

2. predavanje RVM – Kvantitativne metode  
Borut Kodrič, Koper 21.5.2010

Ključ za dostop do e-učilnice: RMD2009

Tekom srečanj bodo zadeve osvežene v smislu, da bodo okleščene. Morda bo dodan še kak rešen primer.

## Literatura

### Literatura - 1

---

- Blejec M., **Uvod v statistiko**, več izdaj...
- Pfajfar L., Arh F., **Statistika 1**, zapiski predavanj, EF, Ljubljana 2002
- Košmelj B., **Statistika 2 – 1. del**, skripta, EF, Ljubljana, 1996

Prvi dve literaturi (za ponovitev, če so luknje v predznanju).

Ključna literatura. V originalu je naslov **Statistika 2-1. del**. Izdan je bil njen ponatis **Statistično sklepanje** (Ekonomska fakulteta v Ljubljani).

### Literatura - 2

---

- Kalton, G., Vehovar V., **Vzorčenje v anketah**, FDV, Ljubljana, 2001
- Košmelj B., **Analiza odvisnosti za vzorčne podatke**, EF, Ljubljana, 1996
- Hair, Black, Babin, Anderson, Tatham, **Multivariate Data Analysis, 6<sup>th</sup> Edition**. New Jersey, Prentice-Hall, Pearson, cop. 2006

V knjigi **Analiza odvisnosti za vzorčne podatke**. V njej je predstavljena regresijska analiza, ki jo bomo kasneje podrobno spoznali.

Poleg regresijske si bomo pogledali tudi uporabo **faktorske analize**, za katero nekega zgledega učbenika v Sloveniji žal ni. V ta namen je navedena tretja literatura, dostopna na FM. Začne z elementarnimi zadevami in kasneje vse do konkretnih vprašanj.

## Literatura - 3

---

- Koprivnik S, Kogovšek T., **Analize podatkov z SPSS-om 12.0**. FDV, 2006
- Rovan J., Turk T., **Analiza podatkov s SPSS za Windows**. Ekonomska fakulteta v Ljubljani.

S programom SPSS se, glede na število ur, ne bomo ukvarjali.

Na izpitu so zastavljena neka vprašanja, zraven pa dobimo računalniški izpis izračunov.

Ta izpis je potrebno poznati. Vse izpise, ki jih moramo poznati, bomo spoznali v smislu interpretacije. Ne bomo se učili, da bi sami prišli do teh izpisov, naučili se pa bomo kaj rezultati, ki nam jih ta program ponudi, pomenijo. To je tudi bistveno.

Izpiti so sestavljeni iz dveh, če ne celo iz treh delov. Dva dela predstavljata pisni del izpita, kjer dobimo nekaj računskih nalog. Računanje je omejeno na seštevanje odštevanje in množenje. Precej izračunov dobimo narejenih s programom SPSS. Morda je treba kako številko še sešteti, ipd. So štiri (4) take naloge. Drugi del pisnega izpita predstavljajo vprašanja z obkroževanjem. Ponavadi so štirje odgovori. Vsako nepravilno obkroženo vprašanje prinaša **minus** točke. *Če nečesa ne znaš raje pusti in ne ugibaj!* V ta namen so uporabljene negativne točke. Izpitni roki so med seboj uravnoteženi (primerljive težavnosti).

## Osnovni pojmi-1

---

- Statistična populacija:
  - množica vseh istovrstnih pojavov – statističnih enot, ki so bili izbrani za namene analize
  - ti pojavi izpolnjujejo opredeljujoče pogoje (časovna, krajevna in stvarna opredelitev)
- Statistična enota: posamezen pojav statistične množice
  - Realne enote
  - Dogodki
  - Dogajanja

### STATISTIČNA POPULACIJA

To je tista množica, ki jo v neki raziskavi, analizi opazujemo. V glavnem bomo operirali z vzorčnimi opazovanji. Tudi če opazujemo zgolj nek vzorec nam mora biti še vedno jasno iz katere populacije je bil ta vzorec izbran. **Naš končni cilj je vedno ta, da naj bi prišli do**

**spoznanj o značilnostih populacije.** Zato moramo na samem začetku vedeti kaj sploh naša populacija je. Katere so tiste enote, ki nas zanimajo.

Navesti moramo tri opredeljujoče pogoje:

- **Stvarni**, ki nam pove za kakšno vrsto enot v naši populaciji gre (da se opredeli ali opazujemo študente, zaposlene, brezposelne ali opazujemo države, gospodinjstva, torej da povemo za kakšno vrsto enote gre).
- **Krajevna opredelitev.** Geografsko opredelimo na katerem geografskem območju naj bi se te enote, ki jih opazujemo, nahajale (največkrat so raziskave omejene na območje Republike Slovenije, lahko gre samo za občino, lahko pa celo na območje EU).
- **Časovna opredelitev.** Populacije so zelo spremenljive. Če recimo opazujemo prebivalce Republike Slovenije. Tiste, ki so bili včeraj rojeni, ki se bodo jutri rodili, tisti, ki so morda že včeraj umrli pa nas še zanimajo ali ne. **Opredeliti moramo na katero stanje oz. obdobje se nanaša populacija.** Ali gre za dan ali opredeljujemo stanje ali neko obdobje, ločujemo statistične enote, ki so posamezni elementi populacije, na tri vrste (če pogledamo njihove značilnosti sta dejansko samo dve vrsti):
  - **Realne enote** (tiste ki v času obstajajo in ugotavljamo njihovo stanje na nek določen dan). (Primeri: Realne enote smo prebivalci. Če zasledimo v medijih objavo podatka o številu prebivalcev RS je ta podatek vedno opremljen z nekim datumom [število prebivalcev na dan 31.12.2009 je bilo tako in tako]). Gre za vse prebivalce, ki so bili do tega dne rojeni oz. še niso umrli, vse ki so do tistega dne priselili in prijavili svoje stalno prebivališče in vsi, ki se do tistega dne še niso odselili.
  - **Dogodki/Dogajanja** gre za opazovanje nekega obsega v nekem obdobju. (Primer dogodkov: če opazujemo populacijo prebivalstva so prebivalci realne enote. Če pa bi opazovali samo rojstva v RS so rojstva dogodki. Če prebivalce opazujemo na nek določen dan, gre za opazovanje dogodkov v nekem določenem intervalu. Podatek o prebivalstvu vedno podajamo na določen dan, število rojstev pa je običajno podano na neko določeno obdobje (število rojstev v mesecu, letu, desetih letih, itd.). Razlika med dogodki in dogajanja je zgolj v vidiku trajanja.

**Dogodki** so neke krajše zaključene celote (rojstvo, priselitev, odselitev).

**Dogajanja** (gre za gradnjo nekega objekta, ladje ali letala kjer se to ne zgodi tudi v roku enega leta. Včasih celo v desetih letih). **Tudi tu opazujemo njihov obseg v nekem intervalu.**

Če bomo želeli opredeliti populacijo prebivalcev RS bi bilo potrebno to opredeliti na dan ko smo to populacijo opazovali.

Če pa opredelimo populacijo kot populacijo rojstev ali priseljencev, to opredelimo na neko določeno obdobje. (Koliko priselitev se je recimo zgodilo v letu 2009 oz. koliko rojstev se je zgodilo v letu 2009).

## Osnovni pojmi-2

---

- Statistične spremenljivke: *značilnosti stat.enot, ki so predmet analize*
  - Krajevne;
  - Časovne;
  - Stvarne:
    - Opisne
    - Številске(nerične)
- Statistični parametri: *značilnosti populacije kot celote*

Zakaj gre pri spremenljivkah? Če so enote tisto kar opazujemo in sestavlja populacijo, so **statistične spremenljivke tisto, kar nas pri posamezni enoti zanima** (lastnosti posamezne enote). Temu pravimo spremenljivka zato ker **so to lastnosti, ki naj bi se od enote do enote razlikovale**. Tiste lastnosti, ki so enake vsaki enoti, je nesmiselno spremljati. (Primer: Če opredelimo populacijo kot populacijo prebivalcev RS, potem je nesmiselno, da vsakega izmed naših enot vprašamo iz katere države je. Če smo se že prej opredelili za prebivalce RS je razumljivo, da vse enote prihajajo iz RS).

**Zanimajo nas samo lastnosti, ki so od enote do enote različne. Zato poudarjamo da so to spremenljivke.**

Spremenljivke delimo na :

- **Krajevne** (posameznika recimo vprašamo kaj od rojstva, bivališča, zaposlitve);
- **Časovne** (datum rojstva, datum zaključka študija, datum prve zaposlitve, itd.)
- **Stvarne** (predstavljajo najširšo množico spremenljivk, kjer so take, ki jih izražamo
  - **opisne** (spol, narodnost, vrsta končane izobrazbe);
  - **številске** (vse kar se izraža s številko in ima to številski pomen). Teh je tudi največ. Ukvarjali se bomo v glavnem s temi.

Če so statistične spremenljivke lastnosti posameznih enot, so **statistični parametri značilnosti populacije kot celote**. (Primer: Opredelimo populacijo kot prebivalstvo republike Slovenije. Vsaka posamezna enota predstavlja posameznega prebivalca. Recimo da nas kot spremenljivka zanima višina mesečnih prejemkov posameznika. Za vsakega bo ta višina različna [nekateri imajo višje, nekateri nižje prejemke]. Če izračunamo **povprečno višino mesečnih prejemkov za celotno prebivalstvo RS**, dobimo eno samo vrednost, kot povprečno vrednost za celotno populacijo. **Temu potem pravimo parameter.**

**TISTO KAR JE ZNAČILNOST POPULACIJE JE PARAMETER, TISTO KAR JE ZNAČILNOST POSAMEZNE ENOTE JE SPREMENLJIVKA.**

Spremenljivka se nanaša na poudarke s katerimi obdelujemo (računamo), rezultati, ki jih dobimo so pa parametri. Obdelujemo podatke, ki se nanašajo na posamezne enote, rezultati, ki jih dobimo, so pa izračuni o tem, kar se nanaša na našo populacijo.

## Osnovni pojmi - primer

---

- V okviru analize življenjske ravni zaposlenih v dejavnosti trgovine smo pri zaposlenih v dejavnosti trgovine na drobno (G52 po SKD klasifikaciji dejavnosti) na dan 31.12.2005 izračunali povprečno plačo na zaposlenega v mesecu decembru 2005. Opredeli kaj je statistična populacija, statistična enota, stat.spremenljivka in stat. parametri.

Ukvarjamo se z analizo življenjske ravni zaposlenih, pri čemer se omejimo na eno izmed dejavnosti. Zaposlene smo opazovali na dan 31.12.2005. Na podlagi zbranih podatkov smo izračunali povprečno plačo na zaposlenega v mesecu decembru. **Skušamo opredeliti kaj je statistična populacija, kaj je enota, kaj je spremenljivka in kaj so razvidni ali možni statistični parametri.**

Če želimo opredeliti kaj je **populacija**, je treba to zelo natančno opredeliti:

- s stvarnim opredeljujočim pogojem;
- s krajevnim;
- s časovnim.

**Stvarni opredeljujoči pogoj** je, da gre za zaposlene v dejavnosti trgovine na drobno.

**Krajevni** predstavlja zaposlene na območju republike Slovenije.

**Časovni pogoj** je, da so bili v delovnem razmerju na dan 31.12.2005. (Če se je nekdo zaposlil šele s 1.1.2006 ni predstavljal naše populacije v trenutku opazovanja).

**CELOTNA OPREDELITEV POPULACIJE** bi se natančno opredeljeno glasila:

**To so zaposleni v dejavnosti trgovine na drobno, na območju Republike Slovenije na dan 31.12.2005.**

**Spremenljivka** (značilnost posamezne enote populacije), ki jo je mogoče razbrati je v tem primeru **višina plače v mesecu decembru**. (spremenljivka ni povprečna plača!). Spremenljivka se nanaša na posamezno enoto in vsaka **posamezna enota (delavec)** je lahko dobil drugačno višino plače.

Ko na podlagi zbranih podatkov, za posameznike izračunamo njihovo **povprečno plačo, dobimo parameter** (značilnost populacije), ki se nanaša na celotno populacijo.

Povzetek za dani primer:

1. **Kaj je statistična populacija?** Zaposleni v dejavnosti trgovine na drobno, na območju Republike Slovenije, na dan 31.12.2005.
2. **Kaj je statistična enota?** Delavec.
3. **Kaj je statistična spremenljivka?** Višina plače v mesecu decembru.
4. **Kaj je statistični parameter?** Povprečna plača.

## Opredelitev populacije - primer

---

- Trgovsko podjetje se odloča za investicijo v nov nakupovalni center v izbranem slovenskem kraju. Odločajo se med dvema lokacijama (v istem kraju). Za podporo odločitvi nameravajo izvesti anketo potencialnih potrošnikov. Opredelite populacijo.

V tem primeru opredelitev populacije ni tako enostavna. Kaj je v tem primeru populacija? Če želimo dati pravilen odgovor je nujno potrebno razumeti kaj je vsebina oz. namen te raziskave. Ta izhaja iz tega za kakšno vrsto centra v tem primeru gre.

Če gradimo nakupovalni center v katerem bomo prodajali blago za široko potrošnjo (od oblačil, obutve, prehrane) bo ta center namenjen lokalnemu okrožju in se morda omejimo na prebivalstvo, ki ima stalno prebivališče v občini v kateri bo ta center lociran.

Če pa gradimo ta nakupovalni center v stilu Emporiuma ali bolj specializirane prodajalne. Namen imamo privabiti kupce iz širšega območja. V tem primeru populacija ne predstavlja samo prebivalce, ki imajo stalno prebivališče v tistem kraju ampak celo prebivalce na celotnem območju RS.

Ni torej čisto trivialno jasno kaj populacija je. Če v startu narobe opredelimo populacijo (primer: načrtujemo izgraditev centra kot je Emporium, ki je osredotočen na prodajo oblačil in obutve pretežno nekoliko višjega razreda. Ker takih kupcev v enem samem kraju ni in če populacijo opredelimo samo kot prebivalce, ki imajo stalno prebivališče na območju npr. mestne občine Ljubljana, lahko dobimo povsem izkrivljene rezultate raziskave. Zato moramo v startu dobro opredeliti populacijo. Vsi rezultati, ki jih dobimo, se bodo nanašali na našo populacijo.

### Etape kvantitativne analize-1

---

- 1) Določitev vsebine in namena analize
  - Opredelitev pojava in osnovnih pojmov
  - Določitev spremenljivk
  - Določitev parametrov
- 2) Izdelava splošnega načrta
  - Natančna opredelitev populacije, opredelitev metod zbiranja podatkov ter določitev metod in postopkov osnovne in analitične obdelave

V prvem koraku je zelo pomembno, da pred zbiranjem podatkov, **jasno opredelimo kaj proučujemo** in kaj želimo s tem raziskovanjem, proučevanjem spoznati, ugotoviti oz. odkriti.

Potem je tudi bistveno lažje opredeliti populacijo (koga moramo opazovati). Jasno je tudi katere značilnosti enot moramo opazovati. Na samem začetku lahko enostavno **določimo spremenljivke**. Če vemo kaj so cilji naše raziskave, vemo tudi približno katero metodo bomo uporabili in kaj moramo izračunati. Potem je relativno trivialna tudi **določitev parametrov**.

Drugi korak je bolj v smislu operacionalizacije tistega kar smo v prvem koraku bolj teoretično določili. Na podlagi opredelitev populacije poiščemo vire, kjer bomo prišli do seznama vseh operacijskih enot. (Primer: če rečemo, da opazujemo prebivalce RS moramo priti do seznama vseh prebivalcev, kot je Centralni Register Prebivalstva [CRP]). Če se osredotočimo na poslovne subjekte, ki delujejo na območju RS, potem kot vir, teh enot, izberemo nek poslovni imenik ali poslovni register. Če opazujemo rojstva na območju RS, pogledamo spet v CRP, kjer se ti dogodki evidentirajo in pridemo do vseh rojenih otrok v nekem obdobju in to predstavlja našo populacijo, itn.).

## Etape kvantitativne analize-2

---

### 3) Zbiranje podatkov

- Primarni vs. sekundarni podatki
- Longitudinalno opazovanje vs. transverzalno opazovanje

### 4) Osnovna obdelava

- Razvrščanje v skupine, preštevanje, agregiranje

### 5) Analitična obdelava

- Ugotavljanje značilnosti populacije  
->izračun parametrov

(3) Pri zbiranju podatkov imamo na voljo dve možnosti. Če raziskujemo neko področje, za katerega obstaja že vrsto podatkovnih virov, običajno uporabimo kar **sekundarne podatke**. To so tisti, ki so bili zbrani že za kakšne druge namene. (Primer: sekundarni podatki so vsi tisti, ki jih zbira Statistični urad).

**PRIMARNI PODATKI** so tisti, ki jih zberemo sami, glede na naše potrebe. Kdaj se poslužujemo enih in kdaj drugih?

### ZNAČILNOSTI SEKUNDARNIH PODATKOV

Priporočajo se za tiste, ki nimajo preveč izkušenj s samostojnim zbiranjem podatkov. Sekundarne podatke običajno zbira, na podlagi neke formalne, zakonodajne podlage, uradna institucija (Statistični urad, Banka Slovenije, ministrstva). V te podatke je običajno vključena zelo velika populacija (Davčni urad, ki odmerja dohodnino. Gre praktično za vse davčne zavezance). Če želimo recimo analizirati življenjsko raven zaposlenih in navedemo kot osrednjo spremenljivko plačo, ob ustreznem dovoljenju, na Davčni upravi lahko dobimo dostop do podatkovne baze dohodninskega sistema).

Ti podatki so bolj verodostojni, bolj popolni in kontrolirani. Podatke zbirajo institucije, ki se že dlje časa ukvarjajo z zbiranjem podatkov in imajo za to usposobljene kadre, izkušnje in ustrezno informacijsko podporo. Običajno so ti podatki na voljo zastonj (spletne strani slovenskega in evropskega statističnega urada (EUROSTAT), organizacija za ekonomsko sodelovanje in razvoj (OECD), svetovna banka) in v trenutku.



Pomanjkljivost je pa recimo, če bi skušali analizirati kakšna je organizacijska klima zaposlenih v tekstilni industriji v Sloveniji, teh podatkov recimo na statističnem uradu ne spremljajo. To ne sodi v okvir njihovih raziskav. V tem primeru se moramo lotiti zbiranja podatkov sami.

**Lahko rečemo da vse tisto kar so prednosti sekundarnih podatkov so zelo velikokrat slabosti primarnih.**

Sekundarne podatke zbira neka usposobljena institucija, ki ima ustaljeno prakso, kadre, itd., se mi kot raziskovalci običajno lotevamo zbiranja podatkov velikokrat z nekoliko manj izkušnjami. Imamo tudi manj kadrov na razpolago. Vsako zbiranje podatkov zahteva čas.

### TRANSVERZALNO OPAZOVANJE

Opazujemo neko množico pojavov zgolj v enem samem obdobju (ne zanima nas kaj je bilo v preteklosti kaj bo v prihodnosti). V angleškem jeziku cross sectional data. **Presečni podatki.**

### LONGITUDINALNO OPAZOVANJE

Tu se omejimo na opazovanje enega samega pojava v daljšem časovnem obdobju. Rezultat takega opazovanja so časovne serije podatkov (time series data).

V realnosti se uporablja tudi kombinacija obeh pristopov. Da opazujemo večje število pojavov in njihovo povezanost v daljšem časovnem obdobju. To so **panelni podatki.**

Večinoma se bomo omejili na **analizo presečnih podatkov**, ki so bili zbrani v transverzalnem opazovanju.

(4) Potem se lotimo obdelave podatkov. Gre za razvrščanje enot v skupine, morda osnovni parametri, kot aritmetična sredina, deleži, indeksi. Gre največkrat za predstavitev našega vzorca ali populacije. Zato da bo bralec raziskave rezultate lažje razumel. **Ta korak nam predstavi kontekst v katerem je bila raziskava izpeljana.**

(5) Sledi uporaba metod kot je regresijska, faktorska, cluster analiza, itd.

## Vrste spremenljivk glede na raven meritve

---

- Nominalne spremenljivke
  - $Y_a \neq Y_b$
- Ordinalne spremenljivke
  - $Y_a > Y_b$
- Intervalne spremenljivke
  - $Y_a - Y_b$
- Razmernostne spremenljivke
  - $Y_a / Y_b$

Če imamo spremenljivko spol, povprečnega spola ne moremo računati. Prav tako spremenljivka stopnja izobrazbe, ki lahko zavzame vrednosti nedokončana osnovna šola, osnova šola, poklicna, srednja, višja, visoka, podiplomska. Tudi tu ne moremo računati povprečne stopnje izobrazbe.



Spremenljivke moramo najprej ločiti (na podlagi vsebine ločujemo krajevne, časovne in stvarne). Tu gre pa za delitev z vidika kaj posamezna spremenljivka omogoča, oz. kje na merski lestvici ta spremenljivka kotira.

## Nominalne spremenljivke

---

- Ali sta opazovani vrednosti enaki ali ne;
- Tudi številske izražene spr. so lahko nominalne;
- Primeri: *spol, kraj bivališča, narodnost...*
- Metode:
  - Razvrščanje v skupine (frekvenčne porazdelitve, strukture)
  - Modus

**Nominalne spremenljivke**, predstavljajo najnižji nivo. Gre bolj zato da znamo odgovoriti ali sta dve vrednosti enaki ali ne (Primer: spol [moški/ženski], moški je enak moškemu, moški pa ni enak ženski. Ne moremo reči povprečen spol, ne moremo dveh spolov sešteti). Edini parameter, ki ga lahko ugotovimo tako da preštejemo kolikokrat se katera vrednost ponovi je **modus**. Ne potrebujemo nobenega obrazca, ugotovimo ga tako da preštejemo katera vrednost se je največkrat ponovila. (v slovenščini se temu pravi gostišnica).

## Ordinalne spremenljivke

---

- Enote lahko razvrstimo po vrednosti spremenljivke tako, da imajo posamezne enote prednost pred drugimi
- Primeri: *stopnja izobrazbe, kakovost proizvoda, letnik študija, ocena...*
- Metode:
  - Ranžirna vrsta in kvantili
  - Mediana, modus

**Ordinalne spremenljivke** so tiste, ki nam omogočijo, da jih razvrstimo po velikosti (stopnja izobrazbe, letnik študija, kakovost proizvoda, itn.). Enote razvrstimo v optimalno ranžirno vrsto, izračunavamo kvantile (najbolj znani kvantil je mediana). Še vedno pa ne moremo ugotoviti kakšna je razlika med dvema vrednostima (lahko rečemo da tisti, ki ima srednjo izobrazbo je to višje stopnje kot pa tisti ki ima osnovno in obratno. Ne moremo pa reči, da je razlika med osnovnošolsko in srednješolsko izobrazbo enako 1).

## Intervalne spremenljivke

---

- Razliko med dvema vrednostima je smiselno izračunati z odštevanjem vrednosti
$$y_a - y_b = d$$
- Primeri: *temperatura, poslovni izid...*
- Metode:
  - Vse mere srednjih vrednosti,
  - Mere variabilnosti,
  - Mere odvisnosti,
  - Mere dinamike.

**Intervalne spremenljivke** šele tu lahko govorimo o številskih spremenljivkah. Pri teh že lahko ugotovimo ali sta dve vrednosti enaki ali ne. Lahko jih razvrstimo po velikosti. Za dve vrednosti lahko izračunamo njuno medsebojno razliko. Praktično lahko uporabljamo že skoraj vse metode (vse mere srednjih vrednosti, tehtana aritmetična, harmonična sredina, mere odvisnosti, varianca, korelacije, itd.). Ne moremo pa računati njunih medsebojnih razmerij (Primer: zimski dan 16.2. Zjutraj ob 7. uri temperatura zraka v Kopru 4 st. Celzija, v Ljubljani pa istega dne ob isti uri -2 st. Celzija. [temp. zraka KP – temp. zraka LJ = 4 – (-2) = 6 st. Celzija]. Nesmiselno pa je računati razmerje [temp. zraka KP / temp. zraka LJ = 4 / -2 = -2, kot indeks recimo -200]. Tega ne moremo interpretirati).

## Razmernostne spremenljivke

---

- Enolično določeno izhodišče (vrednost 0 je upoštevana absolutno)
- Primeri: *večina številskih spremenljivk*
- Metode:
  - vse poznane kvantitativne metode

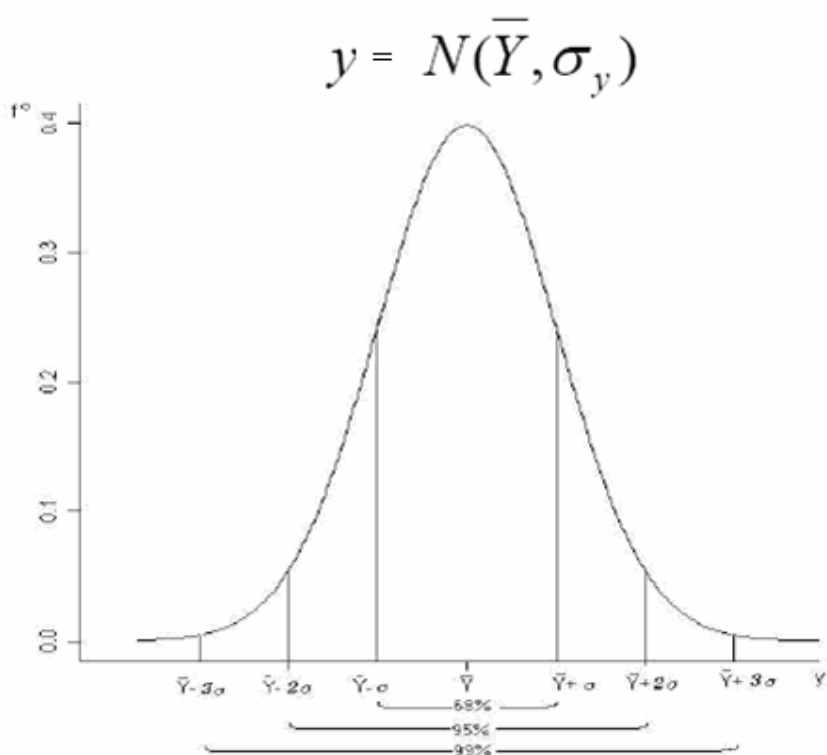
**Razmernostne spremenljivke** Razmerja lahko računamo za tiste spremenljivke, ki so po značilnostih razmernostne spremenljivke. To so tiste, ki lahko zavzamejo samo pozitivne ali samo negativne vrednosti. Vrednost nič (0) pomeni absolutno odsotnost pojava (Primer: če opazujemo gospodinjstva v RS, kot naše enote. Spremenljivka, ki nas pri posameznem gospodinjstvu zanima je pa npr. število otrok. Število je lahko nič, ena, dva, tri, štiri, itn. Če imamo v enem gospodinjstvu pet otrok v drugem pa enega samega otroka lahko rečemo, da so v drugem gospodinjstvu štirje otroci več kot v prvem oz. da je v drugem gospodinjstvu petkrat več otrok kot v prvem.

Recimo da je cilj naše raziskave ugotoviti kako višina mesečne plače vpliva na zadovoljstvo zaposlenih. Vpliv ene spremenljivke na drugo se ugotavlja s pomočjo regresijske analize. Regresijska analiza zahteva pravi številski podatek. Ne moremo dati kot vhodni podatek takšnega, ki je po svoji vsebini zgolj nominalen. (Zadovoljstvo je potrebno izraziti z neko številsko vrednostjo).

Lestvica spremenljivk je hierarhična. Za vsako višjo stopničko veljajo tudi vse prejšnje značilnosti. Za intervalno spremenljivko lahko ravno tako ugotovimo katera vrednost je večja oz. manjša, ali sta dve vrednosti enaki ali ne. Ne moremo pa izračunati razmerij.

## NORMALNA PORAZDELITEV (GAUSSOVA KRIVULJA)

# Normalna porazdelitev

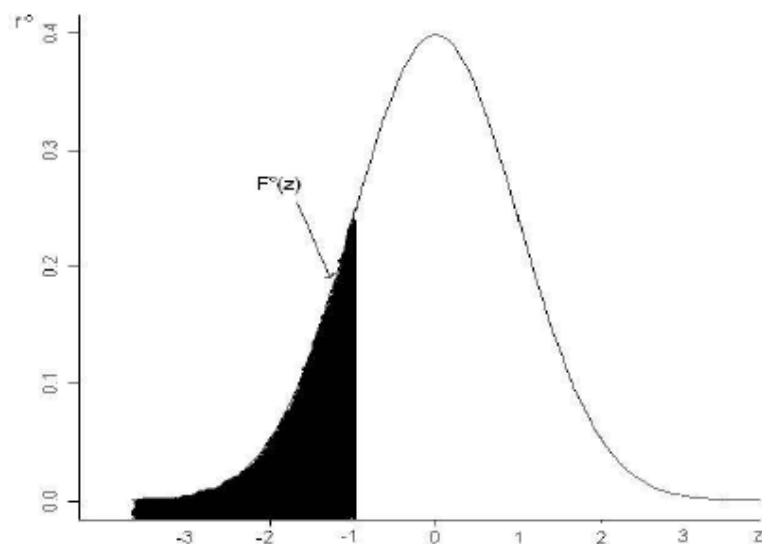


Pomembno je da je normalna krivulja simetrična, levo in desno enake oblike. Da je opredeljena z dvema parametroma. Taka krivulja nam omogoča cel kup hipotetičnih izračunov. Če našo porazdelitev pretvorimo v standardizirano normalno porazdelitev, ki jo bomo pogosto uporabljali lahko za katerokoli spremenljivko ugotovimo kakšna je verjetnost, da njena vrednost leži na nekem intervalu.

## Standardizirana normalna porazd. - 3

---

$$P(y < y_a) = P(z < z_a) = F^o(z)$$



Potem se izhaja iz nekih tabeliranih vrednosti verjetnosti, da med aritmetično sredino minus en standardni odklon plus en standardni odklon leži nekaj več kot 2/3 enot. Če gremo od aritmetične sredine levo in desno za dva standardna odklona lahko iz tabeliranih vrednosti izračunamo, da leži na tem intervalu praktično 95% vseh vrednosti oz. enot, če gremo pa levo in desno tri standardne odklone od aritmetične sredine ležijo tu skoraj vse enote.

## Vzorčna opazovanja

---

- Razlogi:
  - finančna sredstva,
  - čas,
  - zahtevnost popolnega zajetja in možnost napak,
  - opazovanje populacije ni izvedljivo.
- Slučajni vs neslučajni vzorci

Ko se lotimo vzorčenja nikoli ne moremo z gotovostjo nečesa trditi. Vse naše sklepanje moramo nasloniti oz. utemeljiti na nekih porazdelitvah iz katerih bomo izpeljevali neke verjetnosti, da je sklep pravilen oz. napačen. Te verjetnosti so osnovane na nekih teoretičnih porazdelitvah. Zelo pogosto se kot osnova uporablja normalna porazdelitev. Tablice s katerimi operiramo so tiste, v katerih je tabelirano kakšna je kumulativna verjetnost normalne porazdelitve.

Populacije v družboslovju so zelo velikokrat množične. Tudi če je samo 100.000 enot, je raziskava (anketa) časovno potratna, zahtevna in draga. Obdelava množične populacije zahteva velika finančna sredstva (plačilo delovne sile, materialna sredstva, ki se uporabljajo, ankete, kuverte, itd.). Gre tudi za zahteven logistični postopek. Potrebno je usposobiti vse anketarje, da so seznanjeni z metodologijo, z vprašalnikom. Zato se omejimo na vzorec. Dobimo sicer manj kakovostne podatke ampak bo to še vedno dovolj dobra ocena za sprejemanje odločitev. Vzorec predstavlja neko podmnožico celotne populacije. Kako to podmnožico izbrati? Tu se uporabljata dve skupini načinov izbire:

1. **Enostavno slučajno vzorčenje** (enote za izbiro v vzorec so čisto naključne);
2. **Neslučajni vzorci** (vzorci kjer izbiranje ni naključno ampak si vnaprej postavimo neke kriterije. Recimo izberemo tak vzorec, da bo v njem 10 moških pa 10 žensk. Vnaprej se omejimo s kriterijem).

## Slučajni vzorci

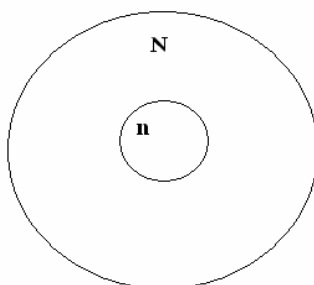
- Vse vzorčne enote imajo enako možnost, da so izbrane v vzorec -> poznana verjetnost za izbiro vzorčne enote v vzorec
- Vse množice enot (vzorci) imajo enako verjetnost, da so izbrani
- Objektivnost pri izbiri enot v vzorec
- Možno ocenjevanje kakovosti vzorčnih ocen s posebnimi kazalci

Ti so **s teoretičnega vidika najbolj zaželeni** na žalost pa najmanjkrat uporabljeni. Slučajni vzorci so tisti, kjer enote izbiramo povsem naključno. Čisto vsaka enota v populaciji ima enako verjetnost, da je izbrana (pridemo že do statistike in kombinatorike).

Imamo N enot v populaciji. Verjetnost da izberemo eno izmed teh N enot je:

$$\frac{1}{N} \leftarrow \begin{array}{l} \text{število enot v} \\ \text{populaciji} \end{array}$$

Če izbiramo enote naključno ima vsaka enota enako verjetnost. Da se tudi matematično dokazati, da ima čisto vsak vzorec N elementov enako verjetnost da je izbran.



Populacija šteje N enot. Vzorec velikosti n (ni en sam!). Število možnih vzorcev je enako:

$$\binom{N}{n}$$

## Verjetnost

$$\frac{1}{\binom{N}{n}}$$

Iz matematične zakonitosti, da ima vsaka enota, oz. iz tega izpeljano tudi vsak vzorec enako verjetnost, je zagotovljena objektivnost: (1) izbire enot v vzorec (če izbiramo povsem naključno je zelo velika verjetnost, da bomo pri tem objektivni). Če imamo populacijo v kateri je  $\frac{3}{4}$  žensk in  $\frac{1}{4}$  moških. Če izbiramo naključno je zelo verjetno, da bo tudi v vzorcu razmerje med moškimi in ženskami približno  $\frac{3}{4}$  proti  $\frac{1}{4}$ .

Če smo objektivni pri izbiranju enot v vzorec, je zelo velika verjetnost, da je ta **VZOREC REPREZENTATIVEN** in zelo dobro odraža značilnosti celotne populacije.

Ne glede na to ali opazujemo vzorec ali populacijo, je naš končni cilj vedno spoznati populacijo. Slučajni vzorci, ki nam omogočajo, da pridemo do reprezentativnih vzorcev, so edina prava pot, da skozi vzorec verodostojno spoznamo populacijo.

Za ocene, ki jih dobimo na podlagi slučajnega vzorca, je mogoče ugotoviti kakšna je njihova kakovost. Zakaj je ta kakovost tako pomembna? Ko spoznamo samo vzorec enot se moramo zavedati, da bo izračun natančen samo za teh 100 enot. Nikoli ne moremo poznati pravega stanja v populaciji.

Če ne moremo presoditi ali je vrednost pravilna ali ne lahko pa z nekimi mehanizmi vsaj ocenimo kakšna je verjetnost, da je ocena pravilna ali ne, nam to ogromno pomeni. Zato je ta značilnost da nam ti vzorci omogočajo neko reprezentativno sliko populacije in drugič da lahko na podlagi teh vzorcev njihove ocene tudi opremimo s podatkom, kakšna je verjetnost, da so te ocene pravilne in kakšna ne, je to ogromnega pomena. Ravno ta preprostost, ta naključna izbira je ključna da so ti vzorci edini, ki s teoretičnega vidika omogočajo, da populacijo objektivno spoznamo.

Vse bo izhajalo iz tega, da bomo predpostavili da je vzorec, s katerim operiramo, enostavni slučajni vzorec. Zakaj so ti vzorci v praksi najmanjkrat uporabljeni? Če želimo oblikovati naključen vzorec iz populacije prebivalcev RS, to pomeni teoretično, da potrebujemo seznam vseh 2Mio prebivalcev (na papirju ali v računalniški obliki), potrebujemo nek generator naključnih števil (excel, funkcija RMD oz random). Če bi kdo objavljala neko javnomnenjsko raziskavo, kjer bo uporabil naključni vzorec, mora iz nabora 2Mio, z nekim računalniškim programom, generirati 1000 naključnih števil. Potem teh 1000 ljudi s tega seznama pobrati in kontaktirati.

Če gremo na ulico pa vsakega desetega izberemo, to v nobenem primeru ni naključen vzorec. Odvisno je kdo je danes sploh hodil v Kopru, drugič je že to, da je bil v Kopru v petek, to ni naključen vzorec. Tehnično zagotoviti slučajnost je včasih zelo težko.

## Neslučajni vzorci-1

---

- Ni znana verjetnost izbore posamezne enote v vzorec
- Ni mogoče izračunati ustreznih kazalcev kakovosti ocene
- Načeloma ni mogoče sklepati iz vzorca na populacijo (subjektivnost izbora enot v vzorec)

**Neslučajna izbira** kjer si v naprej postavimo neke kriterije po katerih vzorec izberemo. V takih primerih ni znana kakšna je verjetnost izbire enote v vzorec. To je tudi ključni dejavnik, da potem tudi za take ocene ni mogoče ugotoviti kako kakovostna je.

1. Imamo oceno za katero nikoli ne moremo vedeti ali je pravilna ali ne (tudi če imamo slučajni vzorec). Lahko pa vsaj približno ocenimo kakšna je verjetnost da je pravilna ali ne;
2. Tu pa tudi tega ne moremo.

### **Zato so s teoretičnega vidika ti vzorci manj zaželjeni:**

1. ker niso objektivni, ker smo imeli neke kriterije po katerih smo izbirali (nismo naključno izbirali);
2. če smo izbirali po nekih kriterijih smo bili na nek način subjektivni. Potem so tudi vse naše ugotovitve bolj ali manj vezane samo na ta vzorec in jih je zelo težko posplošiti na populacijo. Naš končni cilj naj bi pa vedno bil spoznati populacijo.

### **POMANJKLJIVOSTI:**

1. dobimo oceno za katero ne vemo če je pravilna;
2. ne moremo ugotoviti njene kakovosti;
3. dobimo relativno pristransko oceno.

V praksi se kljub temu taki načini zelo pogosto uporabljajo. Če ne drugo je seznam populacijskih enot zelo težko zagotoviti. (Trenutno se v okviru fakultete izvaja aplikativni projekt, katerega cilj je raziskati marketinško kulturo v slovenskih podjetjih, ki se ukvarjajo s storitvenimi dejavnostmi.

- Recimo da je opredeliti, kaj je storitvena dejavnost, enostavno. Pogledamo specifikacijo dejavnosti.
- Dobiti nabor podjetij je tudi še relativno enostavno. Vzamemo nek poslovni imenik in dobimo nabor podjetij v teh dejavnostih.

Če pa je naša enota opazovanja posamezni zaposleni. Kako priti do seznama zaposlenih v vseh teh podjetjih? Razlogi da pridemo do enostavnega slučajnega vzorca, vsaj tistega, ki je teoretično tak kot bi moral biti, je zelo težko.



## Neslučajni vzorci-2

---

- Način oblikovanja:
  - Tipične enote
  - Izbor enot po kvotah
  - Namerno pristranska izbira enot (vnaprej izločimo enote z nezaželenimi lastnostmi)
  - Izbira enot na slepo (primer: izbira zrn v vreči)
  - Izbira najlažje dosegljivih enot

Zato v praksi uporabljajo taki in drugačni načini.

Izbira tipičnih enot, izbira enot po kvotah (10 moških deset žensk)

(Primer: če bi raziskovali kakšne so značilnosti uporabnikov mobilnih telefonov. Pa rečemo da mobilni telefon uporabljajo v glavnem najstniki, ki si dopisujejo z SMS-i, itn. Pozabimo vso ostalo populacijo pa gremo na ulice in bi samo tipične, ki stalno uporabljajo mobilni aparat. Po nekih predsodkih bi se omejili za neke tipične uporabnike).

Zelo podobna je namerno pristranska izbira enot (izbira zrn v vreči). Ni to čisto pravi slučajni vzorec.

Izbira najlažje dosegljivih enot. (Koga bom anketiral ? Seveda svoje sodelavce, svoje starše , svoje sorodnike). Lahko se pričakuje da bodo, zaradi usluge, odgovorili na anketni vprašalnik.

Tak vzorec je pristranski. (gre za krog »naših« ljudi, ki ima lahko čisto neke specifične značilnosti kot pa prebivalstvo celotne Slovenije). Če imamo nek pristranski vzorec z nekimi specifičnimi značilnostmi, bodo značilnosti, ki jih bomo za ta vzorec ugotovili, povsem drugačne kot so značilnosti za populacijo vseh Slovencev. Rezultati bodo imeli zelo majhno uporabno vrednost. Nihče jih ne bo mogel uporabiti, ker niso za splošno uporabni rezultat.

## Osnove vzorčenja

---

- Število možnih vzorcev (brez ponavljanja)

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

- (Vzorčna) ocena parametra

Imamo **populacijo** velikosti **veliki N** iz katere imamo namen izbrati **vzorec mali n** enot.

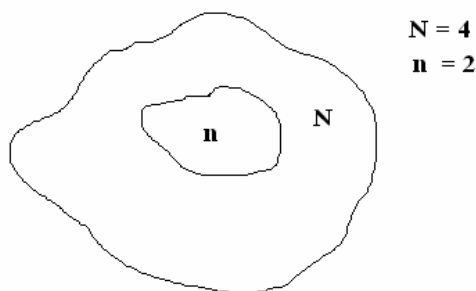
## Vzorčna ocena parametra-3

- Iz populacije 4 enot lahko oblikujemo 6 vzorcev velikosti 2.

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{24}{2 \cdot 2} = 6$$

- Ocena povprečja na podlagi vzorca

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$



$$\binom{4}{2} = \frac{4!}{2! 2!} = \frac{24}{4} = 6$$

Pustimo na strani vprašanje kako določiti velikost vzorca. Je takšen kot je (n enot). Iz populacije N lahko izberemo ogromno različnih vzorcev. Koliko je teh možnih vzorcev? Na podlagi binomskega simbola, ki nam pove kombinacije mali n elementov iz velikih N enot. Zgolj 4 enote, mali n je 2. V tako majhni populaciji in ob tako majhnem vzorcu imamo že šest različnih načinov kako lahko oblikujemo vzorec.

Predstavljajmo si da je:  $N=100.000$  enot, v vzorec izberemo  $n=1000$  enot.

$$\binom{100\ 000}{1000}$$

Dobimo zelo veliko število.

Bolj pomembno je, da vsak izmed teh vzorcev »producira« drugačno oceno.

## Vzorčna ocena parametra-1

---

- Primer: Opazujemo višino bruto plače pri 4 zaposlenih v podjetju X v preteklem mesecu (glej priloženo Excelovo datoteko)

Zaposleni	Plača(SIT)
1	380.000
2	450.000
3	365.000
4	310.000

Imamo populacijo štirih delavcev. Če bi vse štiri delavce povprašali po njihovi plači in bi poznali podatke za celotno populacijo bi izračunali povprečno plačo.

Vse seštejemo in delimo z n. Ugotovimo da je bila njihova povprečna plača 376.250. Temu pravimo **natančna populacijska vrednost**.

## Vzorčna ocena parametra-2

---

- Populacijsko povprečje:

$$\bar{Y} = \frac{\sum_{i=1}^N y_i}{N}$$

- Povprečna bruto plača 4 zaposlenih podjetja X je v preteklem mesecu znašala 376.250 SIT.

Vrnimo se k našim vzorcem.

## Vzorčna porazdelitev ocen parametra

---

Vzorec (zaposleni)	Povprečna plača (SIT)
1,2	415.000
1,3	372.500
1,4	345.000
2,3	407.500
2,4	380.000
3,4	337.500

Imamo šest možnih vzorcev. Poglejmo kakšne vrednosti dobimo na podlagi različnih vzorcev. 376.250 je prava vrednost. Na podlagi prvega vzorca dobimo vrednost, ki je nekoliko večja. V praksi, če bi to bil realen primer bi izmed teh možnih vzorcev dobili v roke enega samega. Kakšna je verjetnost da dobimo en vzorec?

Izmed vseh možnih vzorcev razpolagamo z enim samim, na podlagi tega vzorca dobimo eno izmed možnih vrednosti za katero nikoli ne vemo če je prava.

Slučajni vzorci nam pa omogočajo orodje s pomočjo katerega lahko ocenimo kako kakovostna je ta ocena.

## Pričakovana vrednost ocene parametra in vzorčna varianca-1

---

- Pričakovana vrednost ocene parametra:

$$E(\bar{g}) = \frac{\sum_{v=1}^{\binom{N}{n}} g_v}{\binom{N}{n}} \quad E(\bar{y}) = \frac{\sum_{v=1}^{\binom{N}{n}} y_v}{\binom{N}{n}}$$

- Vzorčna varianca:

$$Var(\bar{g}) = \frac{\sum_{v=1}^{\binom{N}{n}} (g_v - E(\bar{g}))^2}{\binom{N}{n}} \quad Var(\bar{y}) = \frac{\sum_{v=1}^{\binom{N}{n}} (y_v - E(\bar{y}))^2}{\binom{N}{n}}$$

Temu orodju pravimo **VZORČNA VARIANCA** oz. iz nje izpeljana še bolj pomemben kazalec, ker se v praksi to zelo pogosti uporablja, **STANDARDNA NAPAKA OCENE PARAMETRA**.

## Vzorčna varianca in standardna napaka ocene parametra

---

- Vzorčna varianca

$$Var(\bar{y}) = \frac{\sum_{v=1}^N (\bar{y}_v - E(\bar{y}))^2}{\binom{N}{n}} = 830.728.167$$

- Standardna napaka ocene parametra

$$SE(\bar{y}) = \sqrt{Var(\bar{y})} = 28.822,4$$

To je pojem, ki nam na nek način kaže kako kakovostne so naše ocene. Preden razumemo na kaj se ta kakovost nanaša, moramo razumeti kaj pomeni **standardni odklon**?

**STANDARDNI ODKLON nam kaže variabilnost spremenljivke v populaciji.** (Primer: v razredu nas je 10. Vsakega bi vprašal kakšno ima višino plače. Če bi izračunal kakšne so razlike v plači med posamezniki bi to razliko izrazil s standardnim odklonom. Če bi bile razlike med posamezniki velike bi bil standardni odklon velik, če bi bile razlike majhne bi bil standardni odklon majhen).

Enako kot nam standardni odklon kaže variabilnost spremenljivke nam standardna napaka kaže variabilnost potem.

Imamo šest različnih vzorcev in te ocene se med seboj razlikujejo. Koliko se te ocene med sabo razlikujejo nam kaže nam kaže standardna napaka ocene (to nam kaže vzorčna varianca iz nje se izračuna standardni odklon in se uporablja standardno napako ocen, ki je pač bolj pogosta).

Zakaj je to kazalec kakovosti?

Če bi bile te ocene med seboj zelo podobne bi to pomenilo, da je variabilnost ocen majhna. Pomeni da so ocene bolj kakovostne. Če so si ocene med sabo zelo podobne je potem praktično vseeno kateri vzorec dobimo. Taka ocena je zelo malo tvegana. Bolj bi bile ocene med sabo različne manj vseeno je kateri vzorec dobimo.

Večja kot je standardna napaka bolj se ocene med sabo razlikujejo, bolj tvegana je ta ocena in s tem je posledično manjša kakovost take ocene. In obratna logika. Bolj ko so si ocene med sabo podobne manjša je variabilnost ocen, bolj vseeno je kateri vzorec dobimo, ker so si med sabo podobni, z nobeno nič ne tvegamo, ocena je bolj kakovostna.

## Ocenjevanje parametrov

---

- Stopnja tveganja  $\alpha$  in stopnja zaupanja ( $1 - \alpha$ )

---

- Ocenjevanje parametrov z mejami (intervali) zaupanja (veliki vzorci)
- Statistično preskušanje domnev o vrednostih parametrov (mali vzorci)

Gre za predpostavko, ki pravi sledeče:

Če imamo dovolj velike vzorce (v teoriji naj bi dovolj velik vzorec predstavljal vsaj 100 enot) se vzorčne ocene parametra porazdeljujejo normalno. Kaj to pomeni?

Če bi imel v ozadju 100.000 ljudi (za D.N. izračunaj koliko različnih vzorcev velikosti 1000 ali 100 ljudi, pomeni da dobimo toliko različnih ocen). Če bi te ocene skušali na nek način prikazati kakšna je verjetnost da kakšna ocena pojavi, bi ugotovili, da se vse te ocene porazdeljujejo v obliki normalne porazdelitve. Velika večina ocen bi bila točno tam, kjer je tudi v resnici prava vrednost, ki je mi sicer ne poznamo (ampak teoretično). Tam kjer je prava vrednost bi bilo največ ocen. Ustrezno bi se porazdeljevale okrog aritmetične sredine v intervalu aritmetična sredina minus ena standardna napaka plus ena standardna napaka.

Standardna napaka je nekakšen standardni odklon za ocene.

**STANDARDNI ODKLON kaže variabilnost spremenljivke.**

**STANDARDNA NAPAKA kaže variabilnost ocen.**

Povprečje je prava vrednost, minus ena standardna napaka, povprečje plus ena standardna napaka. V tem intervalu naj bi ležalo 68% vseh vzorcev. Izmed vseh vzorcev je 68% takih, ki nam dajo oceno, ki je zelo blizu pravi vrednosti.

Lahko rečemo tudi: 68% verjetnost je da dobimo oceno, ki je nekje v intervalu od prave vrednosti navzdol in navzgor za eno standardno napako. Od prave vrednosti minus dve standardni napaki navzgor in plus dve standardni napaki navzdol leži 95% vseh vzorcev oz. od prave vrednosti minus tri standardne napake in plus tri standardne napake leži 99% (praktično skoraj vsi vzorci).

To je osnova ko dobimo neko oceno. Za to oceno bomo ocenili standardno napako in iz tega izpeljali mehanizem, da bomo ocenili kakšna je verjetnost, da je ta ocena pravilna.

Pri tem se ves čas naslanjamo na predpostavko, zaenkrat si bomo pogledali pristop ko imamo dovolj velike vzorce, kasneje tudi pristop kaj pa če nimamo dovolj velikih vzorcev (v praksi se dogaja tudi to).

Kako se lotimo nekega ocenjevanje. Želimo izračunati povprečno plačo za nek vzorec.

Kako bomo to naredili?

## Ocenjevanje parametrov

---

- Stopnja tveganja  $\alpha$  in stopnja zaupanja ( $1 - \alpha$ )
- 
- Ocenjevanje parametrov z mejami (intervali) zaupanja (veliki vzorci)
  - Statistično preskušanje domnev o vrednostih parametrov (mali vzorci)

Dva pristopa:

1. **OCENJEVANJE INTERVALOV Z OBLIKOVANJEM INTERVALOV ZAUPANJA oz. oblikujemo neke meje**, v katerih je zelo velika verjetnost, da se ocena nahaja. To je pristop ko razpolagamo z velikimi vzorci (po teoriji so to vsi tisti vzorci ko imamo več kot 100 enot).

Zakaj gre? Interval zaupanja je ocena, ki jo dobimo v obliki nekega intervala in v tem intervalu z neko verjetnostjo napovemo, da se nahaja tisto kar ocenjujemo. Tega ne poznamo, zato na podlagi naše ocene bolj ali manj sklepamo. S pomočjo mehanizma intervala zaupanja postavimo nek interval v katerem z določeno verjetnostjo pričakujemo da ta naša ocena je. Neko tveganje vedno je, česar se moramo zavedati.

Tveganje tudi vpliva na natančnost. Če želimo manj tvegane ocene bodo te tudi manj natančne in obratno, če želimo bolj natančno oceno moramo biti pripravljeni za prevzem večjega tveganja (gre za nek trade off).

V glavnem se bomo ukvarjali z dvostranskimi ocenami. Gre za to da postavimo neko spodnjo mejo, neko zgornjo mejo in rečemo. Na tem intervalu je zelo velika verjetnost da leži tisto kar ocenjujemo.

Kar naj bi se naučili je oblikovati **intervalno oceno za aritmetično sredino**, oblikovati **intervalno oceno za vsoto vrednosti** in oblikovati **intervalno oceno za delež**. Gre za pristope, ki so relativno enostavni.

2. **PREIZKUŠANJE DOMNEV O VREDNOSTIH PARAMETROV** - pristop kjer operiramo z majhnimi vzorci (pomeni teoretično manj kot 100, v praksi je lahko 50, 40, 30, 20). Ocenjevanje parametrov z intervali zaupanja se tu ne obnese.

**Temu pravimo tudi preizkušanje domnev oz. testiranje hipotez.**

V praksi veliko bolj uporaben pristop.



## Vrste vzorčnih ocen

---

- Točkovna ocena:  $g$
- Intervalna ocena:
  - Dvostranska ocena
$$g_s < G < g_{s,\alpha}$$
$$g - z * SE(g) < G < g + z * SE(g), \alpha$$
  - Enostranska ocena
$$g_s < G, \alpha \text{ oziroma } g - z * SE(g) < G, \alpha$$
$$G < g_{s,\alpha} \text{ oziroma } G < g + z * SE(g), \alpha$$

**G** - pomeni **tisto kar ocenjujemo** in to je **populacijska vrednost parametra**. (Primer: če ocenjujemo povprečno plačo za neko populacijo delavcev, je torej to  $(\bar{Y})$ , oznaka za povprečje. Če ocenjujemo vsoto vrednosti bo ta **G** vsota vrednosti. To je neka generična oznaka.

**g** – je **tisto kar smo ocenili** (tisto kar smo ocenili ampak na podlagi tega kar smo ocenili nekaj ocenjujemo).

Ocenili smo zgolj povprečje na podlagi vzorca, na podlagi tega pa skušamo ugotoviti kakšno je v resnici povprečje v populaciji.

Ocenjujemo povprečje zgolj na podlagi vzorca ampak si s tem želimo samo pomagati, ker na populaciji tega ne moremo opazovati, da bi sklepali kakšno je v resnici povprečje v celotni populaciji.

Intervalna ocena je v taki obliki, da moramo od naše ocene nekaj odšteti in naši oceni moramo nekaj prišteti. **SE**, ki se tu pojavlja je oznaka za standardno napako (SE – standard error).

$$g - z * SE(g) < G < g + z * SE(g), \alpha$$

Standardna napaka igra svojo vlogo.

Če smo rekli, da je standardna napaka ocene nek kazalec kakovosti je tu lepo razvidno, da ta standardna napaka vpliva na natančnost ocene.

Dobimo neko oceno parametra. Če je ta standardna napaka velika, pomeni da moramo od naše ocene veliko odšteti in veliko prišteti. Pomeni da dobimo širok interval. In obratno. Če je standardna napaka ocene majhna, moramo malo odšteti pa malo odšteti, dobimo ozek interval. Manjša kot je standardna napaka, bolj ozek interval dobimo, večja kot je standardna napaka, širši interval dobimo.

Logično je, da ožji kot je interval bolj natančna je ocena.

*(Primer: ocenjujem, da je povprečna plača zaposlenih v tekstilni industriji, v Sloveniji, med 750 in 810 EUR; kar je zelo majhna razlika, kot pa če rečemo, da se strinjamo, da je povprečna plača zaposlenih v tekstilni industriji v Sloveniji med 600 in 900 EUR. Ta ocena je zelo nenatančna. Živeti s 600 EUR/mesec ali pa z 900EUR/mesec je pa velika razlika).*

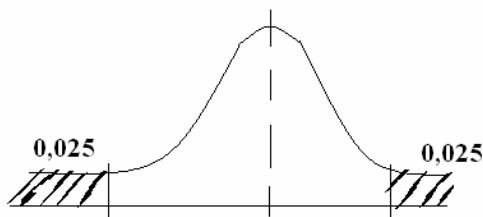
Standardna napaka je eden izmed temeljnih kazalcev kakovosti. Večja kot je standardna napaka manj natančno oceno dobimo (dobimo širši interval) in obratno manjša kot je standardna napaka (dobimo ožji interval), bolj natančna je ocena in zaradi tega bolj kakovostna.

**z je vezan na STANDARDIZIRANO NORMALNO PORAZDELITEV.**

z je izveden iz procentov na Gaussovi krivulji.

V praksi je **zgornja stopnja tveganja**  $\alpha = 5\% = 0,05$ , da so take ocene še uporabne. Nad to stopnjo se ne gre, ker je to že preveliko tveganje.

Če to apliciramo na normalno porazdelitev.



2,5% levo in desno, ker so zelo malo verjetne. Koliko so verjetne nam kaže višina. Skrajnosti se nekako odrečemo, ker so malo verjetne.

Rečemo: Z 5% tveganjem in s 95% verjetnostjo napovedujem, da se moj parameter nahaja na tej oceni.

**z-ju ustreza vrednost 1,96.** To so **standardizirane normalne porazdelitve.**

Naša naloga bo da ocenimo katerikoli že parameter na podlagi vzorca. Za to oceno dobimo standardno napako in če vzamemo stopnjo tveganja 0,05.

*Obrazce se na izpitu dobi kot prilogo. Obrazci, ki so vezani na kombinatoriko in verjetnost pa običajno ni zraven.*

## Ocenjevanje parametrov-3

- Primer: Proučevali smo mesečno porabo sladkorja pri 4-članskih gospodinjstvih v RS v juniju 2002. V ta namen smo opazovali vzorec 75 4-članskih gosp. (glej priloženo Excelovo datoteko).
- Na podlagi izbranega vzorca oceni povprečno poraba sladkorja 4-članskih gosp. v RS. Oceno podaj pri stopnji tveganja  $\alpha=0,05$ .
- Ob upoštevanju, da je bilo v letu 2002 158.145 4-članskih gosp. (vir: Popis 2002) oceni njihovo skupno porabo sladkorja.
- Kolikšen delež vseh 4-članskih gospodinjstev je porabilo manj kot 5 kg sladkorja?

Zanima nas poraba sladkorja v slovenskih 4-članskih gospodinjstvih. Če želimo biti natančni moramo to populacijo natančno opredeliti. Gre za 4-članska gospodinjstva (stvarna opredelitev), to so gospodinjstva, katerih člani imajo stalno prebivališče na območju RS (krajevna opredelitev). Časovna opredelitev, da smo ta gospodinjstva opazovali dovolj natančno. Ker so gospodinjstva tudi realne enote bi moralo biti to opredeljeno natančno na dan 30.6.2002.

Pri tej populaciji smo pa opazovali spremenljivko (značilnosti posameznih enot) poraba sladkorja v mesecu juniju 2002. Spremenljivka se res nanaša na cel mesec ampak smo to spremenljivko opazovali pri gospodinjstvih, ki so obstajala na dan 30.6.2002.

Če bi šli pogledati na Statistični urad je bilo teh gospodinjstev v letu 2002 158.158. Tu se pragmatično odločimo da bomo izbrali naključno izbran vzorec 75, ki ni velik, delajmo se kot da je.

Na podlagi tega vzorca bomo skušali oceniti povprečno porabo sladkorja 4-članskega gospodinjstva. (excelova Tabela 1). Navedenih je vseh 75 vrednosti. Naša naloga je na teh 75 vzorčnih enotah oblikovati intervalno oceno, ki naj bi predstavljala oceno za vseh 158.145 gospodinjstev.

### MEHANIZEM OBLIKOVANJA INTERVALNE OCENE

Najprej zapišemo splošno. Ocenjujemo populacijsko vrednost parametra, kdaj bi se nahajala na intervalu od točkovne ocene (g)...

splošni zapis:

$$g - z * SE(g) < G < g + z * SE(g), \alpha$$

ker nas naloga sprašuje, da naj ocenimo povprečno porabo sladkorja. Zato namesto g-ja pišemo parameter kot povprečje. To povprečje naj bi se nahajalo na intervalu med našo oceno povprečja minus z krat standardna napaka te ocene oz. oceno povprečja plus z krat standardna napaka te ocene povprečja.

$$\bar{y} - z * SE(\bar{y}) < \bar{Y} < \bar{y} + z * SE(\bar{y}), \alpha$$

## Ocenjevanje parametrov-1

---

- Ocena aritmetične sredine

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad SE(\bar{y}) = \frac{\sigma_y}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad se(\bar{y}) = \frac{s_y}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

- Ocena variance in stand. odklona

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \quad s_y = \sqrt{s_y^2}$$

Potrebujemo:

1. **koliko znaša povprečna ocena**

Seštejemo vrednosti pri vseh enotah in to delimo s številom vseh enot v populaciji

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = 4,91$$

čisto enak obrazec sledi

$$\bar{Y} = \frac{\sum_{i=1}^N y_i}{N}$$

Povprečje, ki ga dobimo na teh 75. podatkih je 4,91. (Naša ocena povprečja).

Pomeni da teh 75 gospodinjstev, ki smo jih opazovali, v povprečju porabi 4,91kg sladkorja v nekem mesecu. To pa ni naš končni namen. Ta je, oceniti koliko vsa gospodinjstva (158.145) porabijo. Ker operiramo z enim samim vzorcem in nikakor ne moremo presoditi ali je ta vrednost pravilna ali ne. S to vrednostjo kot tako si nič ne pomagamo. Zato je priročen mehanizem intervalne ocene, kjer potrebujemo še ostala dva elementa.

Naloga pravi naj to oceno oblikujemo pri stopnji tveganja  $\alpha=0,05$  iz tega avtomatsko sledi, da je  $z=1,96$ .

2. **koliko znaša z**  
pri  $\alpha=0,05 \rightarrow z=1,96$

3. **koliko znaša SE (standardna napaka)**

standardna napaka naj bi se izračunala tako da standardni odklon (za populacijo) delimo s korenomo velikosti vzorca krat še nek korekcijski faktor. Ta korekcijski faktor se običajno izpušča zato ker če je mali n v odnosu do velikega N (velikost vzorca glede na velikost populacije relativno majhna, kar v tem primeru je 75 glede na 158.145) relativno majhen. V teh primerih je člen pod korenomo blizu 1 in če nekaj blizu 1 korenimo dobimo še bolj blizu 1, potem je ta člen itak skoraj 1 in se v praksi ta člen izpušča.

$$SE(\bar{y}) = se(\bar{y}) \approx \frac{s_y}{\sqrt{n}} = 0,036 \quad (\text{standardna napaka, ki kaže kakovost ocene})$$

Nastopi vprašanje. Če potrebujemo standardni odklon za populacijo. Kako naj do njega pridemo, če populacije ne poznamo.

Zato vidimo da tudi standardno napako kot tako, če jo želimo uporabljati običajno v praksi moramo oceniti in jo zato tudi pišemo z malo.

Informacija: Vse kar označujemo z velikimi črkami se nanaša na populacijo, vse kar pa z malimi, se nanaša na vzorec.

**N** je število enot v populaciji.

**n** je število enot v vzorcu.

$\bar{Y}$  je povprečje v populaciji.

$\bar{y}$  je povprečje v vzorcu.

Standardne napake natančno kot take ne bomo izračunali ampak jo bomo zgolj ocenili. Vzamemo najbolj skrajno desno različico (standardna napaka je približno enaka oceni standardnega odklona deljeno s korenomo velikosti vzorca).

$$se(\bar{y}) \approx \frac{s_y}{\sqrt{n}}$$

$\sigma_y$  je standardni odklon za populacijo,  $s_y$  je pa standardni odklon ocenjen na podlagi vzorca.

Excel ima vgrajeno funkcijo STDEV (standard deviation). Za izračun je bila uporabljena ta funkcija, dobljen je bil standardni odklon, standardni odklon je bil deljen s korenomo velikosti vzorca in dobimo standardni odklon.

Rečemo: Ocenjujem, da je povprečna poraba sladkorja na intervalu:

$$4,91 - 1,96 * 0,036 < \bar{Y} < 4,91 + 1,96 * 0,036, \alpha = 0,05$$

pri stopnji tveganja

$$4,84 < \bar{Y} < 4,98, \alpha = 0,05$$

**INTERPRETACIJA:**

*Na podlagi vzorca 75-ih gospodinjstev ocenjujemo, da je bila povprečna poraba sladkorja med vsemi štiričlanskimi gospodinjstvi v RS v mesecu juniju 2002 med 4,84 in 4,98 kg.*

*Oceno postavljamo pri stopnji tveganja  $\alpha = 0,05$ .*

Na podlagi vzorca ocenjujemo za celotno populacijo. Interval, ki smo ga dobili 4,84 4,98, nekje na tem intervalu (kjerkoli) naj bi se nahajala prava vrednost aritmetične sredine.

Kaj nam ta stopnja tveganja pove?

Vseh vzorcev je:

$$\binom{N}{n}$$

Izmed teh vzorcev je  $\alpha$  takih, na podlagi katerih dobimo napačno oceno.

$1 - \alpha$  je pa takih, na podlagi katerih dobimo pravo oceno.

To si mi zadamo sami ali pa nam da naročnik. Kolika je še tista zgornja sprejemljiva stopnja tveganja. Ta stopnja tveganja ne predstavlja nič drugega kot to kakšen je odstotek vzorcev med vsemi možnimi, ki nam dajo napačno oceno. Zakaj to pomeni tveganje?



Imamo množico vseh možnih vzorcev, med njimi je 5% takih, ki nam dajo napačno oceno in 95% takih, ki nam dajo pravo oceno. Mi naključno izbiramo enega izmed vzorcev in nikoli ne vemo katerega imamo v rokah. Lahko imamo vzorec, ki je iz množice na desni (v tem primeru je naša ocena pravilna), lahko pa iz množice na levi (v tem primeru je ta ocena napačna). Zato ker ne vemo ali je naša ocena pravilna ali ne, je treba zraven vedno napisati *Oceno postavljamo pri stopnji tveganja  $\alpha = 0,05$* . To uporabniku da informacijo da mi ocenjujemo da je povprečje na intervalu, ki smo ga dobili (4,84 in 4,98) ampak obstaja 5% verjetnosti, da smo s to oceno zgrešili. Gotovosti tudi v vzorčenju ni.

Če bi nekdo dejal da je zanj teh 5% preveč in bi želel bolj gotovo oceno, lahko stopnjo tveganja zmanjšamo. Rečem, oblikoval oceno, ki bo tvegana samo 1%. V takem primeru se nam z poveča interval pa razširi. Res dobimo manj tvegano oceno ampak je ocena tudi manj natančna.

Kaj vse vpliva na natančnost ocene (na širino intervala zaupanja)?

1. **stopnja tveganja**. Nižja kot je  $\alpha$  se z poveča, povečal se bo odklon, natančnost se bo zmanjšala.

$\downarrow \alpha \rightarrow \downarrow z \rightarrow \uparrow \text{odklon} \rightarrow \downarrow \text{natančnost}$

## 2. variabilnost v populaciji

$se(y) = \frac{sy}{\sqrt{n}}$  standardna napaka za povprečje

Ta je približno enaka razmerju med standardnim odklonom in korenem velikosti vzorca. Tu notri sta pa skrita še ostala dva dejavnika, ki vplivata na natančnost.

Drugi dejavnik je standardni odklon in ta nam v bistvu samo odraža, kakšna je variabilnost v populaciji.

Če se naša gospodinjstva med seboj bolj razlikujejo po porabi sladkorja bo večji standardni odklon in se bodo bolj naše ocene povprečne porabe bolj razlikovale in obratno. Če so naša gospodinjstva med sabo zelo homogena (vsak porabijo približno enako sladkorja), bodo ne glede na to kateri vzorec izberemo, ocene s tem približno enake, pomeni manjši standardni odklon, manjša bo standardna napaka.

Večja kot je variabilnost v populaciji vpliva to na večjo standardno napako, to pomeni manjša natančnost.

$\uparrow \text{variabilnost v populaciji} \rightarrow \uparrow se \rightarrow \downarrow \text{natančnost}$

Na ta dejavnik mi nimamo nikakršnega vpliva. To kakšna je variabilnost v populaciji je taka kot pač je.

## 3. velikost vzorca

Edini dejavnik s katerim vsaj teoretično lahko manipuliramo (v praksi to takoj potegne za sabo denar in čas) je pa **n**.

Večji kot je n (večji vzorec kot vzamemo), pomeni večji imenovalec, pomeni da se sama standardna napaka zmanjša, s tem se odklon zmanjša, je ta ocena bolj natančna (gledamo enačbo točke 2).

$\uparrow \text{velikost vzorca}(n) \rightarrow \downarrow se \rightarrow \uparrow \text{natančnost}$

Večji kot je vzorec manjša je standardna napaka pomeni večja je natančnost.

Če bi nekdo želel doseči bolj natančno oceno je odgovor lahko OK ampak bomo potrebovali toliko večji vzorec. Ta toliko večji pa ni tako enostavno, ker imamo spodaj kvadratni koren. Pomeni, če želimo dvakrat manjšo standardno napako potrebujemo štirikrat večji vzorec. Če želimo desetkrat manjšo standardno napako potrebujemo stokrat večji vzorec. To pa seveda ni zastoj in enostavno.

Teoretično je to dejavnik s katerim lahko vplivamo.

Če kdo vpraša s čim lahko povečamo natančnost, je najbolj prost odgovor:

- s povečanjem velikosti vzorca;
- stopnjo tveganja v omejenem obsegu (lahko povečujemo stopnjo tveganja vse dokler ne pridemo do zgornje sprejemljive, da je takšna ocena še prava);

## Ocenjevanje parametrov

---

- Stopnja tveganja  $\alpha$  in stopnja zaupanja  $(1 - \alpha)$
- Ocenjevanje parametrov z mejami (intervali) zaupanja (veliki vzorci)
- Statistično preskušanje domnev o vrednostih parametrov (mali vzorci)

(vrnemo se na prosojnico) Ocenjevanje parametrov z mejami zaupanja (samo pri velikih vzorcih). Zakaj? Zato ker pri malih vzorcih postane standardna napaka tako velika, da je ocena tako široka, da je praktično neuporabna.

Na vsakem vzorcu smo ocenili kakšen je interval. S 5% tveganjem napovedujemo, da bo povprečno gospodinjstvo porabilo med 4,84 in 4,98kg sladkorja.

### 3. alineja naloge

Ob upoštevanju, da je bilo v letu 2002 vseh gospodinjstev 158.145, oceni njihovo skupno porabo sladkorja.

Če izhajamo iz obrazcev za aritmetično sredino v populaciji se ta obrazec glasi, da je **aritmetična sredina** enaka **vsoti opazovanih vrednosti** deljeno z **velikostjo populacije** (to je definicija). Vsota vrednosti je naš **veliki Y**. Če enačbo "premešamo" ugotovimo da je Y **aritmetična sredina** krat N.

$$\bar{Y} = \frac{(i=1 \text{ do } N) \sum y_i}{N} = \frac{Y}{N}$$

$$Y = (i=1 \text{ do } N) \sum y_i = \bar{Y} * N$$

Če to razmerje drži za populacijo bo zelo verjetno na podlagi vzorca podobno.

Velika oznaka – populacija

Mala oznaka – vzorec

Izjemoma pri **vsoti vrednosti** kljub temu ohranimo veliko črko in damo zraven zaznambo da je to ocena na podlagi SL (slučajnega vzorca).

$$Y_{SL} = N \cdot \bar{y}$$

ocena na podlagi slučajnega vzorca

( $\bar{y}$  vsota verjetnosti)

V ozadju je bila storjena neka napaka, je standardna napaka (**standard error**) te ocene je N krat standardna napaka ocene pri povprečju:



$$se(Y_{SL}) = N \cdot se(\bar{y})$$

Tokrat ocenjujemo vsoto vrednosti. Ne ocenjujemo povprečje.

Potrebujemo **točkovno oceno na podlagi vzorca**, od nje odštejemo **z krat standardna napaka te ocene**, in točkovni oceni prištejem plus z krat standardna napaka. Ocena ki jo oblikujemo pri stopnji tveganja  $\alpha$ .

$$Y_{SL} - z \cdot se(Y_{SL}) < Y < Y_{SL} + z \cdot se(Y_{SL}), \alpha$$

$Y_{SL} = N \cdot \bar{y} = 158.145 \cdot 4,91 = 776.492$  (to je ocena vsote vrednosti)  
točkovna ocena  $Y_{SL}$  je  $N$  krat ocena aritmetične sredine (izračunano v excelovi prilogi).

$se(Y_{SL}) = N \cdot se(\bar{y}) = 158.145 \cdot 0,036 = 5.710,45$   
Standardna napaka za to oceno je pa  $N$  krat standardna napaka ocene.

Ker naloga pravi da tudi tokrat upoštevamo  $\alpha=0,05$ , pomeni da bo tudi tokrat  $z = 1,96$ .

$$776.492 - 1,96 \cdot 5.710,45 < Y < 776.492 + 1,96 \cdot 5.710,45; \alpha=0,05$$

$$11192,482$$

$$11192,482$$

$$765299,518$$

$$787684,482$$

$$765.299 < Y < 787.685; \alpha=0,05$$

Naročnik je rekel, da je stopnja tveganja  $\alpha = 0,05$  pomeni, da je zanj zgornja sprejemljiva stopnja tveganja.

Držati se moramo zgornje meje tveganja zato je pa logika, da **spodnje meje vedno zaokrožujemo navzdol** (pomeni da gremo raje malo v levo), **zgornje meje pa obvezno navzgor** (raje gremo rahlo v desno). S tem sicer resda dobimo malenkost manj natančni interval, ampak, če govorimo o malenkostih bomo pa v vsakem primeru ostali znotraj sprejemljivega centra.

#### INTERPRETACIJA:

***Ocenjujemo da je bila skupna poraba vseh 4-članskih gospodinjstev v RS med 765.299 in 787.685kg sladkorja.***

***Oceno postavljamo pri stopnji tveganja  $\alpha = 0,05$ .***

Me vsemi možnimi vzorci je 5% takih, ki nam dajo napačno oceno, 95% vzorcev pa takih, ki nam dajo pravilno oceno. Žal mi ne moremo vedeti iz katere skupine vzorcev je bil naš vzorec izbran.

## OCENA DELEŽA OZIROMA ODSTOTKA

- Ocena deleža, odstotka in števila enot z določeno vrednostjo

$$p = \frac{n_a}{n} \quad se(p) = \sqrt{\frac{p(1-p)}{n-1}} \sqrt{\frac{N-n}{N}}$$
$$p\% = 100 * p = 100 * \frac{n_a}{n} \quad se(p\%) = \sqrt{\frac{p\%(100-p\%)}{n-1}} \sqrt{\frac{N-n}{N}} = 100 * se(p)$$
$$N_a = N * p \quad se(N_a) = N * se(p)$$

Gre za isto vsebino samo nekoliko drugače se izrazi.

Če govorimo o deležu ga ocenimo tako številu enot damo neko lastnost. Mali  $n_a$  je število enot, ki imajo neko lastnost, primerjamo s številom vseh enot v vzorcu.

Če pa govorimo o odstotku gre pač enostavno za to, da delež še pomnožimo s 100.

Po vsebini pomeni to eno in isto.

Enkrat gre za to da izrazimo vrednost od 0 do 1, enkrat pa od nič do p(?).

Razlika je malenkost v izračunu standardne napake. Čeprav sta v formuli dva korena,

$$se(p) = \sqrt{\frac{p(1-p)}{n-1}} \sqrt{\frac{N-n}{N}}$$

je drugi koren spet nek korekcijski faktor in običajno ta drugi člen izpuščamo.

$$se(p\%) = \sqrt{\frac{p\%(100-p\%)}{n-1}} \sqrt{\frac{N-n}{N}} = 100 * se(p)$$

V spodnjem je p izražen v odstotkih. Med obema obrazcema je minimalna tehnična razlika. Čisto vseno, če gre za vprašanje ocenite delež ali pa odstotek, gre za eno in isto zadevo, samo drugačen način izražanja in tehnična malenkost vpliva na način izračuna.

Poglejmo kako bi se tega lotili v našem primeru.

Ocenimo p% (p odstotek).

p% je na intervalu med točkovno oceno odstotka minus z krat standardna napaka te ocene in na drugi strani točkovna ocena odstotka plus z krat standardna napaka te ocene, pri neki dani stopnji tveganja.

$$p\% - z * se(p\%) < P\% < p\% + z * se(p\%), \alpha$$

Točkovno oceno odstotka dobimo tako, da število enot ustreza naši zeleni lastnosti primerjamo s številom vseh enot v vzorcu in to pomnožimo s 100.

$$p\% = \frac{n_a}{n} * 100$$

v našem primeru pogledjmo kako je zastavljeno vprašanje.

Kolikšen odstotek vseh štiričlanskih gospodinjstev je porabilo manj kot 5kg sladkorja?

(glej Tabelo 1: Poraba sladkorja pri 75 gospodinjstvih; od vrstice 4 pa do vrstice 50, 46 gospodinjstev)

$$p\% = \frac{46}{75} * 100 = 61,3 \%$$

Če 46 gospodinjstev primerjamo z vsemi 75-timi krat 100 dobimo rezultat. Če izračunamo za to oceno še standardno napako (uporabimo samo prvi koren).

$$se(p\%) = \sqrt{\frac{p\% (100 - p\%)}{n - 1}}$$

$$= \frac{61,3 (100 - 61,3)}{75 - 1} = 5,6$$

Imamo točkovno oceno, standardno napako ocene, v nalogi je  $\alpha = 0,05$  je tudi tokrat  $z = 1,96$ .

$$61,3 - 1,96 * 5,6 < P\% < 61,3 + 1,96 * 5,6 ; \alpha = 0,05$$

$$50,2 < P\% < 72,5 ; \alpha = 0,05$$

če izračunamo dobimo spodnjo mejo 50,2 če zaokrožimo navzdol in zgornja meja znaša 72,5 če zaokrožimo navzgor, pri  $\alpha = 0,05$ .

#### INTERPRETACIJA:

**Ocenjujemo, da je bilo med vsemi štiričlanskimi gospodinjstvi v RS, med 50,2 in 72,5 odstotkov gospodinjstev, ki so porabila manj kot 5kg sladkorja. Oceno postavljamo pri stopnji tveganja  $\alpha = 0,05$ .**

---

Treba se je naučiti oceniti aritmetično sredino v intervalu zaupanja, naučiti se oceniti vsoto vrednosti in naučiti se oceniti ali delež ali odstotek, kar je pač eno in isto. Tri različne tipe intervalov zaupanja nekako morate poznati. Ta mehanizem gospodinjstev je ves čas enak. Vedno izhajate iz neke točkovne ocene, od katere odštejete odklon oz. točkovni oceni prištejete odklon. Kako odklon dobite? Tako da upoštevate **z, ki je vezan na tveganje**, največkrat kot rečeno bo  $\alpha = 0,05$  tako, da bo ta **z največkrat 1,96**. (Naslednjič bo pokazano na tablicah kako dejansko **z** ugotovimo, **če slučajno  $\alpha$  ne bi bila 0,05** ampak neka nestandardna).

Standardna napaka se pa podobno kot vsak parameter drugače oceni. Zato imate obrazce. Na izpitu ne dobite podatke v taki obliki. Na izpitu dobite že delne rezultate pripravljene na podlagi programa SPSS (izpis). Iz teh delnih rezultatov potem oblikujete interval.

Izpitno vprašanje je zastavljeno s podobnim besedilom vi bi pa recimo že dobili izračunano standardno napako in če je vprašanje oceniti. Eventualno edino pri deležih je včasih potem treba izračunati. Pri deležih dobite neko tabelico pa recimo a, b, c, jih imate 10, 12, 33, pa pravi oceni delež teh, ki imajo b. Pa sami izračunate  $12/43 + 12 = 55$  dobite **p%** pa iz tega **se(p%)** pa v bistvu to interpretirati. Izračuni so minus, plus, pri deležu je še deljenje pa koren. In to je vrhunec vsega

računanja na izpitu. Vse ostalo je bolj vitalno, da stvari razumete, kasneje, ko bomo prišli do regresijske analize, recimo dobite na izpitu že celotni regresijski model izračunan, vse ocene, samo vprašanje se glasi tako, da morate vi znati iz tega izpisa vzeti neko vrednost, ki jo vprašanje sprašuje in .. Ključni poudarek je na tem, da znate stvari razumeti in jih uporabiti.