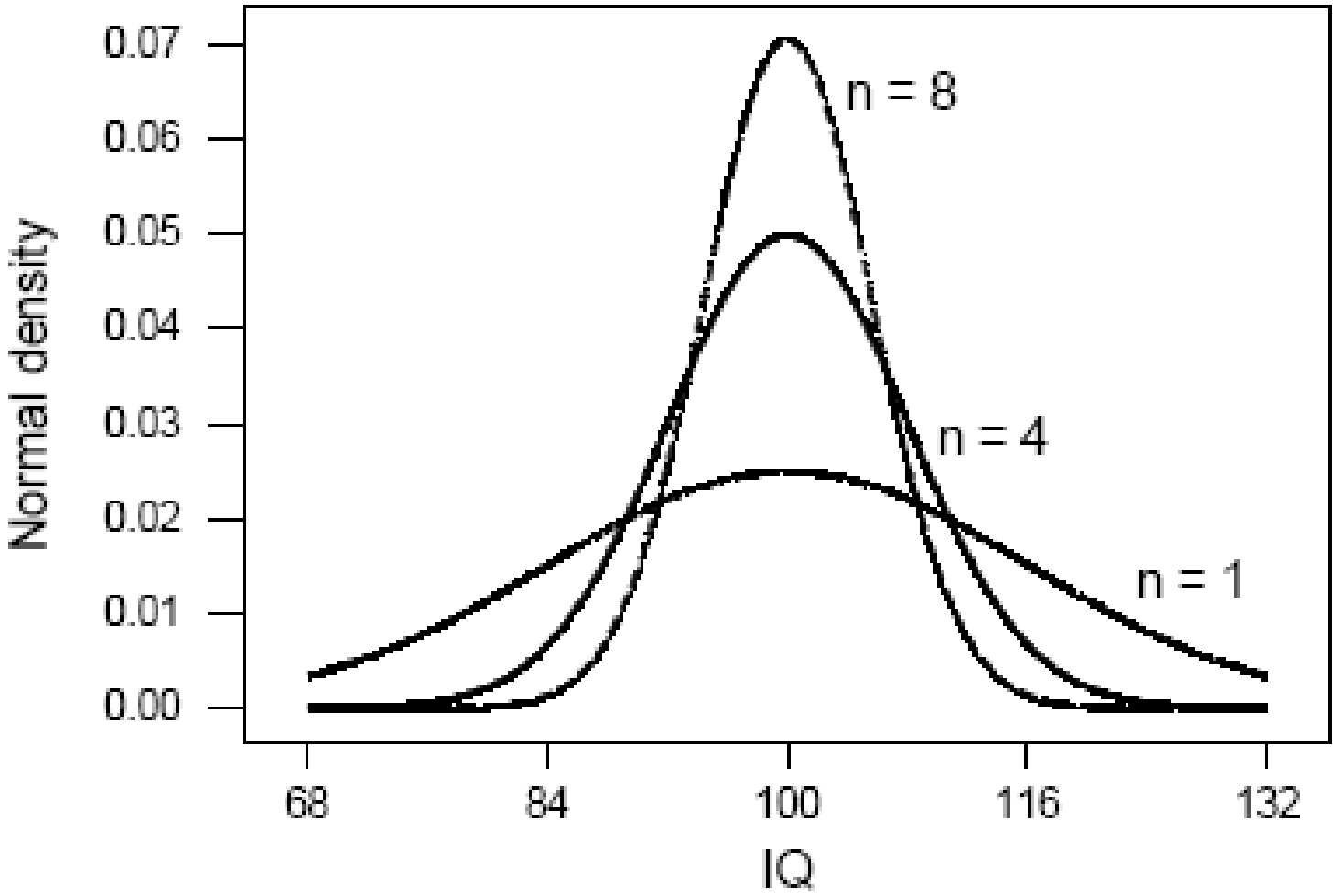
Standardna pogreška aritmetičke sredine

- standardna devijacija distribucije aritmetičkih sredina uzoraka iste veličine oko prave aritmetičke sredine, tj. oko aritmetičke sredine populacije (pravilo 2 u teoremu centralnih granica)
- ako imamo neizmjerno velik broj slučajno odabranih uzoraka, njihove aritmetičke sredine *distribuiraju se normalno* sa središnjom vrijednosti koja odgovara pravoj aritmetičkoj sredini populacije i s *raspršenjem* koje odgovara **standardnoj pogrešci aritmetičke sredine** (ako je N>30)
- Ako je N<30, onda se aritmetičke sredine distribuiraju po **t-distribuciji**

Standardna pogreška aritmetičke sredine

- varijabilitet aritmetičkih sredina neizmjerno velikog broja uzoraka slučajno odabranih iz populacije, tj. *veličina* standardne pogreške aritmetičke sredine ovisi o:
  - VARIJABILNOSTI MJERENE POJAVE U POPULACIJI – što neka pojava manje varira u populaciji, raspršenje aritmetičkih sredina neizmjerno velikog broja uzoraka bit će manje
  - VELIČINI UZORKA – što je veći N uzorka, standardna pogreška aritmetičke sredine je manja

Distribucije aritmetičkih sredina neizmjerno velikog broja uzoraka različite veličine: n=1, n=4 i n=9



Preuzeto s: https://onlinecourses.science.psu.edu/stat414/node/173

Standardna pogreška aritmetičke sredine

- formula za izračunavanje standardne pogreške aritmetičke sredine

$$\sigma_{Mp} = \frac{\sigma_p}{\sqrt{N}}$$

- σ_{Mp} – prava standardna pogreška aritmetičke sredine
- σ_P – standardna devijacija populacije
- N – veličina uzorka

Standardna pogreška aritmetičke sredine

- no, nama je vrlo rijetko, gotovo nikad, poznata veličina standardne devijacije populacije, pa se zato koristimo *standardnom devijacijom reprezentativnog uzorka* koji nam služi kao **procjena** standardne devijacije populacije
- analogno tome imamo formulu za izračunavanje **procjene standardne pogreške aritmetičke sredine**

$$\sigma_M = \frac{SD}{\sqrt{N}}$$

- σ_M – procjena standardne pogreške aritmetičke sredine
- SD – standardna devijacija uzorka
- N – veličina uzorka

Razmislite...

- Kako istraživači mogu smanjiti veličinu standardne pogreške aritmetičke sredine?

Standardna pogreška aritmetičke sredine

- Odgovor: jedino **povećanjem veličine uzorka N** jer na varijabilitet neke pojave u populaciji ne možemo utjecati
- povećanjem veličine uzorka, standardna pogreška aritmetičke sredine opada proporcionalno *drugom korijenu* broja mjerenja (N)
- npr. M dobivena iz 25 rezultata nije 25 puta bolja procjena prave aritmetičke sredine od vrijednosti dobivene samo jednim mjerenjem, već je ona 5 puta bolja procjena populacijske vrijednosti

Standardna pogreška aritmetičke sredine

- kako je ona u osnovi standardna devijacija, za standardnu pogrešku aritmetičke sredine vrijede iste zakonitosti kao i za standardnu devijaciju
- Primjer. Ako standardna pogreška neke aritmetičke sredine iznosi 1, onda to znači da se aritmetičke sredine uzoraka iste veličine grupiraju oko *prave* aritmetičke sredine po normalnoj raspodjeli kojoj standardna devijacija iznosi 1.
- Razmislite...
  - Ako smo napravili istraživanje na nekom reprezentativnom uzorku ispitanika, znamo li koliko iznosi prava aritmetička sredina?

Standardna pogreška aritmetičke sredine

- odgovor: ne znamo!
- no, znamo koliko iznosi aritmetička sredina *uzorka* i ona nam predstavlja najbolju *procjenu* prave aritmetičke sredine koju imamo
- koliko je naša aritmetička sredina uzorka dobra procjena populacijske aritmetičke sredine, ovisi o tome koliko se M-ovi neizmjerno velikog broja uzoraka međusobno razlikuju, odnosno ovisi o veličini standardne pogreške aritmetičke sredine
- što je pogreška veća, naša procjena je manje točna

Standardna pogreška aritmetičke sredine

- **<u>osnovna logika:</u>** ako standardna pogreška aritmetičke sredine iznosi 1, onda se prava aritmetička sredina (populacije) može nalaziti negdje u rasponu od „dobivena aritmetička sredina” ±3 (ali *uz 99.7 % sigurnosti!*)

Granice pouzdanosti

- Primjer: Ako smo pri nekom mjerenju na uzorku od 1000 ispitanika dobili M=90 i SD=10, standardna pogreška aritmetičke sredine iznosi $\sigma_M=1$
- ako je $\sigma_M=1$, onda je:
  - 68 % vjerojatno da se prava aritmetička sredina nalazi u rasponu od M uzorka $\pm1\sigma_M$ (89-91)
  - 95 % vjerojatno da se prava aritmetička sredina nalazi u rasponu od M uzorka $\pm2\sigma_M$ (88-92)
  - 99.7 % vjerojatno da od se prava aritmetička sredina nalazi u rasponu od M uzorka $\pm3\sigma_M$ (87-93)

Granice pouzdanosti

- dakle, gotovo sa stopostotnom sigurnošću možemo tvrditi da se prava aritmetička sredina iz ranije navedenog primjera nalazi u *intervalu* 90 ±3, tj. negdje između 87 i 93
- takav interval, koji izvodimo iz vrijednosti standardne pogreške aritmetičke sredine, naziva se **granicama pouzdanosti**
- granice pouzdanosti – raspon u kojem možemo uz veću ili manju vjerojatnost *očekivati* da se nalazi prava vrijednost neke karakteristike koju smo izmjerili na *uzorku*

Određivanje granica pouzdanosti

- Primjer: Slučajni uzorak od 100 zaposlenika velike tvrtke sudjelovao je u ispitivanju karakteristika zaposlenika. Prosječna dob radnika je M=36.4 godine, a raspršenje je SD=11 godina.
  - U kojem rasponu se kreće prosječna dob svih zaposlenika te firme?
- N=100, M=36.4, SD=11, σ_M=1.1
- granice pouzdanosti:
  - uz 95 % sigurnosti: $M \pm 1.96 \cdot \sigma_M \rightarrow 36.4 \pm 1.96 \cdot 1.1 \rightarrow 34.2 – 38.6$
  - uz 99 % sigurnosti: $M \pm 2.58 \cdot \sigma_M \rightarrow 36.4 \pm 2.58 \cdot 1.1 \rightarrow 33.6 – 39.3$

10. Testiranje hipoteza i pojam statističke značajnosti

Testiranje hipoteza

- postupak odlučivanja jesu li rezultati dobiveni u našem istraživanju, i na *uzorku*, u skladu s rezultatima prijašnjih istraživanja ili teorijskim pretpostavkama koje se mogu primijeniti na čitavu *populaciju*
- **hipoteza** – predikcija o odnosu među analiziranim pojavama koja se bazira na neformalnom opažanju, rezultatima prethodnih istraživanja ili teoriji
- hipoteze mogu biti:
  - nul-hipoteza
  - afirmativna hipoteza
  - direktivna hipoteza

Testiranje hipoteza – usporedba dvije aritmetičke sredine

- jedan od najčešćih slučajeva u statističkoj obradi podataka jest uspoređivanje dviju (ili više) aritmetičkih sredina i testiranje razlike među njima
- Primjer 1: usporedba spacijalnih sposobnosti arhitekata i profesora hrvatskog jezika
- Primjer 2: usporedba ukupnog broja zapamćenih riječi prezentiranih vizualno i auditivno na skupini studenata psihologije
- odgovore na prethodna pitanja omogućuju nam statistički postupci **testiranja razlika među aritmetičkim sredinama**

Logika uspoređivanja dviju aritmetičkih sredina

- čak i kad iz ***iste*** populacije izvučemo 2 uzorka slučajnim odabirom i potom izračunamo njihove aritmetičke sredine, vrlo je vjerojatno da će se one razlikovati u svojoj veličini, što je rezultat djelovanja pukog slučaja prilikom odabiranja članova u uzorak (tj. posljedica je pogreške uzorkovanja)
- kako mi u praksi vrlo rijetko provodimo istraživanja na čitavoj populaciji, već uglavnom to činimo na uzorcima, postavlja se pitanje je li neka pronađena razlika između skupine A i B rezultat djelovanja slučajnih faktora prilikom odabira uzorka ili skupine A i B zapravo pripadaju **različitim populacijama**
- kada imamo dva mjerenja na jednoj skupini ispitanika (npr. razlike u broju zapamćenih riječi kada su one prezentirane auditivno i vizualno), glavno je pitanje je li NZV (način prezentacije) djelovala na ZV (broj zapamćenih riječi) (u ovom slučaju rezultati pojedine ekperimentalne situacije čine različite populacije!)

Statistička značajnost

- ako dobivena razlika između aritmetičkih sredina ***nije*** rezultat slučajnih faktora (dakle, skupine A i B pripadaju različitim populacijama) kažemo da je ona **STATISTIČKI ZNAČAJNA**

Statistička značajnost

- razumijevanje pojma **statističke značajnosti**
  - neka činjenica utvrđena na uzorku (ili uzorcima) *nije slučajna*, već ona, uz određeni rizik donošenja pogrešnog zaključka, ***postoji i u populaciji***
  - ako je neka razlika između dvije aritmetičke sredine statistički značajna, ona nije slučajna već postoji i u populaciji – istu razliku dobili bismo i kad bismo ispitali sve članove tih dviju populacija
  - riječ „značajno” ima drugačije značenje u svakodnevnom životu – u svakodnevnom govoru riječ “značajno” upućuje na nešto *važno i veliko*, a statistički značajna razlika zapravo ne mora uopće biti važna ili velika!

Statistička značajnost

- u znanosti kao što je psihologija, gdje postoji dosta velik varijabilitet pojava koje mjerimo, vrlo je važno ustanoviti je li neka pojava odraz djelovanja slučajnih čimbenika ili je možemo smatrati „općim zakonom”
- provjeravanje je li neka razlika statistički značajna od velike je važnosti jer omogućuje *generalizaciju* činjenica utvrđenih na uzorku na čitavu populaciju
- ako utvrdimo da je neka pojava statistički značajna, odnosno da vrijedi i u populaciji, dobiveni rezultati dobivaju na *znanstvenoj vrijednosti*

Primjer istraživanja

- Neke istraživače zanimalo je razlikuju li se ljudi različitih profesija u kognitivnim sposobnostima. Stoga su na slučajnom uzorku od 500 diplomiranih inženjera arhitekture i slučajnom uzorku od 500 profesora hrvatskog jezika primijenili test spacijalnih sposobnosti pri čemu su utvrđeni sljedeći deskriptivni pokazatelji: M_A=120, SD_A=10; M_H=109, SD_H=12.

Primjer istraživanja

- **Istraživačko pitanje:** postoji li razlika utvrđena na ova dva uzorka i u populaciji, odnosno postoji li razlika u spacijalnim sposobnostima između dipl. ing. arhitekture i profesora hrvatskog jezika?
- **Istraživačka hipoteza (direktivna)**: diplomirani inženjeri arhitekture imaju bolje razvijene spacijalne sposobnosti od profesora hrvatskog jezika: $\mu_A > \mu_H$
- **Nul-hipoteza:** nema razlike u stupnju razvijenosti spacijalnih sposobnosti između dipl. ing. arhitekture i profesora hrvatskog jezika, odnosno nema razlike između ovih dviju populacija: $\mu_A = \mu_H$

Nul-hipoteza vs. afirmativna hipoteza

- nul-hipoteza znači da u populaciji nema razlike među pojavama koje mjerimo
  - npr. zanima nas jesu li u testu spacijalnih sposobnosti prosječno uspješniji muškarci ili žene – prema **nul-hipotezi** nema razlike u spacijalnim sposobnostima između muškaraca i žena, odnosno *distribucije* spacijalnih sposobnosti svih muškaraca i svih žena (dakle, čitavih populacija), međusobno se ne razlikuju, tj. imaju iste aritmetičke sredine i iste standardne devijacije – drugim riječima, *preklapaju se*
  - ako je nul-hipoteza točna, muškarci i žene pripadaju istoj populaciji (s jedinstvenom distribucijom)

Nul-hipoteza vs. afirmativna hipoteza

- **Afirmativna hipoteza**: postoji statistički značajna razlika između dvije aritmetičke sredine
- može biti:
  - *jednosmjerna* (*one-tailed*) – razliku ćemo pronaći u jednom smjeru (npr. muškarci postižu bolji prosječni rezultat na testu spacijalnih sposobnosti od žena) – naziva se još i ***direktivna***
  - *dvosmjerna* (*two-tailed*) – razliku ćemo pronaći, i to u bilo kojem smjeru (muškarci i žene razlikuju se u prosječnom rezultatu na testu spacijalnih sposobnosti)

Testiranje hipoteza

- hipoteze testiramo određenim statističkim postupkom (npr. t-testom) na podatcima dobivenim na uzorku
- rezultat tog testiranja može biti *prihvaćanje* ili *odbacivanje* hipoteze
- **iako je većina hipoteza koje postavljaju istraživači po svojoj prirodi afirmativna i direktivna, statistički se testira nul-hipoteza, i to najčešće dvosmjernim testom (za razliku od jednosmjernog testa)!**

Dvosmjerno testiranje

- dvosmjernim testom testiramo $\mathbf{H_0}$, odnosno provjeravamo vrijedi li neka razlika između dvije aritmetičke sredine pronađene na uzorcima i u populaciji
- pri tom nas ne zanima *smjer* razlike, već samo postoji li ona u populaciji
- npr. zanima nas razlikuje li se visina djece u Portugalu od visine djece u Hrvatskoj
  - po slučaju smo odabrali 800 malih Hrvata i 800 malih Portugalaca iste dobi i izmjerili njihovu visinu
  - ustanovili smo da razlika između dvije aritmetičke sredine ovih dviju skupina iznosi 3 cm u korist malih Hrvata

Dvosmjerno testiranje

- ta razlika je proglašena statistički značajnom uz nivo rizika manji od 5 % - **to znači da bi se, u slučaju kada među populacijama ne bi postojala razlika, takva (ili veća) razlika od 3 cm mogla slučajno pojaviti najviše u 5 % slučajeva**
- gledamo razlike koje su *apsolutno* veće od 3 cm (±3 cm) i koje padaju u regiju vrijednosti većih (u apsolutnom smislu) od onih koje se pojave u 2.5 % slučajeva na *svakoj* strani krivulje ($p<0.05 \rightarrow$ 2.5+2.5=5 %), odnosno od 0.05 % slučajeva na *svakoj* strani krivulje ($p<0.1 \rightarrow$0.5+0.5=1 %)

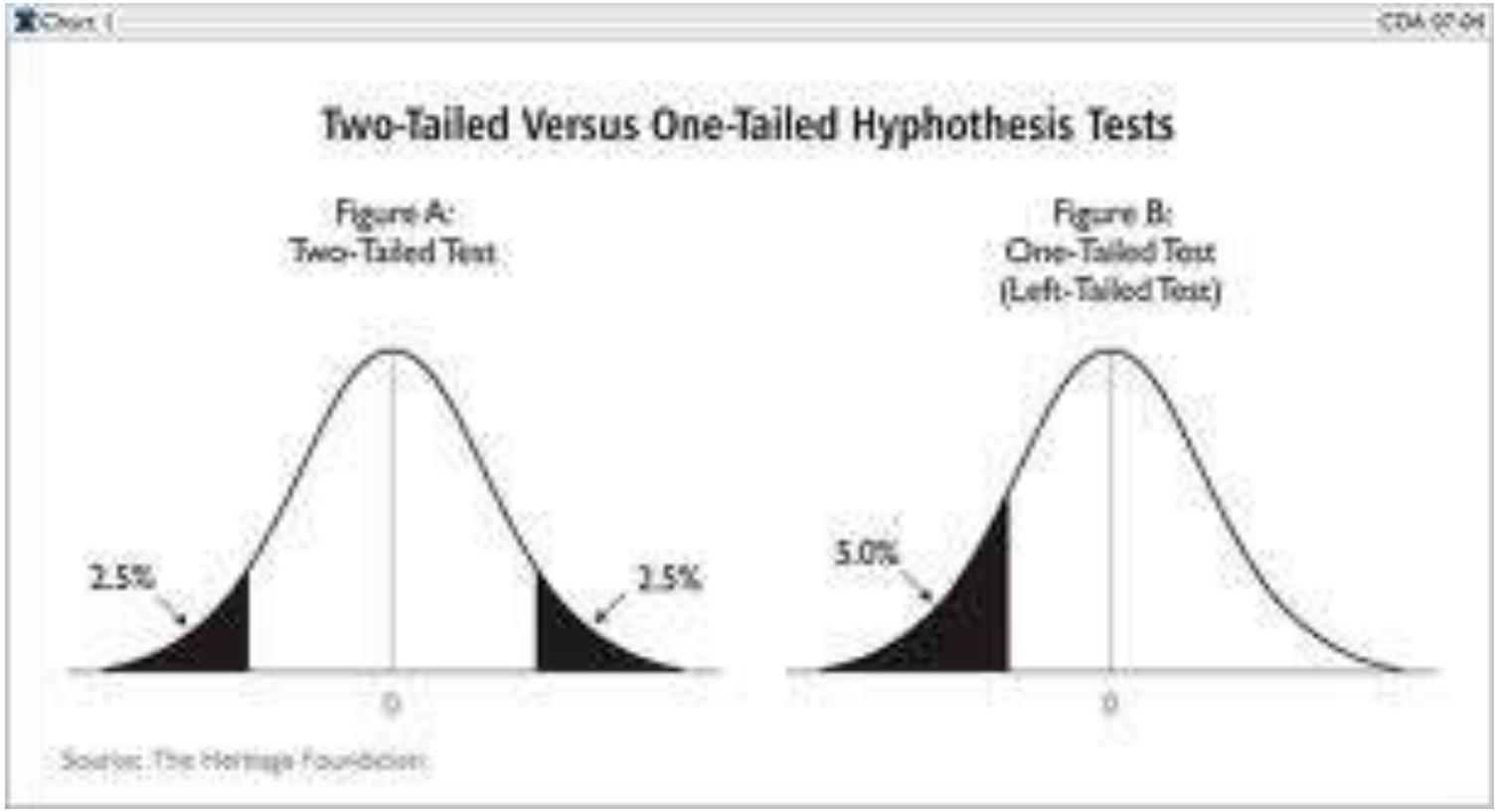
Jednosmjerno testiranje

- npr. ako na skupini od 1000 bolesnika primijenimo novi lijek za reguliranje visokog krvnog tlaka, može nas zanimati *je li nakon primjene lijeka došlo do statistički značajnog smanjenja krvnog tlaka*
- to praktički znači da ako smo nakon primjene lijeka utvrdili povećanje krvnog tlaka, tu razliku ne bismo ni testirali!
- jednosmjernim testom testiramo uglavnom **direktivne hipoteze**

Jednosmjerno testiranje

- da bismo utvrdili je li smanjenje krvnog tlaka statistički značajno uz nivo rizika od 5 %, pomoću t-tablica (ili tablica z-vrijednosti kad su uzorci veliki!) provjeravamo je li dobivena t-vrijednost veća od granične uz $p<0.05$ (ali samo s jedne strane t-distribucije!) → ako je **t>1.64**, (kod velikih uzoraka) dolazi do statistički značajnog smanjenja visokog krvnog tlaka nakon primjene novog lijeka
- ako se odlučimo za rizik od 1 %, dobivenu t-vrijednost uspoređujemo s graničnom uz $p<0.01$ → **t>2.30** (kod velikih uzoraka) dolazi do statistički značajnog smanjenja visokog krvnog tlaka
- granične vrijednosti kod jednosmjernog testa niže su od graničnih vrijednosti kod dvosmjernog, ali **ukupna razina rizika,** tj. statističke značajnosti ostaje ista!

Jednosmjernio vs. dvosmjernio testiranje hipoteza



Preuzeto s: http://willgervais.com/blog/2014/9/2/a-tale-of-two-tails

Jednosmjerni vs. dvosmjerni test

- kod testiranja značajnosti razlika u pravilu se upotrebljava dvosmjerni test, a samo u iznimnim i opravdanim slučajevima jednosmjerni test
- način testiranja (jednosmjerno ili dvosmjerno) treba odrediti ***prije*** nego što su nam poznati rezultati mjerenja – dakle, u fazi planiranja istraživanja ako zato postoji teorijsko i/ili praktično opravdanje
- budući da je kritična vrijednost u jednosmjernom testu *manja* od kritične vrijednosti dvosmjernog testa, neki istraživači, nakon što ne pronađu statistički značajnu razliku u dvosmjernom testu, naprave jednosmjerni test i tako eventualno „postignu" statističku značajnost, što je **neopravdano!**

Testiranje hipoteza

- **prihvaćanje nul-hipoteze** znači da ne postoji statistički značajna razlika između dvije skupine u ispitivanoj pojavi
  - međutim, to ne znači posve sigurno da ta razlika uistinu ne postoji u populaciji, već to samo znači da mi *ne možemo* tvrditi da postoji jer je nismo uspjeli potvrditi našim istraživanjem
- **odbacivanje nul-hipoteze** znači da postoji statistički značajna razlika između dvije skupine u ispitivanoj pojavi
  - međutim, utvrđivanje razlike ne znači da je zbog toga automatski prihvaćena *direktivna* hipoteza jer nju nismo testirali!

11. Provjeravanje razlika među aritmetičkim sredinama

Razlika između dvije aritmetičke sredine

- značajnost neke razlike između dvije aritmetičke sredine možemo provjeravati:
  - uz pomoć **„granica pouzdanosti”** (najčešće se koristi za testiranje razlike između aritmetičke sredine uzorka i neke unaprijed poznate (populacijske) vrijednosti
  - uz pomoć **t-testa**

Provjeravanje razlika između aritmetičkih sredina: **granice pouzdanosti**

- određivanje razlikuje li se dobivena aritmetička sredina uzorka od neke druge (najčešće populacijske) ***fiksne*** vrijednosti
- to je najčešće neka poznata vrijednost koja predstavlja parametar populacije i za koju nam ne moraju biti poznati podatci o varijabilitetu
- **logika:** ako interval, tj. granice pouzdanosti, uz određenu razinu sigurnosti (najčešće 95 % i/ili 99 %) obuhvaća zadanu (fiksnu, populacijsku) vrijednost, prihvaćamo nul-hipotezu, tj. zaključujemo da se prosječna vrijednost utvrđena na uzorku ne razlikuje *statistički značajno* od populacijske vrijednosti

Provjeravanje razlika između aritmetičkih sredina: **granice pouzdanosti**

- **Primjer**: U nekom mjerenju 144 djece dobivena je prosječna težina M=34 kg i SD=4.8 kg. Razlikuje li se dobivena aritmetička sredina statistički značajno od vrijednosti 32 kg koja se smatra normom za djecu te dobi?
  - standardna pogreška aritmetičke sredine: σ_M=0.4
  - određivanje intervala u kojem se nalazi *prava aritmetička sredina* pomoću **granica pouzdanosti**
    - 34±1.96·0.4 → 33.22 – 34.78 (95 % sigurnost)
    - 34±2.58·0.4 → 32.97 – 35.03 (99 % sigurnost)
  - populacijska vrijednost X=32 nalazi se izvan oba intervala, što znači da se *aritmetička sredina uzorka statistički značajno razlikuje od populacijske vrijednosti, tj. norme* ***uz razinu sigurnosti od 99 %***
  - to zapravo znači da uzorak na kojem je provedeno istraživanje pripada populaciji različitoj od ove za koje imamo norme (i u to tvrdimo s 99 %-tnom sigurnošću)

t-test

- t-odnos se općenito odnosi na razlomak kod kojeg se u brojniku nalazi neka vrijednost, a u nazivniku pogreška te vrijednosti
- t-odnos je najpoznatiji kod testiranja značajnosti razlika između dvije aritmetičke sredine i u tom slučaju on pokazuje *koliko je puta neka razlika veća od svoje pogreške*

$$t = \frac{\Delta}{\sigma_{\Delta}}$$

Δ - razlika između M-ova 2 uzorka

σ_{Δ}- standardna pogreška razlike

Standardna pogreška razlike između dvije aritmetičke sredine

- zamislimo sljedeće:
- iz dvije *različite* populacije uzimamo neizmjerno velik broj parova uzoraka (iz svake populacije odabiremo uzorke *iste* veličine) i svaki put za svaki uzorak izračunamo aritmetičku sredinu
- potom odredimo razlike između pojedinih parova aritmetičkih sredina i napravimo distribuciju tih razlika
- te će se razlike distribuirati po *normalnoj raspodjeli* kojoj je aritmetička sredina ***prava* razlika među aritmetičkim sredinama** dviju populacija, a pripadajuće raspršenje, tj. standardna devijacija te distribucije naziva se **STANDARDNA POGREŠKA RAZLIKE IZMEĐU DVIJE ARITMETIČKE SREDINE**

Standardna pogreška razlike između dvije aritmetičke sredine

$$\sigma_{M1-M2} = \sqrt{\left(\sigma_{M1}^{2} + \sigma_{M2}^{2}\right)}$$

– σ_{M1-M2} - standardna pogreška razlike
– σ_{M1} – standardna pogreška aritmetičke sredine prvog uzorka
– σ_{M2} – standardna pogreška aritmetičke sredine drugog uzorka

Standardna pogreška razlike između dvije aritmetičke sredine

- standardna pogreška razlike po svom smislu predstavlja **standardnu devijaciju** pa za nju vrijede isti zakoni kao i za standardnu devijaciju:
  - 68 % je vjerojatno da se prava razlika nalazi u intervalu dobivena razlika ±1 standardna pogreška razlike
  - 95 % je vjerojatno da se prava razlika nalazi u intervalu dobivena razlika ±2 standardna pogreška razlike
  - 99.7 % vjerojatno da se prava razlika nalazi u intervalu dobivena razlika ±3 standardna pogreška razlike

Standardna pogreška razlike između dvije aritmetičke sredine

- uzimamo li uzorke iz dviju populacija koje se međusobno ni po čemu **ne razlikuju** (dakle, zapravo je riječ o *istoj* populaciji!), također bismo dobili normalnu distribuciju razlika između dviju aritmetičkih sredina → uzorci se međusobno razlikuju po aritmetičkim sredinama čak i kad potječu iz iste populacije zbog djelovanja slučajnih faktora
- ali je *aritmetička sredina tih razlika jednaka 0* – jer je riječ o uzorcima koji zapravo pripadaju istoj populaciji
- kada je *aritmetička sredina razlika ≠ 0*, ona je *statistički značajna,* tj. imamo uzorke koji pripadaju dvjema različitim populacijama

t-distribucija

› zamislimo sljedeći pokus:
  – iz *iste* populacije izvlačimo mnoštvo velikih uzoraka
  – izračunamo aritmetičku sredinu svakog uzorka
  – odredimo razlike između aritmetičkih sredina pojedinih uzoraka - d
  – izračunamo t-vrijednost za svaku pojedinu razliku

→ ako su uzorci dovoljno veliki, razlike između aritmetičkih sredina i pripadajuće t-vrijednosti distribuirale bi se po **normalnoj raspodjeli** (0±3t), kojoj standardna devijacija približno odgovara izračunatoj standardnoj pogrešci razlike

→ *t* predstavlja jednu vrstu z-vrijednosti: kao i z, t je odstupanje izraženo u terminima standardne devijacije → t=1 znači da je neka nađena razlika među aritmetičkim sredinama na jednoj standardnoj devijaciji raspršenja svih slučajnih razlika koje se mogu dogoditi (pritom *prava* razlika iznosi 0)

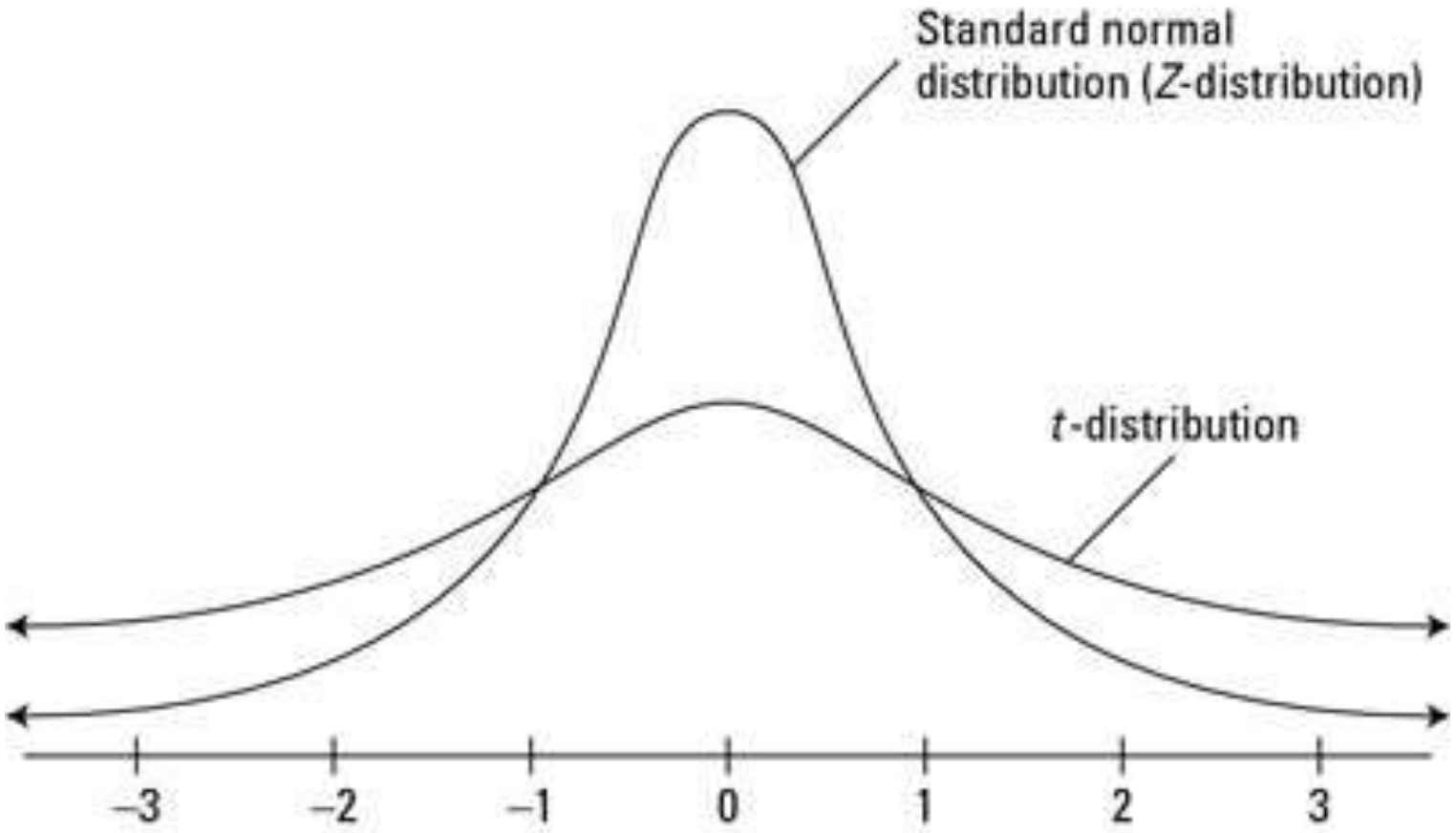
t-distribucija

- ako isti pokus ponovimo na *malim uzorcima,* razlike između aritmetičkih sredina (d) oko prave razlike u populaciji također će se distribuirati po *normalnoj* raspodjeli
- ali distribucija njihovih pripadajućih t-vrijednosti neće biti normalna – t-vrijednosti će se distribuirati po ***t-raspodjeli***
- t-raspodjela naziva se još i „Studentova" distribucija jer se njezin autor Gosset potpisivao pseudonimom "Student"
- oblik t-raspodjele ovisi o veličini uzorka, tj. o stupnjevima slobode

t-distribucija

- zvonolika i simetrična (kao i normalna distribucija), ali šira od nje, to više što je N manji
- u normalnoj distribuciji 99.74 % svih rezultata obuhvaćeno je rasponom između M±3SD
- kod t-distribucija taj je postotak to *manji* što je manji uzorak – t-distribucija je šira od normalne distribucije tim više što je uzorak manji

t – distribucija



t-distribucija je šira na krajevima u usporedbi s normalnom distribucijom.

Preuzeto s:http://www.dummies.com/education/math/statistics/how-to-tell-a-z-distribution-from-a-t-distribution/

Stupnjevi slobode

- engl. *degrees of freedom* – df
- broj događaja koji pri izračunavanju neke statističke vrijednosti mogu slobodno (nezavisno) varirati
- npr. ako u uzorku imamo 10 rezultata i izračunamo M, samo 9 devijacija pojedinačnih rezultata od aritmetičke sredine može varirati; 10. devijacija mora iznositi točno onoliko koliko je potrebno da suma svih devijacija od M iznosi 0
- npr. ako zbrajanjem 4 broja moramo dobiti 10, tada prva tri broja biramo kako želimo, ali posljednji mora biti jednak razlici između zbroja prva tri broja i broja 10

Stupnjevi slobode

- njihov broj ovisi o tome što je sve zajedničko u dvije skupine rezultata koje uspoređujemo – npr. ako testiramo je li neka distribucija „normalna”, onda opažene distribucije i idealne (očekivane) distribucije imaju isti N, istu M i istu SD, pa zato stupnjevi slobode iznose *broj kategorija – 3*

Razina rizika ili nivo značajnosti – *p*

- vjerojatnost da ćemo dobiti t-vrijednost koja će u *apsolutnom* smislu biti jednaka ili veća od određene granične t-vrijednosti
- kada su **uzorci veliki**, vjerojatnost da ćemo dobiti t koji će u apsolutnom smislu biti veći od 1.96 iznosi 5 %; vjerojatnost da ćemo dobiti t koji će u apsolutnom smislu biti veći od 2.58 iznosi 1 %

- 5 % i 1 % predstavljaju konvencionalne **razine rizika,** tj. p

Razina rizika

- uvriježeni kriteriji:
  - **razina rizika od 5 %:** razlika mora biti barem 1.96 puta veća od svoje pogreške (**t>1.96**) da bi se proglasila statistički značajnom → svaka razlika koja se nalazi izvan intervala $0 \pm 1.96 \cdot \sigma_{M1-M2}$ statistički je značajna
  - **razina rizika od 1 %:** razlika mora biti barem 2.58 puta veća od svoje pogreške (**t>2.58**) da bi se proglasila statistički značajnom → svaka razlika koja se nalazi izvan intervala $0 \pm 2.58 \cdot \sigma_{M1-M2}$ statistički je značajna

  - Napomena: ovo vrijedi samo kad su uzorci veliki!!!

Razina rizika

- **p<0.05** = dobivena vrijednost je statistički značajna uz razinu rizika manju od 5 % - vjerojatnost da će se (toliko velika) vrijednost po slučaju pojaviti u populaciji manja je od 5 %.
- **p<0.01** = dobivena vrijednost je statistički značajna uz razinu rizika manju od 1 % - vjerojatnost da će se (toliko velika) vrijednost po slučaju pojaviti u populaciji manja je od 1 %.
- **p<0.05, p>0.01** = dobivena vrijednost je statistički značajna uz razinu rizika manju od 5 %, ali ne i uz razinu rizika manju od 1 % - vjerojatnost da će se (toliko velika) vrijednost po slučaju pojaviti u populaciji manja je od 5 %, ali veća od 1 %.
- **p>0.05** = dobivena vrijednost nije statistički značajna, tj. vjerojatnost da će se (toliko velika) vrijednost po slučaju pojaviti u populaciji veća je od 5 % (što je uglavnom neprihvatljiv rizik)

Pogreška tipa I

- pogreška koja nastaje kad odbacimo nul-hipotezu, iako je ona zapravo točna
- neku razliku proglasimo statistički značajnom, iako stvarno nema razlike među populacijama
- javlja se uz *blaži kriterij,* tj. uz veće razine značajnosti (5 ili 10%)
- **α=razina rizika**

Pogreška tipa 2

- prihvaćamo nul-hipotezu koja stvarno nije istinita
- mi proglasimo neku razliku statistički neznačajnom, iako ona stvarno postoji u populaciji
- javlja se uz *stroži* kriterij (npr. uz razinu značajnosti od 1 % ili 0.1 %)

Odabir razine značajnosti/rizika

- razinu značajnosti odabiremo u skladu sa situacijom:
  - kada želimo biti gotovo potpuno sigurni da neki postupak ima učinka (jer je veoma skup), treba uzeti stroži kriterij i manji nivo rizika (npr. 1 % ili 0.1 %)
  - radi li se pak o postupku koji nije ni kompleksan ni skup ni opasan, a mogao bi imati pozitivan utjecaj na situaciju, odlučujemo se za blaži kriterij
- u znanosti se najčešće nastoji izbjeći pogreška tipa I, pa se češće susrećemo sa strogim kriterijem – dakle, prednost se daje potvrđivanju nul-hipoteze

Tablice t-vrijednosti

- sadrže granične t-vrijednosti za određenu razinu rizika i određene *stupnjeve slobode*:
  - t-vrijednost koju smo mi dobili u svom istraživanju mora biti **jednaka ili veća** od one u tablici da bismo taj parametar (npr. razliku između dvije M) mogli proglasiti statistički značajnim **uz odabranu razinu rizika** (najčešće je to razina rizika od 1 % ili 5 %)

Tablice t-vrijednosti

- što je uzorak manji, imamo *manje stupnjeva slobode*, a granična t-vrijednost je veća – kriterij za *proglašavanje neke razlike statistički značajnom je stroži*
- primjer: kad imamo 10 stupnjeva slobode, granična t-vrijednost uz **nivo rizika** 0.05 iznosi 2.23 (što je veće od granične vrijednosti 1.96 koja se pronalazi kod velikih uzoraka)
- to znači da kod uzorka veličine N=10 razlika između dvije aritmetičke sredine mora biti barem 2.23 puta veća od svoje pogreške da bi je mogli smatrati statistički značajnom na razini od 0.05

t-test

- kada dobijemo **t>1.96**, to znači da je razlika između dvije aritmetičke sredine statistički značajna uz rizik od pogrešnog zaključka manji od **5 %** (u 5 od 100 slučajeva, što je malo vjerojatno) → dakle, odabravši ovaj kriterij, odlučili smo se za RAZINU ZNAČAJNOSTI ili NIVO RIZIKA manji od 5 % (sa sigurnošću većom od 95 % možemo tvrditi da dobivena razlika nije rezultat djelovanja slučajnih faktora)
- kada nam je važno uzeti *stroži kriterij*, razliku između dvije aritmetičke sredine uz **t>2.58** proglašavamo statistički značajnom uz rizik pogrešnog zaključka manji od **1 %** (u 1 od 100 slučajeva) → odabrali smo razinu značajnosti ili nivo rizika manji od 1 % (sa sigurnošću većom od 99 % možemo tvrditi da razlika nije slučajna)

t-test na jednom uzorku

- Koristi se za provjeravanje razlika između aritmetičke sredine uzorka i prave aritmetičke sredine (populacije), pri čemu nam je poznata vrijednost μ, ali ne σ populacije
- računa se po formuli:

$$t = \frac{M - \mu}{\sigma_M}$$

df=N-1

M – aritmetička sredina
μ – aritmetička sredina populacije
σ_M – standardna pogreška aritmetičke sredine

t-test na jednom uzorku

- **Primjer**: Na nekom sveučilištu studenti u prosjeku uče 17 h tjedno. Jedan student primijetio je da njegovi kolege iz studentskog doma uče više od 17 h tjedno. Na skupini od 16 studenata koji žive u domu utvrdio je da u prosjeku uče 21 h tjedno (SD=6.80). Razlikuju li se studenti koji žive u domu od studenata sveučilišta općenito u broju sati koje tjedno provedu u učenju?
- Pomoću formule dobivena je t-vrijednost koja iznosi 2.35; za N=16 imamo 15 stupnjeva slobode; uz razinu rizika od 1 % i uz 15 stupnjeva slobode granična vrijednost t iznosi 2.947, a uz razinu rizika od 5 % granična t-vrijednost iznosi 2.132
- Naša t-vrijednost veća je od granične vrijednosti uz razinu rizika manju od 5 %, ali ne i uz rizik manji od 1 % – to pišemo ovako: **t=2.35, df=15, p<0.05, p>0.01**
- Interpretacija: ***Utvrđena je statistički značajna razlika u broju sati provedenih u učenju tjedno između studenata koji žive u studentskom domu i ostalih studenata na sveučilištu uz razinu rizika od pogrešnog zaključka manju od 5 %, ali ne i uz razinu rizika manju od 1 %***
- praktično, to znači da sa sigurnošću od 95 % možemo tvrditi da ova razlika postoji i u populaciji, ali to isto ne možemo tvrditi s 99 %-tnom sigurnošću, odnosno vjerojatnost da se ovako velika razlika pojavi u populaciji po slučaju je manja od 5 %, ali veća od 1 %.

t-test: veliki nezavisni uzorci

- ako je neka razlika između dviju aritmetičkih sredina *približno dva do tri puta* veća od svoje vlastite pogreške, onda je možemo smatrati statistički značajnom jer je vrlo malo vjerojatno da će se tako velika razlika dogoditi slučajno (koliko iznosi ta vjerojatnost?)

t-test: veliki nezavisni uzorci

› koliko je puta neka razlika veća od svoje pogreške možemo odrediti tako što ćemo neku *razliku podijeliti s njezinom pogreškom* → tako se dobiva **t-odnos** ili t-vrijednost

$$t = \frac{M_1 - M_2}{\sqrt{\sigma_{M1}{}^2 + \sigma_{M2}{}^2}}$$

df=(N-1)+(N-1)

M_1 – aritmetička sredina uzorka 1
M_2 – aritmetička sredina uzorka 2
σ_{M1} – standardna pogreška aritmetičke sredine uzorka 1
σ_{M2} – standardna pogreška aritmetičke sredine uzorka 2

t-test: veliki nezavisni uzorci

- statistička značajnost t-testa:
  - **$t<1.96$, $p>0.05$**
  - **$1.96<t<2.58$, $p>0.01$, $p<0.05$**
  - **$t>1.96$, $p<0.05$**
  - **$t>2.58$, $p<0.01$**

  - granične vrijednosti t=1.96 i t=2.58 vrijede *samo ako su uzorci vrlo veliki* (npr. N=500) – što je N manji, granična vrijednost t mora biti veća da bi se neka razlika proglasila statistički značajnom (**pogledati u t-tablicu!)**

t-test: veliki nezavisni uzorci

- ako je razlika između dviju aritmetičkih sredina uzoraka statistički značajna, to znači da se *aritmetičke sredine* dviju populacija međusobno razlikuju
- to ne znači da se svi individualni rezultati jedne populacije razlikuju od svih individualnih rezultata druge populacije – može biti „prekrivanja” distribucija
- Razmislite...
  - Koliko velika treba biti razlika u aritmetičkim sredinama (izražena u standardnim devijacijama) da se distribucije *gotovo* uopće ne preklapaju?

t-test: mali nezavisni uzorci

- s obzirom na to da gotovo nikad ne raspolažemo s vrijednošću standardne devijacije populacije, prilikom izračunavanja standardne pogreške razlike koristimo se *standardnom devijacijom uzorka*
- kod velikih uzoraka dobivamo uglavnom točnu *procjenu* standardne devijacije razlika mnoštva uzoraka oko prave razlike među populacijama, tj. standardne pogreške razlike
- stoga, činjenica što u računu t-testa ne koristimo pravu standardnu devijaciju populacije ne uzrokuje odstupanje od normaliteta t-distribucije te vrijedi pravilo da t-vrijednost mora biti *veća* od 1.96 da bi neku razliku proglasili statistički značajnom ($p<.05$)
- međutim, kod *malih uzoraka* **razlike** između aritmetičkih sredina (neizmjerno) velikog broja uzoraka distribuiraju se normalno oko prave razlike, no izračunati **t-odnosi** ne distribuiraju se normalno, već po t-distribuciji – obavezno trebamo koristiti **t-tablice**!

t-test: mali nezavisni uzorci

- pod pretpostavkom da oba uzorka potječu iz iste populacije (nul-hipoteza), određuje se **zajednička standardna devijacija** za oba uzorka
- pretpostavlja se da se na taj način može dobiti *točnija* procjena standardne devijacije populacije
- no, zajedničku standardnu devijaciju opravdano je računati samo ako se varijance dvaju uzoraka međusobno ne razlikuju statistički značajno – uvjet *homogenosti varijanci*
- ipak, neka istraživanja pokazuju da će t-test dati relativno točne rezultate unatoč tomu što varijance nisu homogene, ako su oba uzorka *jednako* velika – u tom će slučaju, čak i ako se varijance razlikuju statistički značajno, pogreška u računu biti neznatna

t-test: mali nezavisni uzorci

$$t = \frac{M_1 - M_2}{\overline{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \qquad \overline{\sigma} = \sqrt{\frac{{\sigma_1}^2 (N_1 - 1) + {\sigma_2}^2 (N_2 - 1)}{(N_1 - 1) + (N_2 - 1)}}$$

$df=(N_1-1)+(N_2-1)$

M_1 – aritmetička sredina uzorka 1
M_2 – aritmetička sredina uzorka 2
N_1 – broj rezultata u uzorku 1
N_2 – broj rezultata u uzorku 2
Σ^2_1 – varijanca uzroka 1
Σ^2_2 – varijanca uzroka 2

t-test: veliki zavisni uzorci

- situacija u kojoj imamo dva mjerenja na istim ispitanicima (npr. mjerimo VR skupini od 100 ispitanika u 2 navrata: 1. mjerenje – bez utjecaja alkohola; 2. mjerenje – pod utjecajem alkohola)
- s obzirom na to da u oba mjerenja sudjeluju *isti* ispitanici, za očekivati je da će izmjerene vrijednosti iz 1. i 2. mjerenja biti u **korelaciji,** tj. da će biti povezane – ispitanici koji brzo reagiraju na podražaj vjerojatno će imati kraće vrijeme reakcije u obe situacije mjerenja

t-test: veliki zavisni uzorci

- rezultati su u korelaciji kad:
  - isti ispitanici sudjeluju u eksperimentalnoj i kontrolnoj situaciji (2 situacije mjerenja)
  - kad su upotrijebljeni *ekvivalentni parovi*

t-test: veliki zavisni uzorci

$$t = \frac{M_1 - M_2}{\sqrt{{\sigma_{M1}}^2 + {\sigma_{M2}}^2 - 2r\sigma_{M1}\sigma_{M2}}}$$

df=N-1

M_1 – aritmetička 1. mjerenja
M_2 – aritmetička sredina 2. mjerenja
σ^2_{M1} – standardna pogreška aritmetičke sredine rezultata 1. mjerenja
σ^2_{M2} – standardna pogreška aritmetičke sredine rezultata 1. mjerenja
r – korelacija između rezultata 1. i 2. mjerenja

t-test: mali zavisni uzorci

- METODA DIFERENCIJE
  - isključuje potrebu računanja korelacije između dvije varijable
  - predstavlja pravi algoritam za provjeru razlike između aritmetičkih sredina zavisnih uzoraka (iako se ne koristi kod velikih uzoraka zbog kompliciranog izračuna – kod velikih zavisnih uzoraka umjesto toga se uvrštava koeficijent korelacije kao mjera kovariranja rezultata između dva mjerenja)
  - sastoji se u tome da se individualne razlike parova rezultata uzmu kao uzorak te se određuju aritmetička sredina, standardna devijacija i standardna pogreška

t-test: mali zavisni uzorci

› koraci u provedbi metode diferencije:

1. određivanje razlike *D* među rezultatima istog ispitanika u dvije situacije mjerenja, tj. određivanje individualnih razlika parova rezultata ($D=x_1-x_2$)
2. zbroj svih individualnih razlika, uzimajući u obzir i predznak, podijeljen brojem ispitanika, tj. parova, daje nam prosječnu razliku, tj. diferenciju između aritmetičkih sredina u dvije situacije mjerenja ($M_D=\sum D/N=M_1-M_2$)

t-test: mali zavisni uzorci

- koraci u provedbi metode diferencije:

3. Izračunavanje standardne devijacije razlika $\boldsymbol{\sigma_D}$
4. Izračunavanje standardne pogreške aritmetičke sredine odnosno standardne pogreške razlike:
   - s obzirom na to da je M_D zapravo *prava razlika*, standardna pogreška aritmetičke sredine zapravo je ***standardna pogreška razlike***:

$$\sigma_{M_D} = \frac{\sigma_D}{\sqrt{N}}$$

t-test: mali zavisni uzorci

› koraci u provedbi metode diferencije:

5. izračunavanje t-vrijednosti:

$$t = \frac{M_D}{\sigma_{M_D}}$$

df=N-1

12. Korelacija

Što je korelacija?

- zavisnost, povezanost, asocijacija između dvije pojave
- statistička definicija: *sukladnost u variranju dviju ili više varijabli*
- primjeri: povezanost između ekonomskog statusa i zdravstvenog stanja, povezanost između visine i težine u ljudi, povezanost između inteligencije i školskog uspjeha, povezanost između savjesnosti i školskog uspjeha itd.
- Karl Pearson (1857. – 1936.) – engleski matematičar koji je razradio računski postupak za izračunavanje stupnja povezanosti i izrazio stupanj povezanosti brojem, koji je nazvao **KOEFICIJENT KORELACIJE – r (Pearsonov r)**
- **r** je najčešće korišteni koeficijent korelacije

Smisao korelacije

- **DIJAGRAM RASPRŠENJA**
  - tablični ili grafički prikaz korelacije u koordinatnom sustavu
  - prikaz vezanog raspršenja u dvije varijable x i y
  - svaka statistička jedinica (tj. ispitanik) prikazana je jednom točkom s dvije koordinate:
    - rezultat u X varijabli
    - povezani rezultat u Y varijabli

Dijagram raspršenja

- CRTA ili PRAVAC REGRESIJE
  - crta koja aproksimira točke (koje predstavljaju pojedine statističke jedinice, tj. ispitanike) na jednu zajedničku liniju
  - spaja točke s koordinatama određenim FIKSNIM VRIJEDNOSTIMA u jednoj varijabli (X) i PARCIJALNIM SREDNJIM VRIJEDNOSTIMA u drugoj varijabli (Y)
  - pokazuje tip odnosa između varijabli:
    - pravac – linearna korelacija
    - crta – zakrivljena korelacija

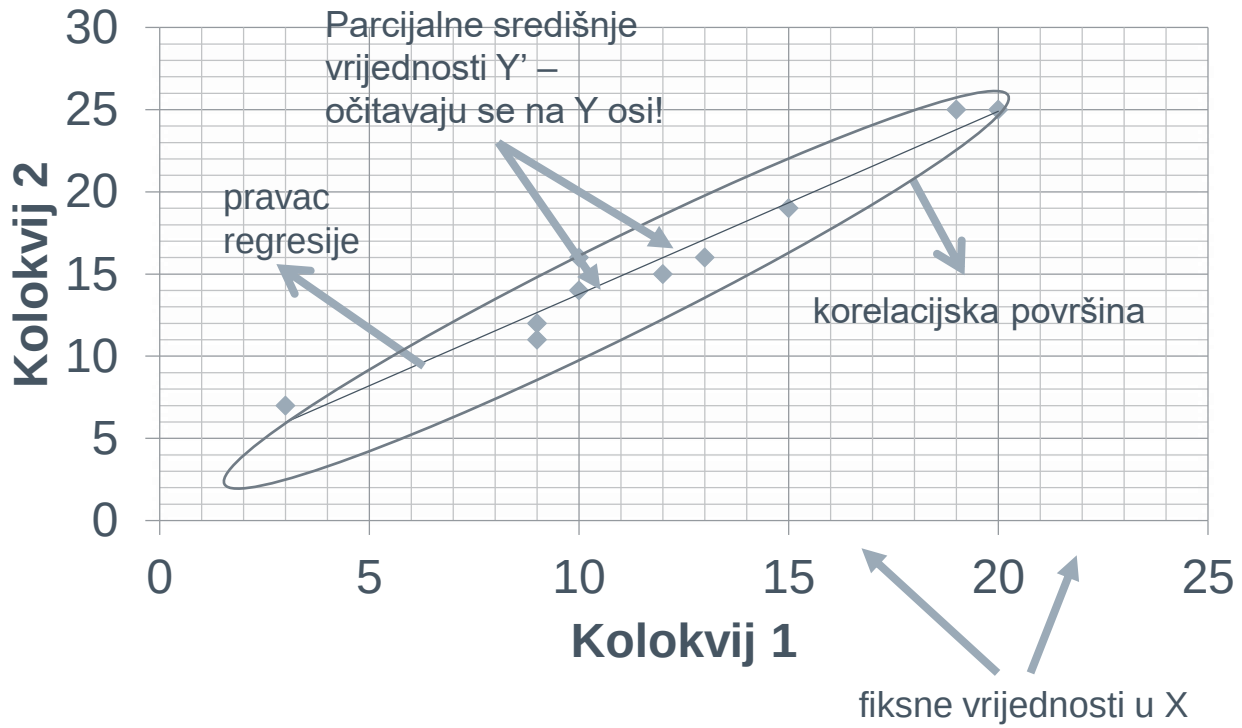
Dijagram raspršenja

- FIKSNE VRIJEDNOSTI
  - polazišne, arbitrarno ili namjerno odabrane vrijednosti u **varijabli X *(nezavisnoj, prediktorskoj),*** za koje se promatraju sve vrijednosti koje su uz njih vezane u drugoj ***varijabli Y (zavisnoj, kriterijskoj)***
- PARCIJALNE SREDNJE VRIJEDNOSTI
  - *aritmetička sredina* rezultata u jednoj varijabli vezanih za određenu fiksnu vrijednost u drugoj varijabli
  - sredina između svih vrijednosti u Y varijabli koje postižu različiti ispitanici za određenu vrijednost u X varijabli
  - raspršenje rezultata u Y varijabli oko parcijalne srednje vrijednosti (za neku fiksnu vrijednost u X), može se izraziti *standardnom devijacijom* (REZIDUALNI VARIJABILITET)

Dijagram raspršenja

- KORELACIJSKA POVRŠINA
  - „površina” koju u koordinatnom sustavu zauzimaju individualni rezultati pojedinih elemenata u dvije varijable, između kojih se računa korelacija
  - površina koja označava raspršenja pojedinih statističkih jedinica oko parcijalne srednje vrijednosti u Y za neku fiksnu vrijednost u X

Dijagram raspršenja



Smisao korelacije

- Potpuna korelacija ili funkcionalna veza
- Djelomična korelacija
- Korelacija = 0

Potpuna korelacija

- postoji kad u svakoj vrijednosti u jednoj varijabli X odgovara *samo jedna* vrijednost u drugoj varijabli Y
- ako linearnom ***porastu*** prve varijable odgovara linearni ***porast*** druge varijable, i to tako da je jedna određena vrijednost prve varijable uvijek povezana s korespondentnom vrijednošću druge varijable, korelacija je POTPUNA i POZITIVNA: **r=+1**
- Razmislite...
  - Kako će izgledati pravac regresije i korelacijska površina?

Potpuna korelacija

- ako linearnom ***porastu*** prve varijable odgovara linearni ***pad*** druge varijable, i to tako da je jedna određena vrijednost prve varijable povezana s jednom korespondentnom vrijednošću druge varijable, korelacija je POTPUNA I NEGATIVNA: **r=-1**
- Razmislite...
  - Kako će tada izgledati pravac regresije i korelacijska površina?

Djelomična korelacija

- postoji kovarijabilitet kod promatranih varijabli, ali određenoj vrijednosti varijable X odgovara *više* različitih vrijednosti varijable Y
- ako linearnom ***porastu*** prve varijable uglavnom odgovara linearni ***porast*** druge varijable, i to tako da je jedna određena vrijednost prve varijable povezana s više vrijednosti druge varijable, korelacija je DJELOMIČNA I POZITIVNA: **0<r<1**
- Razmislite...
  - Kako će izgledati pravac regresije i korelacijska površina?

Djelomična korelacija

- ako linearnom porastu prve varijable uglavnom odgovara linearni pad druge varijable, i to tako da je jedna određena vrijednost prve varijable povezana s više vrijednosti druge varijable, korelacija je DJELOMIČNA i NEGATIVNA: **-1<r<0**
- Razmislite...
  - Kako će izgledati pravac regresije i korelacijska površina?

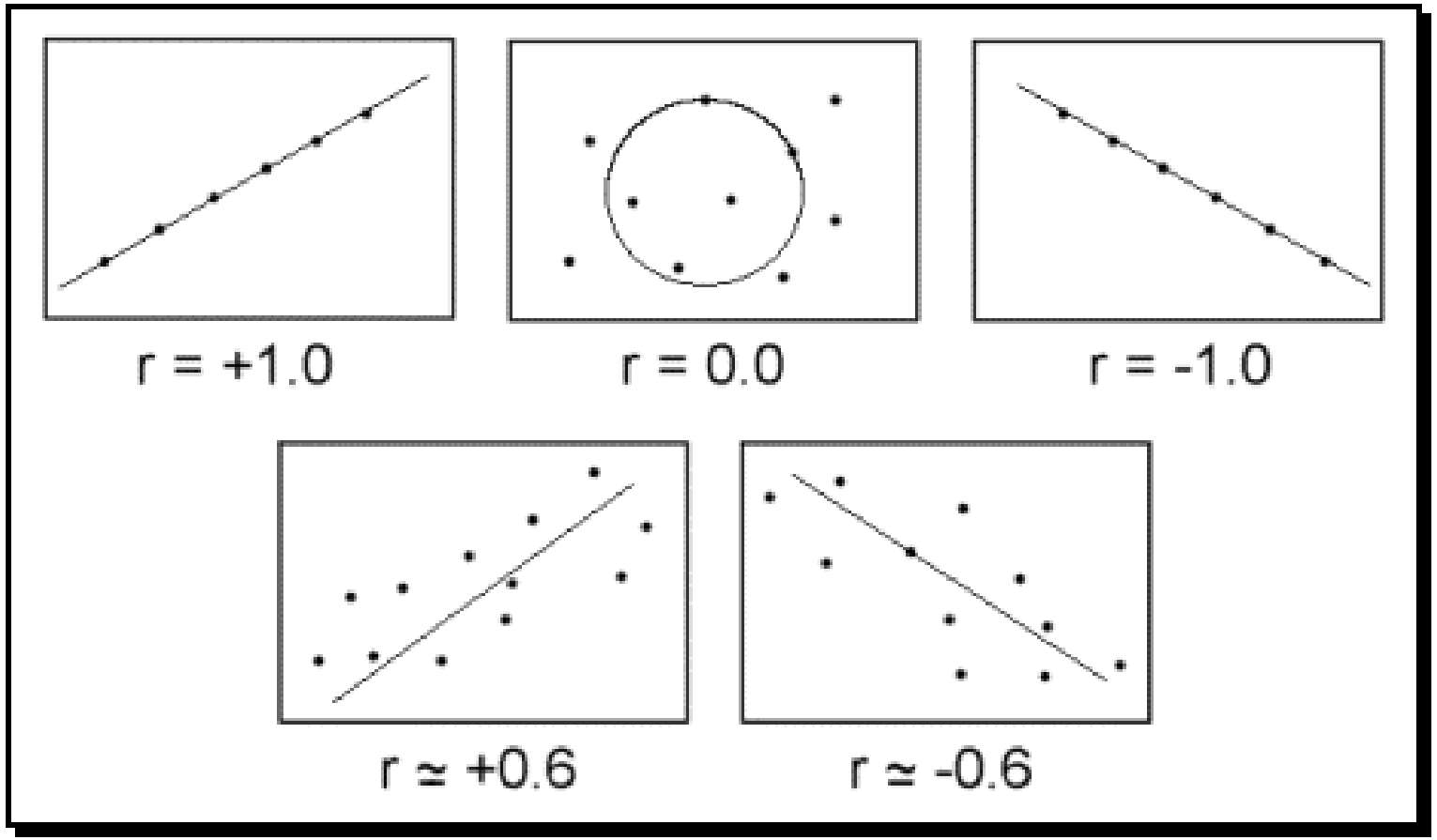
Korelacija = 0

- svakoj vrijednosti u X varijabli mogu odgovarati *sve* vrijednosti u Y varijabli
- jednoj određenoj vrijednosti X varijable odgovara ***bilo koja*** (od mogućih) vrijednosti druge varijable: **r=0**
- među varijablama postoji samo SLUČAJAN varijabilitet, tj. nema kovariranja → promjene u jednoj varijabli odvijaju se nezavisno od promjena u drugoj varijabli
- Razmislite...
  - Kako će tada izgledati pravac regresije i korelacijska površina?

Izgled korelacijske površine i stupanj povezanosti

- ako je korelacija potpuna, korelacijska površina stopit će se u *crtu* regresije jer će svi ispitanici koji su postigli neku vrijednost u X varijabli postići samo jedan, isti korespondentan rezultat u Y varijabli
- ako korelacije nema, tj. ona iznosi 0, korelacijska površina pretvorit će se u *krug* – raspršenje točaka oko parcijalne srednje vrijednosti u Y varijabli bit će maksimalno za svaku vrijednost u X varijabli
- što je korelacija veća, tj. što se više udaljuje od nulte, krug više nalikuje *elipsi* koja je to uža što je korelacija veća – raspršenja točaka oko parcijalne srednje vrijednosti se smanjuju

Dijagram raspršenja kad je korelacija: potpuna i pozitivna (r=+1.0), nulta (r=0.0), potpuna i negativna (r=-1.0), umjerena i pozitivna (r=+0.6) i umjerena i negativna (r=-0.6)



Preuzeto s: https://allpsych.com/researchmethods/correlation/

Rezidualni varijabilitet

- varijabilitet rezultata koji je *preostao* u varijabli Y kad se promatra povezanost s rezultatima u X
- što je korelacija između X i Y varijable veća, rezidualni varijabilitet je manji jer se točke bolje aproksimiraju na crtu regresije
- Razmislite...
  - Koliko će iznositi rezidualni varijabilitet ako je r=0, odnosno ako je r=1?

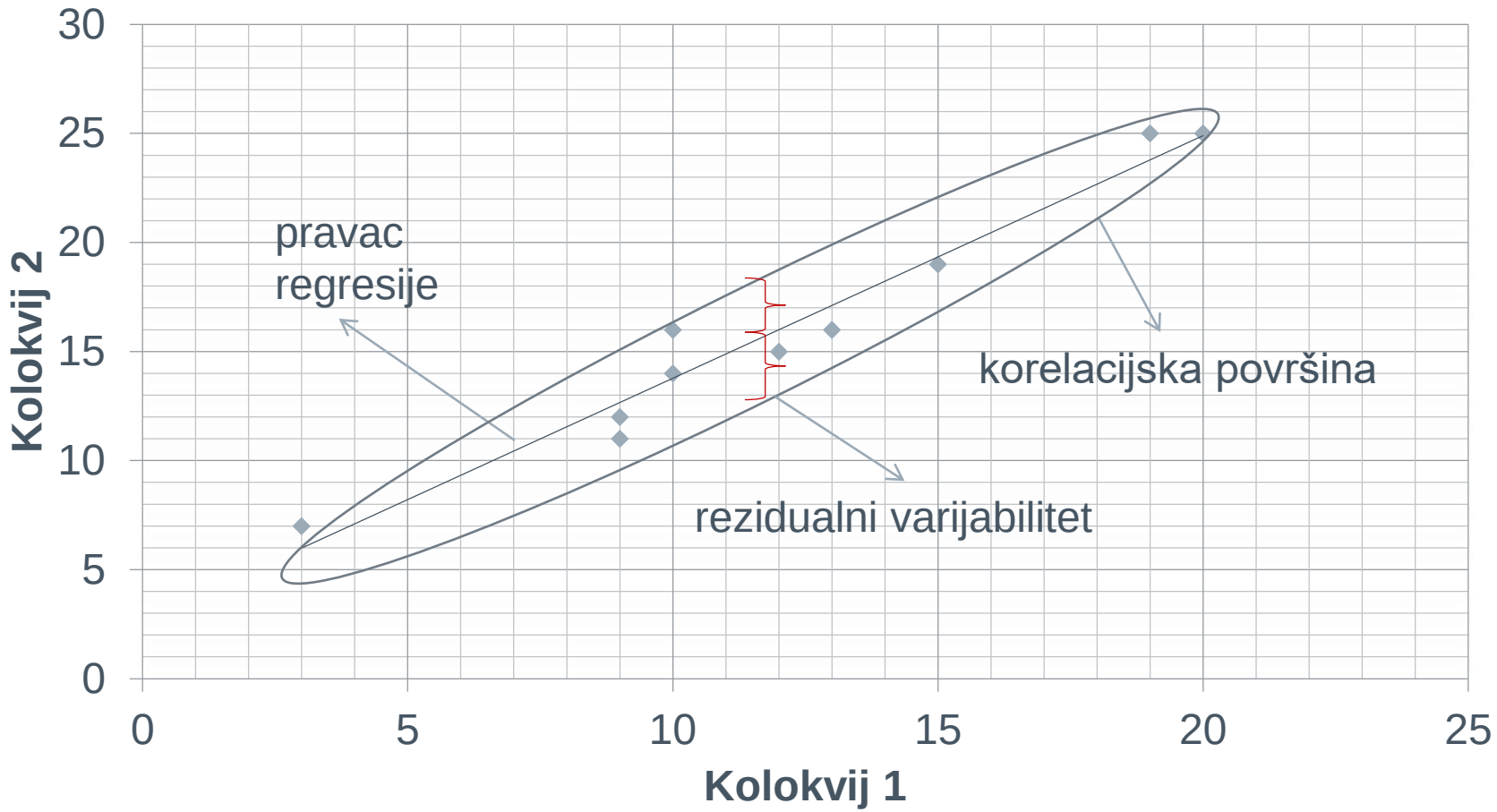
Rezidualni varijabilitet

- ako je korelacija potpuna, sve točke prianjaju na crtu regresije pa je rezidualni varijabilitet = 0
- ako nema korelacije, točke će maksimalno odstupati od crte regresije pa će rezidualni varijabilitet biti maksimalan, tj. jednak standardnoj devijaciji vrijednosti u Y varijabli
- što točke više odstupaju od crte regresije, to je rezidualni varijabilitet veći

Rezidualni varijabilitet

- izražavamo ga KOEFICIJENTOM REZIDUALNOG VARIJABILITETA
  - standardna devijacija rezultata u Y varijabli koji se vežu za jednu određenu (fiksnu) vrijednost u X varijabli
  - što je koeficijent rezidualnog varijabiliteta veći, to je koeficijent korelacije manji
  - naziva se još i STANDARDNA POGREŠKA PROGNOZE

Dijagram raspršenja



Linearna korelacija

- *podjednake* promjene vrijednosti u jednoj varijabli praćene su *podjednakim* promjenama vrijednosti u drugoj varijabli
- crta regresije u dijagramu raspršenja je PRAVAC
- načelno, za svaku korelaciju možemo odrediti dva pravca regresije:
  1. na apscisu nanosimo vrijednosti A varijable, a na ordinatu vrijednosti B varijable
  2. na apscisu nanosimo vrijednosti B varijable, a na ordinatu vrijednosti A varijable
- određivanje dva pravca regresije za jednu korelaciju ima svoj smisao samo ako obje varijable mogu biti i ***prediktorske*** i ***kriterijske*** (npr. korelacija između godina školovanja i verbalnih sposobnosti)

Zakrivljena korelacija

- povezanost koja nije linearna (npr. korelacija između intenziteta rasvjete i radnog učinka u nekom preciznom poslu)
- može biti dvojaka:

1. zakrivljena povezanost u kojoj crta regresije *ne* mijenja svoj smjer
2. zakrivljena povezanost u kojoj crta regresije mijenja svoj smjer

- nije dopušteno izračunavati Pearsonov r!
- prije odluke o izračunu određenog koeficijenta korelacije, korisno je grafički prikazati dobivene rezultate pomoću dijagrama raspršenja

Korelacija i kauzalnost

- ako su neke dvije varijable u korelaciji, logično je za pretpostaviti da postoji neki uzrok njihovog kovariranja
- ali, korelacija nam sama po sebi ne govori ništa o smjeru kauzalnosti (što je uzrok, a što posljedica)!
- Mogući smjerovi kauzalnosti:

1. **X uzrokuje Y** (npr. varijacije u školskom uspjehu uzrokuju varijacije u samopoštovanju)
2. **Y uzrokuje X** (npr. varijacije u samopoštovanju uzrokuju varijacije u školskom uspjehu)
3. **neki treći faktor uzrokuje i X i Y** (tj. sukladnost u njihovu variranju) (npr. varijacije u dječjoj percepciji prihvaćanja od strane majke uzrokuju varijacije i u školskom uspjehu i u samopoštovanju)
4. **X uzrokuje Y, i Y uzrokuje X** (npr. varijacije u školskom uspjehu uzrokuju varijacije u samopoštovanju; a varijacije u samopoštovanju uzrokuju varijacije u školskom uspjehu)

Pearsonov koeficijent korelacije

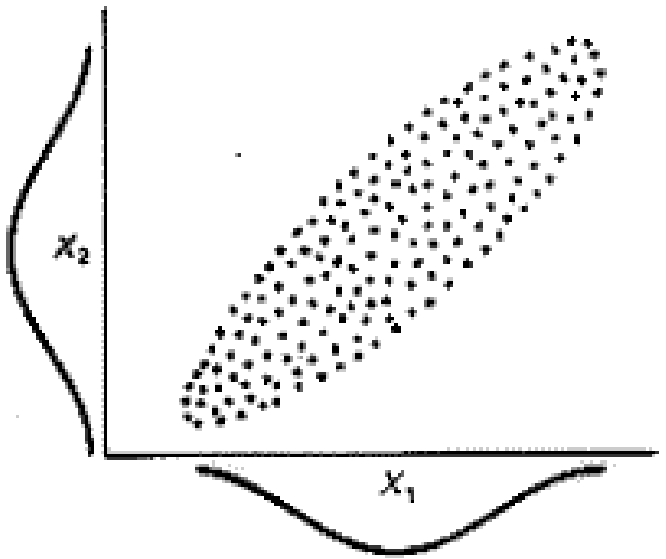
- koeficijent korelacije umnožaka
- „product-moment correlation”
- najčešće korišten koeficijent korelacije

Uvjeti za izračunavanje Pearsonovog r

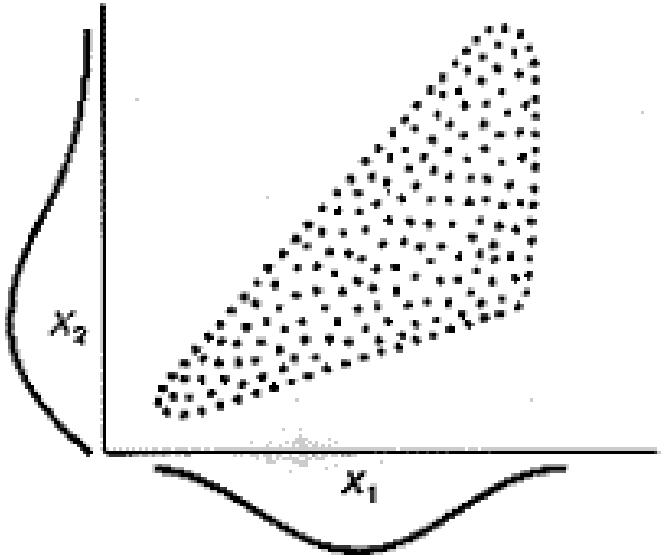
1. rezultati u obje varijable moraju biti prave numeričke vrijednosti, tj. moraju biti izraženi barem na **intervalnoj skali**
2. ako je povezanost među varijablama **linearna** (ako je povezanost zakrivljena, r može iznositi 0, čak i ako povezanost uistinu postoji) – kontrola: grafički prikaz dijagrama raspršenja i pravca regresije
3. ako su distribucije rezultata u varijablama normalne ili barem **simetrične** – asimetričnost distribucije može uzrokovati zakrivljenost korelacije kad nije opravdana upotreba r (kontrola: oblik distribucije rezultata u svakoj varijabli!)
4. zadovoljena pretpostavka o **HOMOSCEDASCITETU,** tj. o homogenosti u variranju

- HOMOSCEDASCITET
  - podjednako variranje vrijednosti u Y varijabli za bilo koju, tj. sve vrijednosti u X varijabli
  - teško ga je kontrolirati pa se obično to i ne čini, već se samo pretpostavlja da nije u znatnijoj mjeri narušen
  - kontrola: dijagram raspršenja

Prikaz homoscedasciteta i heterodasciteta u variranju varijabli koje kovariraju:



Homoscedasticity with both variables normally distributed



Heteroscedasticity with skewness on one variable

Preuzeto s: http://bobhall.tamu.edu/epsy640/Tools/CorrelationInterpretation.html

Racionala za izračunavanje r-a

- linearna korelacija postoji kada su niski rezultati na jednoj varijabli (X) praćeni niskim rezultatima na drugoj varijabli (Y) odnosno visoki rezultati na X varijabli su praćeni visokim rezultatima na Y varijabli
- osnova za određivanje korelacije: konzistentna usporedba rezultata koje ispitanici postižu na dvije različite varijable
- kada bi korelacija bila linearna i potpuna, svaki ispitanik bio bi u obje varijable na jednakim ***mjestima*** u skupu rezultata – koliko je ispitanik A u varijabli X udaljen od prosjeka, upravo toliko je i u varijabli Y udaljen od prosjeka
- Razmislite...
  - Kako ćemo odrediti položaj pojedinih rezultata?

Racionala za izračunavanje r-a

- odgovor: uz pomoć **z-vrijednosti**
- z-vrijednost nekog rezultata pokazuje nam koliko je on udaljen od M u terminima standardne devijacije
- z-vrijednosti jedne varijable direktno su komparabilne sa z-vrijednostima druge varijable bez obzira na vrijednosti njihovih aritmetičkih sredina i standardnih devijacija (tj. na skalu na kojima su rezultati izraženi)

Racionala za izračunavanje r-a

- Npr. ako je **r=+1**, svaki ispitanik koji je postigao z=0.5 u X varijabli imat će rezultat z=0.5 i u drugoj varijabli
- a što ako je **r=-1**?
- što je apsolutna veličina r manja, z-vrijednosti za iste ispitanike u dvije varijable će se više međusobno razlikovati
- ako je **r=0**, razlike između pojedinačnih z-vrijednosti bit će *slučajne*

Racionala za izračunavanje r-a

- ako za svakog ispitanika napravimo umnožak z-vrijednosti dobivenih na X varijabli i z-vrijednosti dobivenih na Y varijabli dobivamo uvid u smjer i veličinu korelacije
- dakle, visinu korelacije možemo promatrati i kao prosječan umnožak z-vrijednosti dviju varijabli X i Y:
  - kad je r=+1, $z_x=z_y$, suma umnožaka je maksimalna i pozitivna ($-z_x \cdot(-z_y)=z_xz_y$; $z_x \cdot z_y=z_xz_y$)
  - kad je r=-1, $z_x=-z_y$, suma umnožaka je također maksimalna, ali je predznak negativan ($-z_x \cdot z_y=-z_xz_y$; $z_x \cdot(-z_y)=-z_xz_y$)
  - ako je r=0, $z_x \neq z_y$, prosječna suma umnožaka z_xz_y iznosi 0 (negativni i pozitivni umnošci samo su slučajni pa je njihova M=0)

Racionala za izračunavanje r-a

- dakle, ako je suma umnožaka z-vrijednosti na X i Y varijabli velik i pozitivan broj, korelacija je pozitivna
- ako je suma umnožaka velik i negativan broj, korelacija je negativna
- ako se vrijednost sume umnožaka kreće oko 0, korelacije nema
- ali ostaje pitanje: koliko je neka korelacija zapravo velika odnosno mala, tj. koliko iznosi snaga povezanosti između dvije varijable?

Formula za izračunavanje **Pearsonovog koeficijenta korelacije r**

› odgovor na prethodno pitanje daje određivanje *prosječnog umnoška z-vrijednosti dviju varijabli*

$$r = \frac{\Sigma(z_x \cdot z_y)}{N}$$

zx – z-vrijednost u varijabli X
zy – z-vrijednost u varijabli y
N – ukupan broj rezultata

Moguće vrijednosti Pearsonovog r-a: od -1.0 do +1.0

Testiranje statističke značajnosti r

- prije interpretacije izračunatog r-a moramo testirati njegovu statističku značajnost, odnosno utvrditi postoji li povezanost u populaciji
- razlog: povezanost među varijablama gotovo uvijek izračunavamo na *uzorcima,* pa je dobivena vrijednost koeficijenta korelacije samo procjena *prave* povezanosti u populaciji

Nul-hipoteza

- H_0: nema povezanosti među varijablama, tj. povezanosti između varijabli iznosi 0
- ako je H_0 točna, varijabilitet u Y varijabli ne može se objasniti varijabilitetom u X varijabli jer su oba varijabiliteta međusobno nezavisna, tj. slučajna
- ako H_0 nije točna, postoji povezanost među varijablama – promjene u Y možemo objasniti promjenama u X

Testiranje statističke značajnosti r

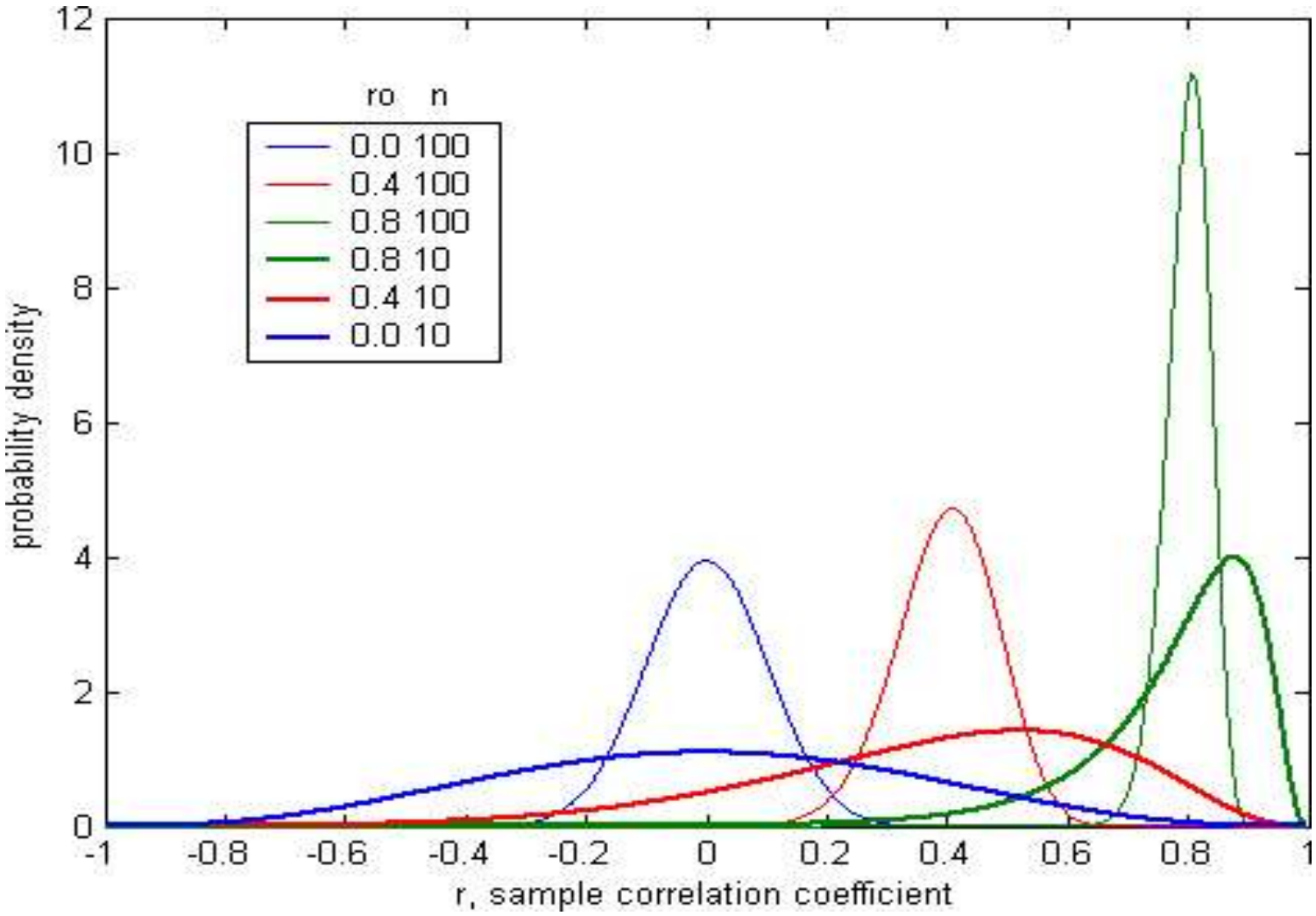
- **standardna pogreška koeficijenta korelacije r**
  - pogreška kojoj se izlažemo kada iz koeficijenta dobivenog na uzorku zaključujemo o povezanosti u populaciji
  - standardna devijacija distribucije koeficijenata korelacije velikog broja uzoraka iste veličine oko *pravog* koeficijenta korelacije
  - može se izračunati pomoću formule:

$$\sigma r = \frac{1}{\sqrt{N}}$$

Testiranje statističke značajnosti r

- ako nam je poznata vrijednost standardne pogreške koeficijenta korelacije, mogli bismo pomoću *granica pouzdanosti* odrediti je li neki izračunati r statistički značajan
- problem: koeficijenti korelacije velikog broja uzoraka određene veličine uglavnom se ne distribuiraju po normalnoj raspodjeli
- distribucija koeficijenata korelacije normalna je u samo jednom slučaju: kad je korelacija u populaciji 0!
- izgled distribucije koeficijenata korelacije r, dobivena na velikom broju uzoraka, ovisi o ***stvarnoj korelaciji u populaciji*** i o ***veličini uzorka,*** tj. parova

Distribucija koeficijenta korelacije r



Izgled distribucije ovisi o veličini uzorka i veličini povezanosti.

Preuzeto s: http://www.alivelearn.net/?p=709

Testiranje statističke značajnosti r

- zbog asimetričnosti distribucije, statističku značajnost koeficijenta korelacije utvrđujemo na dva načina:

1. izračunavanjem t-testa

$$t = r\sqrt{\frac{N-2}{1-r^2}} \qquad df = N-2$$

2. pomoću posebnih tablica koje pokazuju koliko najmanje mora iznositi visina neke korelacije, da bismo je – uz zadanu veličinu uzorka – mogli smatrati statistički značajnom na određenoj razini značajnosti
   - tablica je dvosmjerna što znači da pokazuje graničnu vrijednost r bez obzira na predznak

Interpretacija koeficijenta korelacije r

- visina korelacije (koja se kreće od -1 preko 0 do +1) ukazuje na sukladnost u variranju dvije varijable
- sukladnost u variranju događa se zbog djelovanja nekih ZAJEDNIČKIH FAKTORA dviju varijabli

Interpretacija koeficijenta korelacije r

- KOEFICIJENT DETERMINACIJE: $\mathbf{r_{xy}^2}$
  - postotak *zajedničke varijance dviju varijabli*, odnosno postotak svih faktora koji su odgovorni za dobiveni stupanj sukladnosti u variranju rezultata varijabli X i Y
  - određuje se kvadriranjem koeficijenta korelacije: $\mathbf{r_{xy}^2}$ (npr. r=0.30, r_{xy}^2=0.09 ili 9 % zajedničke varijance; r=0.6, r_{xy}^2=0.36 ili 36 % zajedničke varijance; dakle, povezanost r=0.6 je zapravo *4 puta* snažnija od povezanosti r=0.3!)
  - može se još interpretirati i kao **postotak objašnjene varijance** ili **postotak redukcije pogreške**

Interpretacija koeficijenta korelacije r

- koeficijent determinacije raste ubrzano, tj. geometrijskom progresijom s linearnim porastom veličine koeficijenta korelacije:
  - r=0.2 → 4 % zajedničkih faktora
  - r=0.3 → 9 % zajedničkih faktora
  - r=0.4 → 16 % zajedničkih faktora
  - r=0.5 → 25 % zajedničkih faktora

Interpretacija koeficijenta korelacije r

- odluka o tome je li neki izračunati koeficijent korelacije visok ili nizak ovisi prvenstveno o samoj prirodi varijabli, načinu njihova mjerenja, ali i konceptualnoj pozadini istraživanja (npr. korelacija inteligencije i školskog uspjeha koja iznosi r=0.6 je očekivane visine ako su nam ispitanici djeca nižih razreda OŠ, ali istu vrijednost smatramo visokom ako su nam ispitanici studenti)

13. Hi-kvadrat

Hi-kvadrat test χ^2

- jedan od najpoznatijih *neparametrijskih* testova
- koristi se kod **kategorijalnih podataka (NOMINALNA SKALA)** – podatci su *frekvencije* opažanja u pojedinim kategorijama na nekoj varijabli (npr. broj ljudi koji ima odnosno nema psa)
- Bit hi-kvadrata: *opažene* frekvencije uspoređujemo s *teoretskim* frekvencijama koje bi se očekivale pod nekom određenom hipotezom
- nul-hipoteza: nema razlike između opaženih i teoretskih frekvencija
- Primjer istraživačkog problema: postoje li razlike između muškaraca i žena u obolijevanju od neke bolesti?

Upotreba χ^2 testa

1. kada imamo frekvencije ***jednog uzorka*** pa želimo odrediti odstupaju li one od neke teoretske raspodjele frekvencija
2. kada imamo frekvencije ***dva ili više nezavisnih uzoraka*** pa želimo ustanoviti razlikuju li se uzorci međusobno u opaženim svojstvima
3. kada imamo frekvencije ***jednog uzoraka***, koji ima dihotomna svojstva, u ***dvije situacije*** mjerenja pa želimo ustanoviti razlikuju li se ispitanici u opaženim svojstvima kroz dvije situacije mjerenja

Hi-kvadrat test χ^2

- mjera odstupanja opaženih od teoretskih frekvencija
- koraci određivanja χ^2:
  1. određivanje teoretskih frekvencija za svaku kategoriju podataka
  2. pronalaženje razlika između svake opažene i pripadne teoretske frekvencije
  3. kvadriranje svake takve razlike te dijeljenje svakog kvadrata razlike s pripadnom teoretskom frekvencijom
  4. sumiranje tih vrijednosti za svaku kategoriju

Hi-kvadrat test χ^2

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}$$

df=k-1 (jedna varijabla)
df=(k1-1)(k2-1) (dvije varijable)

f_o – opažene frekvencije
f_t – teoretske frekvencije
k – broj kategorija

O čemu je važno voditi računa kod primjene χ^2 testa?

1. podatci kojima raspolažemo trebaju biti pravi *brojeni* podatci, tj. podatci koji se mogu svesti na *frekvencije* – u račun se ne unose mjerene vrijednosti!
2. individualni podatci koji su prema određenom načelu klasifikacije raspoređeni u pojedine kategorije smiju se u klasifikaciji pojaviti samo *jednom*
3. smije se računati samo iz frekvencija, ali ne i postotaka ili proporcija
4. smije se računati samo kad ni jedna od *teoretskih* frekvencija nije manja od 5

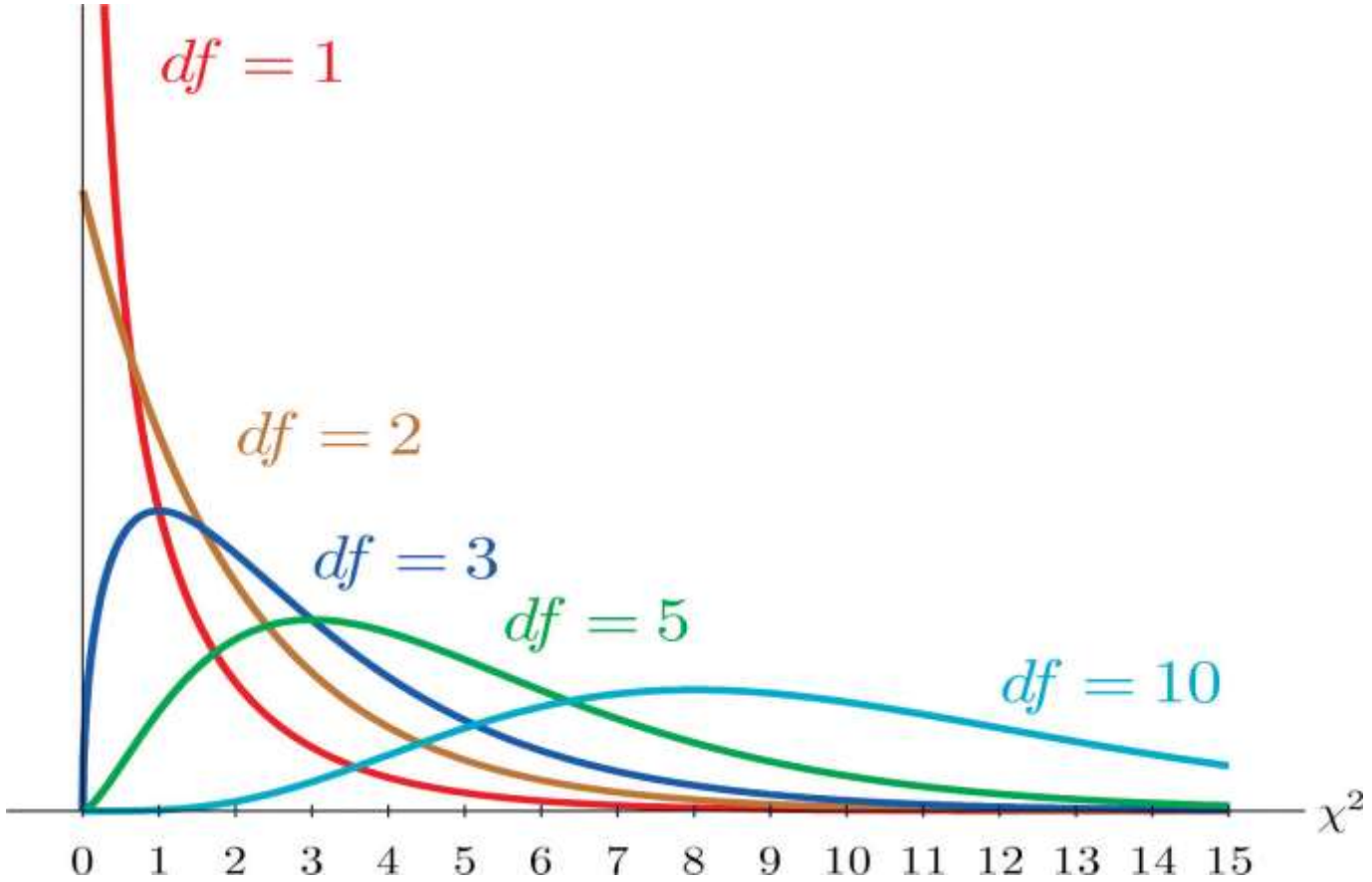
Interpretacija χ^2

- Razmislite...
  - Kad ne bi bilo nikakve razlike između opaženih i teoretskih frekvencija, koliko bi iznosio χ^2?
  - Bi li nul-hipoteza tada bila točna?
- što su veće razlike između opaženih i teoretskih frekvencija, to je χ^2 veći i veća je vjerojatnost odbacivanja nul-hipoteze
- Ali koliko χ^2 mora iznositi da bismo mogli odbaciti nul-hipotezu?

Distribucija χ^2

- raspodjela svih mogućih χ^2 vrijednosti koje se mogu dobiti neizmjerno velikim brojem mjerenja uz određeni stupanj slobode
- što je manje stupnjeva slobode, distribucija je više **pozitivno asimetrična**
- one vrijednosti χ^2 koje su toliko velike da je njihovo slučajno pojavljivanje moguće samo u 1 % ili 5 % slučajeva (desni kraj distribucije!), možemo smatrati statistički značajnim

Distribucija χ^2



Izgled distribucije ovisi o broju stupnjeva slobode.

https://2012books.lardbucket.org/books/beginning-statistics/s15-chi-square-tests-and-f-tests.html

Distribucija χ^2

- praktično pravilo: centralna vrijednost hi-kvadrat distribucije, uz neki stupanj slobode, iznosi otprilike toliko koliko imamo stupnjeva slobode – nul-hipotezu možemo sigurno prihvatiti (i bez uvida u tablicu), ako je dobiveni χ^2 *manji ili jednak* broju stupnjeva slobode

Tablica graničnih vrijednosti χ^2

- uz određeni broj stupnjeva slobode i određenu razinu rizika pokazuje koliko *najmanje* mora iznositi vrijednost χ^2 da odbacimo nul-hipotezu
- granične vrijednosti χ^2 u tablici odnose se na vrijednosti *dvosmjernog testiranja* (iako mi promatramo desni kraj krivulje) – smjer razlike nije važan jer se razlike između f_o i f_t u računu kvadriraju

χ^2: jedan uzorak

- pitanje: razlikuju li se opažene frekvencije ispitanika u pojedinim kategorijama od očekivanih prema nekoj hipotezi – pri tome su svi ispitanici članovi *istog uzorka* (npr. razlikuju li se omjer muškaraca i žena na nekom fakultetu (uzorak) od omjera koji bi se očekivao na temelju podataka o općoj populaciji)

χ^2: dva ili više nezavisnih uzoraka

- kada imamo podatke koji su kategorizirani prema dvije ili više varijabli – pitanje njihove međusobne nezavisnosti
- **tablica kontingencije**
  - pokazuje distribuciju jedne varijable na svakoj razini druge varijable
  - drugim riječima, pokazuje je li distribucija frekvencija unutar kategorija jedne varijable „povezana” s distribucijom frekvencija unutar kategorija druge varijable

χ^2: dva ili više nezavisnih uzoraka

- **Primjer:** Istraživače je zanimalo razlikuje li se broj osuđenika na smrt u Sj. Karolini (u četverogodišnjem periodu) ovisno o rasi osuđenika. Na uzorku od 825 ispitanika utvrđeno je da su 33 bijelca (od njih ukupno 541) osuđena na smrt. Jednako toliko osoba crne rase bilo je osuđeno na smrt (Howell, 2002).

	Smrtna presuda			
Rasna pripadnost		Da	Ne	Ukupno
	Crnci	33	251	284
	Bijelci	33	508	541
	Ukupno	66	759	825

χ^2: dva ili više nezavisnih uzoraka

- **Primjer**: ako nema značajne razlike između bijelaca i crnaca, proporcija smrtne presude morala bi biti jednaka kod obje rase
- koliko će u tom slučaju iznositi očekivane frekvencije?
  - ***očekivane frekvencije*** **dobivamo tako da pomnožimo sumu reda sa sumom stupca i rezultat podijelimo s totalnom sumom frekvencija**

Primjer:

	Smrtna presuda			
Rasna pripadnost		Da	Ne	Ukupno
	Crna rasa	33 (**22.72**)	251 (**261.28**)	284
	Bijela rasa	33 (**43.28**)	508 (**497.72**)	541
	Ukupno	66	759	825

284x66/825=22.72
541x66/825=43.28
284x759/825=261.28
541x759=825

Yatesova korekcija

- uvijek kada imamo 2x2 tablicu (ali i druge oblike tablica *ako je* $f_t<5$), potrebno je provesti **Yatesovu korekciju**
  - svaka opažena frekvencija koja je *veća* od očekivane treba se smanjiti za 0.5, a svaka opažena frekvencija koja je *manja* od očekivane treba se povećati za 0.5, tj. ***svaka se apsolutna razlika između očekivane i opažene frekvencije smanji za 0.5***
- Kakve će posljedice Yatesova korekcija imati za veličinu χ^2 testa? Kojoj hipotezi ide u prilog?
- Yatesova korekcija gubi svoj smisao ako su 1) razlike između f_t i f_o manje od 0.25 jer u tom slučaju oduzimanje vrijednosti 0.5, dovodi do povećanja vrijednosti χ^2, i 2) razlike između f_o i f_t jako velike

Primjer:

› Yatesova korekcija: razliku f_o-f_t umanjujemo za 0.5

f_o	f_t	f_o-f_t	$\lvert f_o-f_t\rvert-0.5$	$(\lvert f_o-f_t\rvert-0.5)^2$	$(\lvert f_o-f_t\rvert-0.5)^2/f_t$
33	22.72	10.28	9.78	95.65	4.21
33	43.28	-10.28	9.78	95.65	2.21
251	261.28	-10.28	9.78	95.65	0.37
508	497.72	10.28	9.78	95.65	0.19
					$\chi^2=6.98$

Primjer:

- određivanje stupnjeva slobode: df=(broj stupaca -1)·(broj redaka -1)
- $\chi^2=6.98$, df=1, $p<0.01$
- Kako glasi interpretacija ovog izraza?
- Postoji statistički značajna razlika u udjelu osuđenih na smrt ovisno o njihovoj rasi uz nivo rizika manji od 1 % – pripadnici crne rase su osuđivani na smrt češće od pripadnika bijele rase.

χ^2: dva ili više nezavisnih uzoraka

- ako imamo više skupina ispitanika od dvije u nekoj varijabli (tj. retku ili stupcu), i računanjem χ^2 testa dokažemo statistički značajnu razliku, χ^2 nam ne govori o tome **među kojim** podskupinama je razlika značajna – tada možemo izračunati χ^2 *samo* za podskupine koje nas zanimaju

χ^2: zavisni uzorci – McNemarov test

- usporedba rezultata iste skupine ispitanika dobivenih u dvije različite situacije mjerenja (npr. prije i poslije)
- između rezultata prvog i drugog mjerenja postoji *korelacija*

χ^2: zavisni uzorci – McNemarov test

- **Primjer:** 100 ispitanika ispitano je testom 1 i testom 2. Postoji li statistički značajna razlika u uspješnosti ispitanika u testu 1 i testu 2? (Petz, Kolesarić i Ivanec, 2012):

	Test 2		
Test 1		Pali	Prošli
	Prošli	5 A	55 B
	Pali	25 C	15 D

- Razlike između 1. i 2. testa vide se A i D kućicama: A+D=totalni broj onih kod kojih se ne slaže uspjeh u prvom i drugom mjerenju (u B i C kućicama postoji poklapanje rezultata)

χ^2: zavisni uzorci – McNemarov test

- Nul-hipoteza: očekivana frekvencija u A i D ćeliji iznosi ½·(A+D) – pola slučajeva promijenilo se u jednom, a pola u drugom smjeru
- u ovom primjeru, mi zapravo testiramo statističku značajnost razlike između dvije proporcije:
  - p_1=(A+B)/N (proporcija onih koji su prošli u testu 1)
  - p_2=(B+D)/N (proporcija onih koji su prošli u testu 2)

χ^2: zavisni uzorci – McNemarov test

- Formula za McNemarov test:

$$\chi^2=(A - D)^2/(A + D)$$

- Yatesova korekcija:

$$\chi^2=(|A - D| - 1)^2/(A + D)$$

$$df=(\text{broj redaka}-1)\cdot(\text{broj stupaca}-1)$$

Primjer:

- $\chi^2=4.05$, df=1, $p<0.05$
- Postoji statistički značajna razlika u uspješnosti u dva testa – drugi test prošlo je više ispitanika.

χ^2: zavisni uzorci – McNemarov test

- potrebno je paziti na smisao kućica A i D: one predstavljaju one ispitanike koji su se *promijenili* – ako je tablica formirana drugačije, formule je potrebno prilagoditi!
- ako su očekivane frekvencije u kućicama A i D manje od 5, ovaj test se ne može primijeniti

Osnovni uvjeti za upotrebu hi-kvadrat testa (Petz, Kolesarić i Ivanec, 2012)

1. χ^2 se može računati samo s frekvencijama
2. suma očekivanih frekvencija mora biti jednaka sumi opaženih frekvencija
3. frekvencije u pojedinim kućicama moraju biti nezavisne – svaka frekvencija u pojedinoj ćeliji mora pripadati drugom ispitaniku
4. nijedna teoretska frekvencija ne smije biti premalena (<5)
5. kad je df=1, potrebno je provesti Yatesovu korekciju

14. Statistička snaga istraživanja i veličina učinka

Veličina efekta ili učinka

- **veličina učinka** (engl. *effect size*) = mjera *veličine razlike* između aritmetičkih sredina dviju populacija ili *veličine povezanosti* među varijablama
- prilikom usporedbe dviju aritmetičkih sredina, veličina učinka ukazuje nam na stupanj preklapanja distribucija dviju populacija – odnosno stupanj njihove „razdvojenosti” koji se može pripisati djelovanju eksperimentalne procedure
- što je preklapanje veće, to je veličina učinka manja

Veličina učinka

- test statističke značajnosti sam po sebi ne daje puno informacija o *veličini* razlike između neke dvije aritmetičke sredine, tj. populacije, ali su ova dva pojma ipak povezana – kako?
- razina rizika određuje naš stupanj sigurnosti – što je rizik manji, sigurniji smo da je razlika statistički značajna
- ali odbacivanje nul-hipoteze uz manju razinu rizika ne znači nužno da je razlika između dvije M veća jer to ovisi i o *broju ispitanika*
- Npr. ako je dobivena razlika statistički značajna uz $p<0.05$ u istraživanju na samo 20 ispitanika, učinak može biti veći nego u istraživanju na 1000 ispitanika gdje je ista razlika značajna uz $p<0.01$

Veličina učinka

- Publication Manual of the American Psychological Association (2009) – standardi prikazivanja rezultata psihologijskih istraživanja;
  - preporuka: uz rezultate testa statističke značajnosti potrebno je prikazati i mjeru veličine učinka

Određivanje veličine učinka

- veličina učinka izražena u mjernim jedinicama ZV – *veličina učinka izračunata na sirovim podacima*
- problem ovako određene vrijednosti jest nemogućnost usporedbe veličine učinka dobivenih u različitim istraživanjima koji su istu ZV operacionalizirala na različit način
- rješenje: standardizacija, tj. dijeljenje veličine učinka izračunate na sirovim podatcima s pripadajućom standardnom devijacijom populacije – *standardizirana veličina učinka*

Cohenov d: t-test za jedan uzorak

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

d – simbol za veličinu učinka
μ_1 – aritmetička sredina Populacije 1 (skupina koja prima eksperimentalnu manipulaciju)
μ2 – aritmetička sredina Populacije 2 (kontrolna skupina)
σ – standardna devijacija populacije

Cohenov d: t-test za nezavisne uzorke

- koju vrijednost **σ** ćemo uvrstiti u formulu za Cohenov d?
- ili bilo koju od dvije koje su nam na raspolaganju (situacija u kojoj testiramo razliku između M-ova dobivenih na dva uzorka) – jer je jedan od preduvjeta za korištenje t-testa homogenost varijanci odnosno pretpostavka o tome da se varijance dviju populacija međusobno ne razlikuju (konceptualno opravdanije)
- ili zajedničku standardnu devijaciju (ako imamo podjednak broj ispitanika u obje skupine) – jer je ta mjera stabilnija odnosno određena je na većem broju ispitanika:

$$\bar{\sigma} = \sqrt{\frac{\sigma_1^2(n_1-1)+\sigma_2^2(n_2-1)}{n_1+n_2}}$$

Cohenov d

- s obzirom na to da su nam najčešće nepoznate vrijednosti prave aritmetičke sredine i prave standardne devijacije, već imamo samo njihove procjene, i Cohenov d može se smatrati *procjenom* veličine učinka
- J. Cohen (1988): konvencije o značenju veličini učinka
  - razlikuju se za različite statističke procedure!
  - bazirane su na učincima pronađenim u psihologijskim istraživanjima i služe kao vodič za odlučivanje važnosti učinka dobivenog u vlastitom istraživanju
  - veličinu učinka uvijek treba razmotriti u relaciji s područjem koje se istražuje kao i s praktičnim i/ili kliničkim značenjem tog učinka

Cohenov d

- Cohenova konvencija za veličinu učinka – usporedba dviju populacija (t-test za nezavisne uzorke):
  - d=0.2; mala veličina učinka – prekrivanje distribucija oko 85 %
  - d=0.5; srednja veličina učinka – prekrivanje distribucija oko 67 %
  - d=0.8; velika veličina učinka – prekrivanje distribucija oko 53 %

Veličina učinka: t-test za zavisne uzorke

$$d = \frac{M_{poslije} - M_{prije}}{SD_{prije}}$$

Cohenova konvencija:
d=0.10 mala veličina učinka
d=0.25 srednja veličina učinka
d=0.40 velika veličina učinka

Veličina učinka: Pearsonov koeficijent korelacije

- Cohen, 1988.
  - r=0.10 (malen učinak)
  - r=0.30 (umjeren učinak)
  - r=0.50 (velik učinak)

Statistička snaga

- **vjerojatnost** da će istraživanje rezultirati statistički značajnim rezultatom *ako* je istraživačka hipoteza točna
- zašto je važna?
  - zbog planiranja veličine uzorka
  - zbog razumijevanja smisla nekog statistički neznačajnog rezultata
  - zbog razumijevanja smisla nekog statistički značajnog rezultata koji nije važan u praktičnom smislu

Statistička snaga

- ako neko istraživanje ima malu statističku snagu, mala je šansa da će rezultat biti statistički značajan čak i ako nul-hipoteza nije točna
- što je statistička snaga manja, veća je šansa za pogrešku tipa II

Određivanje statističke snage

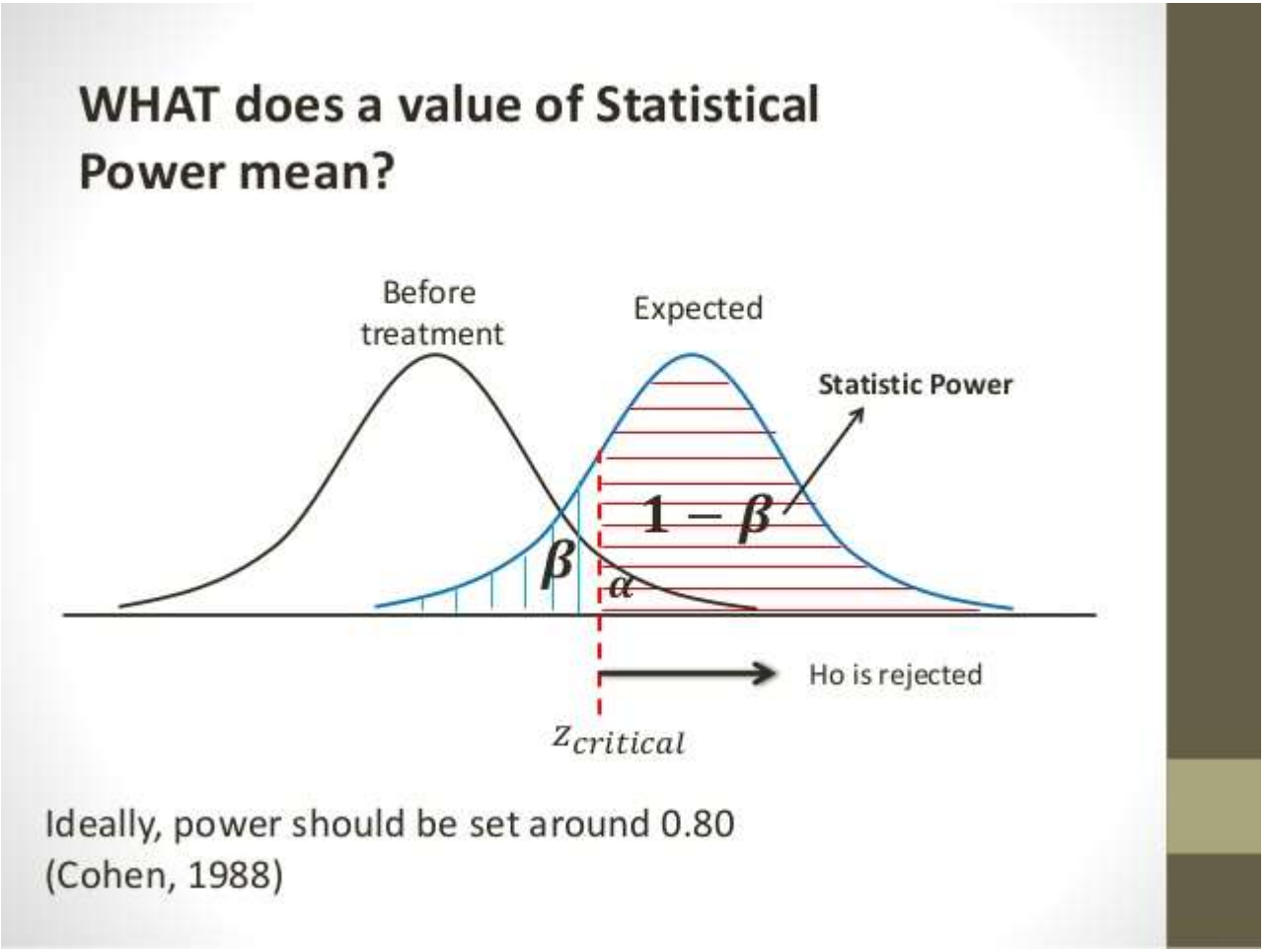
1. „pješice” (kao u primjeru)
2. kalkulatori statističke snage na internetu
3. softverski paketi

   (istraživač unosi vrijednosti različitih aspekata svog istraživanja kao što su poznata populacijska M, predviđena populacijska M, SD populacije, veličina uzorka, razina rizika i jednosmjerno vs. dvosmjerno testiranje)
4. tablice statističke snage (pokazuju statističku snagu istraživanja ovisno o dobivenoj veličini učinka i veličini uzorka) (Cohen, 1988; Kraemer i Thiemann, 1987)

Što određuje statističku snagu?

- **veličina učinka**
- **broj ispitanika**
- **odabrana razina rizika**
- **jednosmjerni vs. dvosmjerni test**
- vrsta statističke procedure

Na što se odnosi vrijednost statističke snage (1-β)?



Preuzeto s: https://www.slideshare.net/husnulkhuluq/20140602-statistical-power-husnul-and-nur

Statistička snaga i veličina učinka

- što je veličina učinka veća, veća je i statistička snaga – što je razlika između dvije M veća, distribucije se manje preklapaju, β je manja, a snaga veća

Statistička snaga i veličina uzorka

- što je veći N, veća je i statistička snaga:
  - što je veći N, manja je standardna pogreška aritmetičke sredine, a što je manja σ_M, distribucija je uža, tj. manje je preklapanje, izračunati *t* bit će veći i bit će veća vjerojatnost da bude statistički značajan (uz niže razine rizika)
- određivanje potrebnog N-a: istraživač unaprijed odabere stupanj statističke snage (npr. 80 %) koju želi imati u istraživanju te pomoću softvera, internet kalkulatora ili tablica odredi potreban broj ispitanika

Statistička snaga i razina rizika, jednosmjerni vs. dvosmjerni test i vrsta statističke procedure

- što je veća razina rizika (npr. $p<0.10$ ili $p<0.20$), veća je i statistička snaga
- ali, veća razina rizika znači i veću vjerojatnost za pogrešku tipa I
- kod jednosmjernog testa, statistička snaga veća je nego kod dvosmjernog testa (ako su svi ostali čimbenici izjednačeni)
- parametrijski statistički postupci imaju veću statističku snagu od neparametrijskih statističkih postupaka

Uloga statističke snage prilikom planiranja istraživanja

- ako provodimo istraživanje koje ima slabu statističku snagu, čak i ako je naša istraživačka hipoteza točna, mala je šansa da ćemo moći odbaciti nul-hipotezu (pa su vrijeme, energija i novac koji su utrošeni na planiranje i provedbu istraživanje – uzaludni)
- koliko statističke snage je dovoljno?
- široko prihvaćeno pravilo: 80 % (Cohen, 1988)
- statistička snaga od 80 % znači da postoji 80 % šanse da će istraživanje rezultirati statistički značajnim rezultatom (ako je razlika uistinu postoji u populaciji)
- kriterij od 80 % često je teško postići (preskupo)

Kako možemo povećati statističku snagu

1. kroz povećavanje očekivane veličine učinka
   - ali to očekivanje mora biti utemeljeno (npr. kroz varijacije u metodi), a ne arbitrarno određeno
2. kroz smanjivanje pogreške
   - moguće postići na dva načina: 1. odabir populacija s manjom varijancom i 2. standardizacija uvjeta testiranja i korištenje preciznijih mjera
3. kroz povećavanje veličine uzorka
   - ponekad nemoguće, ali najčešće skupo

Kako možemo povećati statističku snagu

4. kroz odabir veće razine rizika

› odabir što veće razine rizika, ali da bude razumna s obzirom na vjerojatnost pogreške tipa I – u psihologiji je to $p<0.05$

5. kroz korištenje jednosmjernog testa

› u psihologiji relativno rijetko – jer se i direktivne (istraživačke) hipoteze testiraju dvosmjernim testom

6. kroz primjenu osjetljivijih statističkih procedura (npr. Pearsonov r umjesto Spearmanovog Rho)

Statistička snaga i interpretacija istraživačkih rezultata

- rezultat je statistički značajan
  - kada je snaga velika, istraživanje s vrlo malom veličinom učinka može dati statistički značajan rezultat (vjerojatno posljedica velikog N-a), koji ne mora imati *klinički značaj* odnosno *praktičnu važnost*
  - dakle, kada je rezultat statistički značajan, treba se zapitati je li veličina učinka dovoljno velika da bude korisna ili interesantna (naročito ako istraživanje ima praktične implikacije)

Statistička snaga i interpretacija istraživačkih rezultata

- rezultat nije statistički značajan
  - mogući razlozi: nul-hipoteza je točna ili je statistička snaga bila nedovoljna (jer je npr. bio premali broj ispitanika)
  - stoga, ako je u istraživanju velike snage dobiven statistički neznačajan rezultat, to je znak da je malo vjerojatno da je istraživačka hipoteza točna ili da je učinak zapravo manji od onog kojeg smo predvidjeli

LITERATURA KORIŠTENA ZA PRIPREMU MATERIJALA:

- Aron, A., Coups, E.J., Aron, E.N. (2013). *Statistics for psychology*. Pearson.
- Cohen, B.H., Lee, R.B. (2004). *Essentials for the social and behavioral sciences.* New Jersey: John Wiley & Sons.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage Publications.
- Howell, D.C. (2002). *Statistical methods for psychology*. Duxbury: Wadsworth Group.
- Kolesarić, V. i Petz, B. (2003). *Statistički rječnik*. Jastrebarsko: Naklada Slap.
- Miles, J., Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction*. London: Sage Publications.
- Petz, B., Kolesarić, V. i Ivanec, D. (2012). *Petzova statistika: Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap.