3. predavanje RVM – Kvantitativne metode
Borut Kodrič, Koper 27.5.2010

~~~~~~

1. del

Ponovitev:

*Začeli smo s pristopi ocenjevanja parametrov na podlagi vzorčnih opazovanj. Pogledali smo ključne razloge zakaj se v praksi bolj ali manj operira samo z vzorčnimi podatki. Nakazali smo, da se v primeru analize samega vzorčenja podatkov, uporabljata načeloma dva pristopa. En pristop je ko nek parameter ocenimo s pomočjo intervala zaupanja. Naučili smo se oblikovati intervale zaupanja za tri različne parametre. In sicer za **aritmetično sredino**, za **vsoto vrednosti** in pa za **delež enot**.*

INTERVALI ZAUPANJA SO NAČELOMA UPORABNI SAMO V PRIMERU, KO RAZPOLAGAMO Z VELIKIMI VZORCI. *Tisto kar želimo oceniti je prava vrednost, ki je ne poznamo.*

Kot izhodišče vzamemo **TOČKOVNO OCENO**. *Ta je kakršna je. Drugi element, ki vpliva na natančnost, je* **ODKLON ZAUPANJA**. *Ta je odvisen od stopnje tveganja, ki si jo zadamo.*

$\bar{y} - z * se(\bar{y}) < \bar{Y} < \bar{y} + z * se(\bar{y}), \ \alpha=0,05$

Dovolj manipulacije nimamo, ker smo omejeni z zgornjo mejo, če želimo, da je to tveganje še toliko sprejemljivo, da so take ocene tudi uporabne.

*Edini element, s katerim lahko vplivamo na natančnost, je element, ki vpliva neposredno na standardno napako in ta je **velikost vzorca**.*

STANDARDNA NAPAKA ZA ARITMETIČNO SREDINO *je približno enaka* **OCENA STANDARDNEGA ODKLONA** *deljeno s korenom* **VELIKOSTI VZORCA**.

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

se – **s**tandard **e**rror

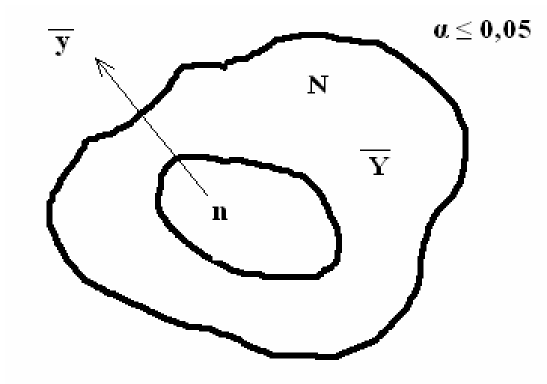
*Pri variabilnosti v populaciji tudi nimamo kaj manipulirati. V končni fazi je res samo velikost vzorca tista, ki opredeljuje odklon zaupanja in s tem samo natančnost ocene. **V praksi se pri majhnih vzorcih (ko je ta n majhen) je tudi imenovalec, kot koren majhne vrednosti, majhen. Pomeni če neko vrednost delimo z majhnim številom dobimo veliko število (glej obrazec zgoraj!).** V takem primeru je ta standardna napaka ocene velika, zato se interval zaupanja razširi. Dobimo tako širok interval zaupanja, da ta nima praktično nikakršne uporabne vrednosti.*

~~~~~~

PREIZKUŠANJE DOMNEV

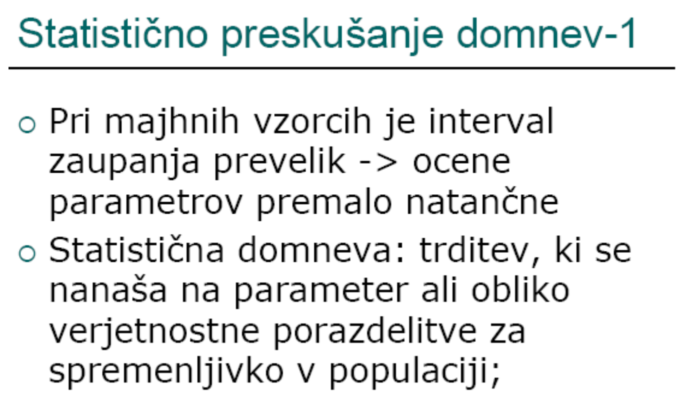
Pristop, ki se v praksi v prevladujoči meri uporablja.

Iz neke **populacije N enot** vzamemo mali **vzorec n enot** in na podlagi njega ocenimo povprečje.



Ocena, ki jo dobimo, na podlagi nekega vzorca, še zdaleč ni nujno enaka (in največkrat tudi ni) pravi vrednosti.

V praksi populacije nikoli ne poznamo, zato je še toliko težje ugotoviti ali je izračunana vrednost prava ali ne. Če bi poznali celotno populacijo, potem lahko v vsakem trenutku rečemo ali številka, ki jo imamo v roki, je prava ocena ali ne.



Čim imamo **opravka z vzorci**, je praktično **vedno prisotno neko tveganje**, da je ocena, ki jo imamo napačna oz. nikoli ne moremo z gotovostjo ugotoviti ali razpolagamo s pravo ali napačno oceno. Da je v praksi še sprejemljivo, to tveganje omejimo na 0,05 oz. 5%.

Statistično preskušanje domnev-2

- Opazovanje vseh enot: ugotovimo ali je trditev **pravilna** ali **nepravilna.**
- Opazovanje vzorca enot: lahko sklepamo ali je trditev **"verjetno pravilna"** ali **"verjetno ni pravilna".**

- V postopku stat.preskušanja domnev preizkušamo ničelno domnevo, ki jo lahko zavrnemo ali sprejmemo.

Domneve, si predstavljamo kot trditev, recimo medijev. To je konkretno zapisana vrednost za nek parameter (nekdo trdi nekaj, nasprotnik trdi nasprotno).

Ničelna in alternativna domneva

- Ničelna domneva (H_0): domneva, ki jo preskušamo na podlagi vzorca

  Primer: $H_0 : M_1 = M_2$

- Alternativna domneva (H_1):
  - Nasprotna ničelni domnevi;
  - Če se izkaže pri preskusu ničelne domneve, da ta verjetno ne velja, potem je smiselno sprejeti alternativno domnevo;

  Primer: $H_1 : M_1 \neq M_2$

Na eni strani bi premier dal izjavo, povprečna plača v RS, v preteklem letu je bila na višini 800 EUR. Nekdo iz opozicije bi pa dejal, da je povprečna plača samo 700 EUR.

ničelna domneva : Ho $\overline{Y} = 700$

alternativna domneva : H1 $\overline{Y} = 800$

Če bi želeli natančno ugotoviti kdo ima prav bi morali vse zaposlene na območju RS (600.000 in več) anketirati ali s strani Davčnega urada dobiti podatke.

Drugi pristop, ki v praksi tudi da zadovoljive rezultate, je da izberemo samo nek vzorec. Na podlagi vzorca pridemo do neke ocene in potem na podlagi te ocene sklepamo kdo ima prav.

Postopek preizkušanja domnev se vedno nanaša na dve domnevi. Eni pravimo ničelna eni alternativna. Ti dve domnevi sta vedno postavljeni na tak način, da sta si v nasprotju. Velja ali ena ali druga ne moreta pa veljati obe hkrati.

Končni rezultat samega preizkusa naj nam bi dal odgovor ali drži ničelna ali alternativna domneva.

Kako domnevi zapišemo je odvisno od vsebine.

Vrste statističnih domnev

- Nesestavljene domneve: za parameter G upoštevana le po ena vrednost
  Primer: $H_0: G = G_0$ in $H_1: G = G_1$

- Sestavljene domneve: bodisi v alternativni, bodisi v obeh domnevah več vrednosti parametra
  Primer: $H_0: G = G_0$ in $H_1: G \neq G_0$ ali
  $H_0: G \leq G_0$ $H_1: G > G_0$

se uporablja veliko pogosteje

NESESTAVLJENI DOMNEVI

sta tedaj ko v ničelno in alternativno domnevo **zapišemo samo eno vrednost.** Tak način je v praksi manj pogost.

SESTAVLJENE DOMNEVE

Ta pristop se veliko pogosteje uporablja. Imamo **ENOSTRANSKE** ali **DVOSTRANSKE PREIZKUSE**.

Za izhodišče vzamemo:

ničelna domneva : Ho $\overline{Y} = 800$

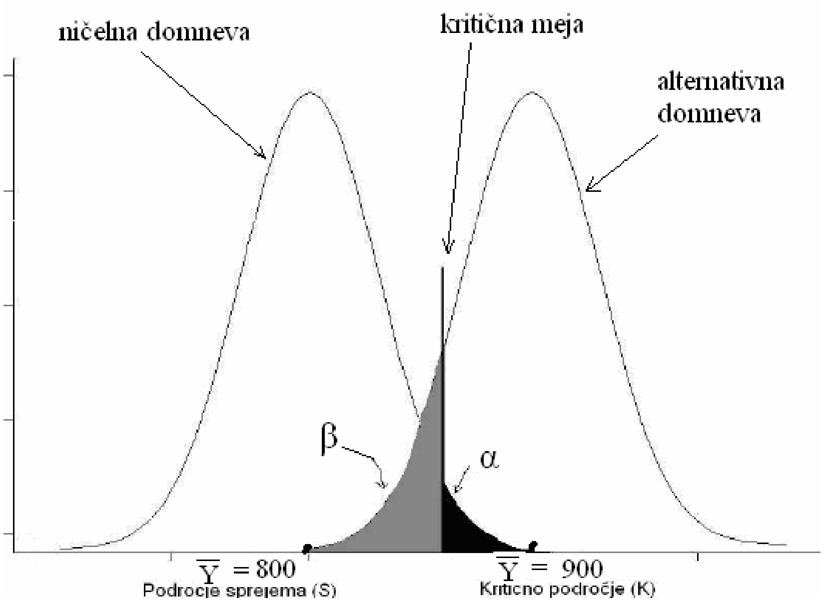
alternativna domneva : H1 $\overline{Y} = 900$

Kadar imamo dovolj velike vzorce (velikost vzorca nad 100), je vzorčna **porazdelitev parametrov normalna**.

Zaradi lažje razlage, v našem primeru predpostavljamo, da imamo opravka z dovolj velikim vzorcem. Lahko predpostavimo, da je vzorčna porazdelitev parametrov normalna okrog vrednosti, ki naj bi veljala v populaciji.

a) če bi v resnici veljalo stanje, ki je v ničelni domnevi (800, premier), bi to pomenilo, da se bodo ocene na podlagi vseh možnih vzorcev stanja populacije, porazdeljevale tako kot kaže normalna porazdelitev proti vrednosti 800.

b) če bi pa veljalo stanje, ko velja alternativna domneva (900, opozicija), bi, ko bi iz take populacije jemali vzorec, na podlagi vseh možnih vzorcev dobili normalno porazdelitev okrog vrednosti 900.



Prvo ključno za razumeti je, da iz neke **populacije N** lahko vzamemo **vzorec velikosti n** na veliko različnih načinov, točneje na spodnje število možnosti.

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

In toliko kot imate naključnih vzorcev, toliko imate različnih ocen. Te ocene so si med seboj različne. Kakšne pa so nam kaže porazdelitev. Velika večina ocen iz neke populacije, v kateri je aritmetična sredina v resnici 800. Nje sicer ne poznamo ker so vse le teoretične predpostavke. Ko vzorčimo iz take populacije bo velika večina ocen res nekje v bližini 800, malenkost manj, malenkost več. Bo tudi nekaj takih, ki bodo krepko presegale to oceno oz. takih, ki bodo pod to oceno.

Če bi (po analogiji) veljala v realnosti alternativna domneva (kjer je povprečje enako 900), bi nam različni vzorci dajali porazdelitev ocen, kjer bi bila velika večina okrog 900, z nekoliko manjšo verjetnostjo bi bile vrednosti, ki bi močno precenjevale oz. podcenjevale to vrednost.

POSTOPEK:

Iz **populacije N elementov** izberete **vzorec mali n elementov** in na osnovi njega ocenite neko aritmetično sredino, v tem primeru povprečno plačo.

Če boste dobili oceno, ki je recimo enaka 920 je vsem razumljivo, da **bi** na podlagi take ocene **sprejeli alternativno domnevo**. Ta vrednost ni enaka temu (900) ampak nam porazdelitev kaže, da je verjetnost da se taka ocena, na podlagi nekega vzorca, ki je izbran iz te populacije pojavi, relativno visoka. Zelo verjetno je, da iz populacije, v kateri je v resnici povprečje 900, izberete vzorec, ki vam da oceno 920. To je precej realna možnost.

Po enaki logiki, če bi dobili oceni, ki bi bila na 790 ali 820 ali 750. Na podlagi katerekoli teh ocen bi z veliko verjetnostjo predpostavili, da velja situacija, ki je zapisana v **ničelni domnevi**.

Dilema nastopi, če je ocena nekje med obema ocenama. Obstaja neka določena verjetnost, da je ta ocena element leve porazdelitve in obstaja verjetnost, da je ocena element desne porazdelitve.

Bistvo vseh postopkov preizkušanja domnev je samo to, da postavimo mejo (ločnico, ki ji pravimo **KRITIČNA MEJA**), kjer bomo rekli:

Če je ocena tostran meje bomo sprejeli **ničelno domnevo**, če je ocena preko meje bomo sprejeli **alternativno domnevo**.

Vsi naši sklepi bodo vedno opremljeni z nekim tveganjem.

Kritična meja razločuje celotno območje vrednost na dva dela in sicer:

- na sliki levo je območje sprejema
- na sliki desno je pa kritično področje

Vse dokler bo **ocena manjša od vrednosti na kritični** meji bomo z veliko verjetnostjo trdili, da velja situacija, ki je predpostavljena v ničelni domnevi. Lahko se bomo tudi zmotili. Vse vrednosti od meje navzdol lahko predstavljajo element ali ene ali druge porazdelitve. Tej stopnji tveganja pravimo β.

Če pa **vrednost preskoči kritično mejo**. Od tu dalje rečem, da sprejmem alternativno domnevo, ampak spet ta sklep ni gotov, ker noben sklep na podlagi vzorca ne more biti gotov. Stopni tveganja v tem primeru pravimo α.

Za razliko od intervalov zaupanja, kjer smo imeli eno samo stopnjo tveganja (da je intervalna ocena pravilna ali ne), se tu pojavljata dve različni stopnji tveganja.

Stopnja tveganja α se nanaša na verjetnost, da ko zavrnemo ničelno domnevo je v bistvu ta v resnici veljala, smo jo lahko napačno zavrnili. In obratno. Stopnja tveganja β pa v bistvu predstavlja verjetnost, da smo sprejeli ničelno domnevo morali bi pa dejansko sprejeti alternativno.

Pravilni in nepravilni sklepi

Sklep na podlagi vzorca	Dejansko stanje	
	H_0 velja	H_0 ne velja
H_0 velja	Sklep je pravilen **(1-α)**	Sklep ni pravilen – napaka 2.vrste **(β)**
H_0 ne velja	Sklep ni pravilen – napaka 1.vrste **(α)**	Sklep je pravilen **(1-β,moč preizkusa)**

Če ničelna domneva v resnici velja in če mi to s poskusom ugotovimo, bo naš sklep pravilen (obkroženo). Verjetnost da je ta sklep pravilen je $(1 - \alpha)$. α je verjetnost da naredimo napako, $(1 - \alpha)$ je verjetnost, da je sklep pravilen.

Prav tako je naš sklep pravilen, če v resnici ničelna domneva ne velja in če bi mi to pravilno ugotovili (obkroženo).

To sta dva pravilna sklepa. Verjetnost za prvega je $(1 - \alpha)$, verjetnost za drugega je $(1 - \beta)$. V teoriji se to označuje (power of the test).

Poleg teh dveh izidov lahko nastopi, da se z našim postopkom preizkušanja domnev enostavno zmotimo.

Če ničelna domneva ne velja (dejansko stanje), mi jo pa potrdimo da velja (sklep na podlagi vzorca), smo s tem naredili napako 2. vrste β.

Druga možnost napake pa je, da ničelna domneva v resnici velja (dejansko stanje), mi pa tega ne znamo pravilno ugotoviti. Naredimo napako 1. vrste α.

Vsak postopek preizkušanja domnev ima te štiri možne izide. Dva možna izida sta pravilna dva sta pa napačna. Nikoli pa z gotovostjo ne moremo vedeti ali je naša ocena pravilna ali ne, zato je vse sklepe potrebno vedno opremljati z podatkom o tem, kakšna je kljub temu možnost, da se to zgodi.

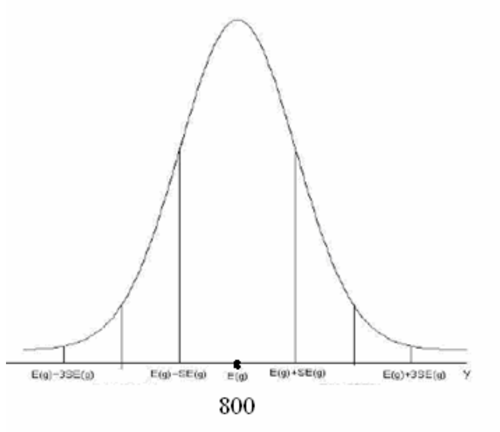
V praksi običajno izhajamo iz zgolj ničelne domneve.

ničelna domneva : Ho $\overline{Y} \geq 800$

alternativna domneva : H1 $\overline{Y} < 800$

Ta postopek je zasnovan tako, da preizkuša NIČELNO in če se le da, naj bi se ničelna zavrnila in sprejela alternativna.

Vrednost, ki smo jo zapisali v **ničelno domnevo** je poznana. Mejna vrednost je 800 zato tudi običajno porazdelitev okrog te ničelne vrednosti znamo narisati. Če bi v resnici veljala ta mejna vrednost 800 potem je ničelna porazdelitev kot spodaj na sliki.



Kaj pa je zapisano v alternativni domnevi? Da je povprečje manjše od 800.

Lege alternativne porazdelitve (v našem primeru prej na desni strani) **v praksi nikoli ne poznamo**. Pomeni da ne moremo natančno oceniti kolikšen je del prekrivanja oz. kolikšna je ta stopnja tveganja. ZARADI TE SPECIFIKE JE V POSTOPKIH PREIZKUŠANJA DOMNEV **STOPNJA β VEDNO NEZNANA**.

Če bi želeli, v nekem trenutku, na podlagi take ocene, sprejeti ničelno domnevo, bi morali ta sklep opremiti s stopnjo tveganja β. Če bi test pokazal da ničelna domneva velja je verjetnost, da smo imeli prav (1 – α), verjetnost da ničelna domneva ne velja pa je β. Ker tega kar bi morali zraven takega sklepa napisati ne poznamo, takega sklepa nikoli ne sprejemamo.

Ko bomo ugotovili, da ničelne domneve ne moremo zavrniti, je kljub temu ne bomo eksplicitno sprejeli. Zato bodo naši sklepi kvečjemu, da bomo ali ničelno domnevo zavrnili in rekli da velja alternativna. Praktično je vedno bolj pomembna tveganost za napako. Ta je pa znana (α), ker nam jo v bistvu predpiše naročnik.

Zato se v praksi vedno skuša postaviti domnevi tako, da tisto kar bi radi potrdili oz. kar se nakazuje, da je večja verjetnost, da bo sprejeto, postavimo v alternativno. V praksi ni čisto vseeno kaj postavimo v ničelno in kaj v alternativno domnevo. V alternativno vedno postavimo tisto kar bi želeli dokazati.

Izida preizkušanja domnev bosta dva.

Kadar bo **ocena v območju sprejema** bomo rekli da **ničelne domneve ne moremo zavrniti**. V tem primeru ne navajamo nikakršnega tveganja ker ga ne poznamo. Pa tudi ni potrebno ker nismo sprejeli nobenega eksplicitnega sklepa.

Kadar bo ocena padla preko **kritične meje** (to vedno določimo na podlagi α) bomo rekli, verjetnost da smo se o tem motili seveda obstaja ampak ker smo rekli, da je α lahko kvečjemu 0,05 ali manj, da v takem primeru velja ničelna domneva je zelo zelo majhna. Rekli bomo: ničelno domnevo zavračamo in sprejemamo alternativno.

Rezime: **Postopek se uporablja kadar imamo majhne vzorce,** zato ker takrat intervali zaupanja zatajijo (postanejo tako široki da praktično nimajo uporabne vrednosti). Zato se v takih primerih raje poslužujemo postopkov preizkušanja domnev. To ne pomeni, da te postopke lahko uporabimo samo na majhnih vzorcih. Nasprotno. Večji kot je vzorec bolj natančni so lahko naši sklepi. Velikost vzorca vpliva tudi na postopek preizkušanja domnev ampak v nekoliko manjši meri. Imajo veliko prednost ker so bolj splošno uporabni kot intervali zaupanja.

Postopek je zasnovan na tem, da med sabo primerjamo dve nasprotni si domnevi, ena je ničelna druga alternativna.

Vsa ta razlaga je bila osnovana na predpostavki, da imamo dovolj velik vzorec.

V PRIMERIH KO IMAMO **MAJHNE VZORCE** (po teoriji 30 ali manj), se uporabljajo neke druge porazdelitve.

Vzorčne porazdelitve pri malih vzorcih-1

- Verjetnost pravilnih in nepravilnih sklepov na podlagi vzorčnih podatkov je mogoče izračunati le, če izhajamo iz vzorčne porazdelitve.
- Če so vzorci majhni (n<30) je treba pri opisu vzorčnih porazdelitev upoštevati druge teoretične porazdelitve

Vzorčne porazdelitve pri malih vzorcih-2

- χ^2 – porazdelitev;
- t – (Studentova) porazdelitev;
- F – porazdelitev;
- Njihova oblika odvisna od velikosti vzorca oziroma stopinj prostosti (m).

Največ se bomo srečevali s **t porazdelitvijo**. Ta je precej podobna normalni porazdelitvi. Tako kot normalna je tudi ta simetrična (zvonasta), za razliko od normalne je ta nekoliko bolj sploščena.

Skupna značilnost vseh teh treh porazdelitev je ta, da se njihova oblika spreminja glede na velikost vzorca.

t-porazdelitev

- Simetrična, a nekoliko bolj sploščena kot standardizirana normalna porazdelitev
- $m=n-1$
- Tabelirane so vrednosti $t_{m,\alpha}$, za katere je verjetnost, da je vrednost slučajne sprem. večja od tabelirane, enaka α

F-porazdelitev

- Osnova pri primerjanju varianc dveh neodvisnih slučajnih vzorcev iz normalne porazdelitve
- $m_1=n_1-1$, $m_2=n_2-1$

Ključno in zelo pomembno je, da boste znali besedilo pravilno razumeti in na podlagi besedila pravilno formulirati ničelno in alternativno domnevo.

Postopek pri statističnem preskušanju domnev - 1

1. Skladno s problematiko postavimo H_0 in H_1.
2. Z ustreznim preizkusom preizkusimo H_0.

Postopek pri statističnem preskušanju domnev - 2

3. Ugotovimo ali je razlika med vzorčno oceno g in vrednostjo parametra v ničelni domnevi G_0 značilna ali ni značilna:
   - ✓ Razlika je značilna, če je vzorčna ocena g v kritičnem področju;
   - × Razlika ni značilna, če je vzorčna ocena v področju sprejema.

(differences are significant, differences are not significant)

`termin značilnost / pomembnost (značilnost naj bi bil hrvatizem)`

V drugem koraku odvisno, od tega na kateri parameter se domneve nanašajo. Vsa dosedanja razlaga je bazirala da se domneva nanaša samo na aritmetično sredino. Domneve lahko preizkušamo za katerikoli parameter. Lahko za aritmetično sredino, za primerjavo med dvema aritmetičnima sredinama, lahko za primerjavo med večjim številom aritmetičnih sredin, lahko je domneva, ki se nanaša na nek regresijski koeficient, lahko domneva, ki bi se nanašala na delež.

Od parametra do parametra se uporabljajo različni preizkusi. Konkretno: za preizkušanje domnev o aritmetični sredini se uporabljata dva preizkusa (z in t preizkus). V praksi pa bolj ali manj samo t preizkus.

Za preizkušanje razlike med dvema aritmetičnima sredinama bomo ravno tako uporabili t preizkus.

Za preizkušanje domnev o razlikah med večjim številom aritmetičnih sredin pa uporabljali F preizkus.

Postopek pri statističnem preskušanju domnev - 3

Če razlika ni značilna običajno le ugotovimo, da ni mogoče zavrniti ničelne domneve. Ničelne domneve običajno ne sprejemamo, ker ne poznamo verjetnosti za napako 2.vrste β.

- Sprejmemo sklep skladno z ugotovitvami v 3. koraku in z opredelitvijo H_0 in H_1.

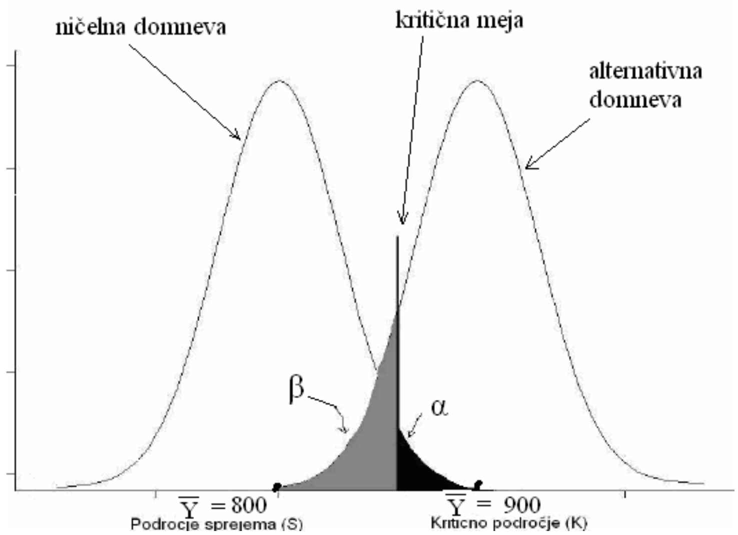
Specifika v koraku 3. Če ocena pade v območje sprejema je naš sklep zgolj ta da ničelne domneve ni mogoče zavrniti. Če pa ocena pade v kritično področje, ničelno domnevo zavrnemo in hkrati sprejmemo alternativno.

V primeru kadar ničelne domneve ne zavrnemo je tudi kljub temu eksplicitno nikoli ne sprejmemo.

Velikokrat boste zasledili ta ocena je značilna, ta ocena ni značilna oziroma v angleškem jeziku

Stopnje značilnosti in stopnja P

- Stopnjo tveganja α, s katero v postopku preskušanja domnev zavrnemo ničelno domnevo, imenujemo stopnja značilnost i.
- V rač.izpisih je navadno navedena točna stopnja tveganja, pri kateri bi zavrnili ničelno domnevo (P, SIGN, Prob.).



Bistvo vsega tega preizkušanja domnev je **ugotoviti kje je tista kritična meja**, do katere bomo bolj verjetno verjeli ničelni oz. preko katere bomo bolj verjetno verjeli alternativni domnevi.

KRITIČNA MEJA SE POSTAVI NA PODLAGI STOPNJE TVEGANJA α.

Naročnik specificira kakšno tveganje je pripravljen sprejeti in vi na podlagi tega tveganja določite kakšna naj bo kritična meja. S postopkom preizkušanja domnev bi ugotovili, če ocena pri taki stopnji tveganja leži v območju sprejema pa ničelne domneve ne zavrnemo. Če ocena leži v kritičnem območju ničelno domnevo zavrnemo in sprejmemo alternativno.

Kako funkcionirajo računalniški programi? Lahko nam izračuna tudi **NATANČNO STOPNJO TVEGANJA.** Samo da razumemo kako do nje pridemo.

To se standardno označuje kot sig ali kot SIGN (significant) – **STOPNJA ZNAČILNOSTI**.

V praksi je α maksimalno 0,05.

Kadar nam bo računalnik izračunal stopnjo značilnosti sig ≤ 0,05 bo to za nas znak, da ocena leži nekje v kritičnem področju, taka kot bi si jo zastavili z ročnim postopkom. V takem primeru bo naš sklep, da H_1 sprejmemo.

Kadar bo pa obratno, ko bo računalniški program izračunal stopnjo značilnosti sig ≥ 0,05 bo za nas znak, da je ocena nekje v območju sprejema in bo to znak da Ho ne moremo zavrniti.

Razlikovanje med **ENOSTRANSKIMI** in **DVOSTRANSKIMI** preizkusi.

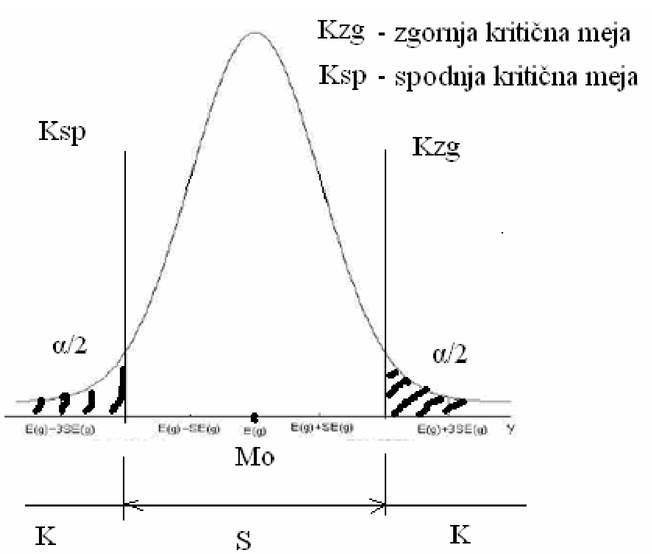
Enostranski ali dvostranski preskus

- Odvisno od vsebine
- Pri enakih pogojih je mogoče z enostranskim preskusom doseči značilno razliko pri nižji stopnji tveganja.

DVOSTRANSKI PREIZKUSI so preizkusi tipa ko v ničelno domnevo zapišemo, da je nek parameter (povprečje ali katerikoli drug parameter) enak neki konkretni vrednosti. V alternativno domnevo pa zapišemo da parameter ni enak tej vrednosti.

$$\text{Ho} \quad \overline{Y} = \text{Mo}$$

$$\text{H1} \quad \overline{Y} \neq \text{Mo}$$



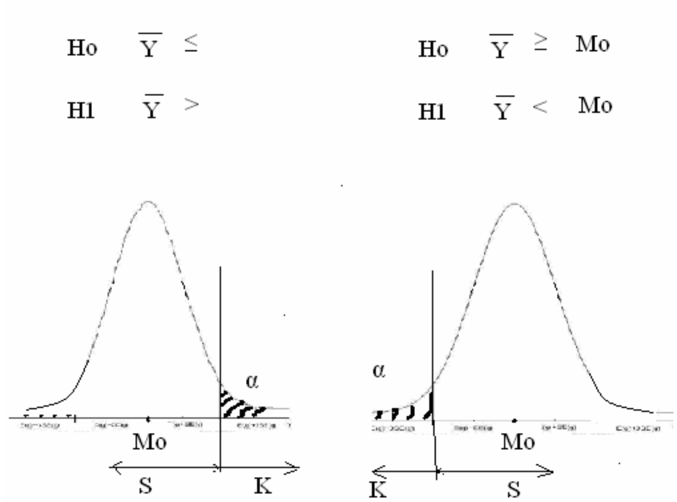
V H1 je zapisano, da se ocene razlikujejo. Ocene so lahko bodisi večje ali manjše. V takem primeru imamo dve kritični področji in s tem tudi dve kritični meji (zgornjo in spodnjo kritično mejo). Če bo ocena toliko večja, da bo padla preko kritične meje in bo to znak da je ocena večja. Oz. da bo ocena padla pod kritično mejo na spodnji strani bo znak, da je ocena znatno manjša. V enem ali drugem primeru bomo ničelno domnevo zavrnili in sprejeli alternativno. V tem primeru ni pomembno ali je vrednost večja ali manjša. Pomembno je samo da se razlikuje. To potegne za sabo, da imamo na vsaki strani zgolj polovično stopnjo tveganja (imamo dejansko dve kritični področji).

Če ocena pade nekam vmes med obe kritični področji bi rekli da je ocena v območju sprejema, ničelne domneve ne moremo zavrniti.

Če pa pade ocena preko kritične meje na zgornjo stran, oz. preko kritične meje na spodnjo stran, ničelno domnevo zavrnemo in sprejmemo alternativno, ki pravi, da se parameter razlikuje od vrednosti, ki je bila zapisana v ničelni domnevi.

ENOSTRANSKI PREIZKUS

Lahko se formulira kot povprečje je manjše ali enako in H1, povprečje je večje.



Oba pristopa sta praktično enaka. Gre samo za zrcalno sliko.

Desni del slike: Ničelna domneva velja bodisi kadar je situacija taka ali katerakoli druga večja vrednost. Če pa s preizkusom ugotovimo, da ocena pade na levo stran kritične meje (kritično področje na levi desno od nje je področje sprejema) je na levi celotna stopnja tveganja α. V tem primeru bi ničelno zavrnili in sprejeli alternativno.

Levi del slike: Situacija je zrcalna. Kritično področje je na zgornji desni strani. Tu je tudi stopnja tveganja α. Kritično področje je desno od kritične meje, levo je področje sprejema.

Katerakoli izmed treh različic se v praksi preizkuša z dvema različnima preizkusoma.

Prvi je t.i. **z-preizkus**, ki **temelji na predpostavki normalne porazdelitve**. Pomeni da lahko z- preizkus uporabljamo samo takrat:

1. **KADAR IMAMO VELIKE VZORCE (NAD 30)**. Takrat je vzorčna porazdelitev normalna in iz tega je izpeljana standardizirana spremenljivka z . V praksi z preizkusa ne moremo uporabiti.

Preskušanje domneve o aritmetični sredini - 1

- Predpostavimo: $y = N(M, \sigma_y)$
- Vrednost za σ_y je znana;
- Za preskus ničelne domneve $H_0 : M = M_0$ je ustrezen z-preizkus:

$$z = \frac{\bar{y} - M_0}{SE(\bar{y})} = \frac{\bar{y} - M_0}{\sigma_y}\sqrt{n}$$

standardni odklon, ki naj bi bil poznan za populacijo

- H_0 zavrnemo, če je s preizkusom izračunana vrednost z večja ali vsaj enaka od teoretične vrednosti za standardizirano normalno porazdelitev pri $\alpha \leq 0,05$

V praksi se zato uporablja univerzalno t-preizkus.

Preskušanje domneve o aritmetični sredini - 2

- Predpostavimo: $y = N(M, \sigma_y)$
- Vrednost za σ_y ni znana;
- Za preskus ničelne domneve $H_0 : M = M_0$ je ustrezen t-preizkus:

$$t = \frac{\bar{y} - M_0}{se(\bar{y})} = \frac{\bar{y} - M_0}{s_y}\sqrt{n}$$

standardni odklon

- H_0 zavrnemo, če je s preizkusom izračunana vrednost t večja ali vsaj enaka od teoretične vrednosti za t- porazdelitev pri stopinjah prostosti m=n-1 in $\alpha \leq 0,05$.

Ta je **UPORABEN NE GLEDE NA TO ALI IMAMO MAJHEN ALI VELIK VZOREC**. Namesto sigme nastopa mali s ipsilon. Za razliko od sigme, kar je vrednost standardnega odklona za populacijo, je mali s ipsilon ocena standardnega odklona na podlagi vzorca. Ta je pa znan. Ko imamo vzorec lahko standardni odklon na podlagi vzorca ocenimo.

P R I M E R

Preskušanje domneve o aritmetični sredini – primer

- Povprečna velikost gospodinjstev se vztrajno manjša. Za neko občino je bilo na podlagi preteklih raziskav ugotovljeno, da so imela gospodinjstva leta 1993 v povprečju 3,274 članov. Ker so želeli ugotoviti ali se trend zmanjševanja velikost gospodinjstev nadaljuje so izbrali vzorec 41 gospodinjstev (*Podatki so v priloženi datoteki*). Kaj so ugotovili?

Zamislimo si neko raziskavo na področju problematike demografije. Raziskovalci proučujejo kako se število rojstev oz. nataliteta manjša. Kako se to kaže v velikosti gospodinjstev.

Leta 1993 so prvič izvedli raziskavo. V njej so ugotovili, da so gospodinjstva štela 3,274 člana. Čez nekaj let so to raziskavo ponovili. Skušali so ugotoviti ali se število povprečno število članov na gospodinjstva še naprej manjša ali se je morda trend padanja ustalil ali se je morda obrnil navzgor. Na podlagi te ponovljene raziskave so dobili naslednje rezultate:

Leta 1993 so ugotovili, da je bila **povprečna velikost gospodinjstva** : **3,274** (na voljo so imeli podatke za celotno populacijo), zato so tudi natančno ugotovili kakšna bi bila velikost gospodinjstev.

Leta 2003 so raziskavo ponovili. Opazovali so samo **vzorec n = 41** in za teh 41 gospodinjstev **so ocenili** da so ta gospodinjstva štela v povprečju **2,88** člana (populacije v tem primeru ne poznamo).

Naša naloga je dati odgovor na vprašanje ali se trend zmanjševanja števila članov gospodinjstva nadaljuje. **Če želimo to odgovoriti bi bilo logično, da si v alternativno domnevo postavimo, da je povprečje manjše od 3,274**.

Leta 1993 : $\overline{Y} = 3,274$ podatki so bili dobljeni na celotni populaciji

Leta 2003 : $\overline{y} = 2,88$ ocena na vzorcu $n = 41$

1. korak

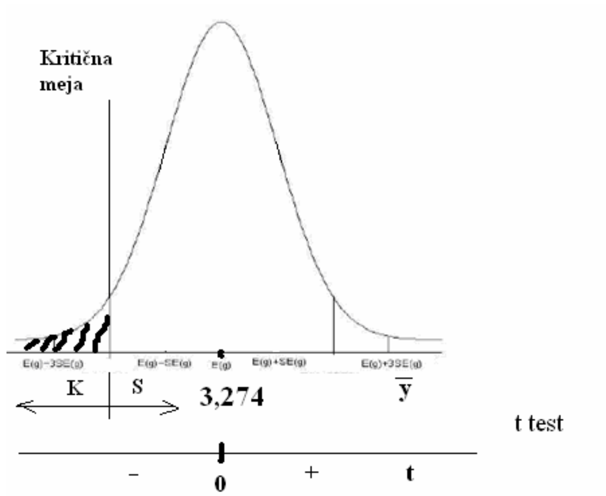
Ho $\overline{Y} \geq 3,274$ **ničelna domneva**

H1 $\overline{Y} < 3,274$ **alternativna domneva**

Če bo naša raziskava dokazala, da je povprečje manjše kot je bilo leta 1993 je to dokaz, da se velikost gospodinjstev še naprej manjša. Če tega ne bomo uspeli dokazati, se je trend upadanja ustalil.

2. **korak** - skiciramo

Priporočljivo je, da si zadevo skiciramo (ne glede ali se rešuje ročno ali z računalnikom).



Napravimo si skico ničelne porazdelitve. Porazdelitev okrog vrednosti, ki ste jo postavili v ničelno domnevo. **V ničelno domnevo smo postavili 3,274**. To naj bi bila prava vrednost. In da si na to skico takoj označimo, kje naj bi ležalo kritično področje.

Če bomo ničelno domnevo zavrnili takrat kadar bo vrednost manjša (pomeni da je kritično področje na levi strani). Če se bo **ocena 2,88** pokazala, da leži na levi od kritične meje potem ničelno domnevo v tem področju zavrnemo. Če se bo pokazalo, da ta ocena 2,88 pade že preko meje kritičnega področja, rečemo da ničelno domnevo zavrnemo, sprejmemo alternativno kar je znak da se velikost gospodinjstev še naprej manjša.

3. korak – T-test

Narediti T test pomeni dve stvari:

- **pomeni izračunati t** tako da vzamemo povprečno oceno na vzorcu 41 gospodinjstev s katero razpolagamo (2,88) minus Mo (vrednost parametra v ničelni domnevi = 3,274), deljeno z oceno standardnega odklona (1,382) krat koren velikosti vzorca (n=41).

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Clani	41	2,88	1,382	,216

povprečna ocena na vzorcu 41 gospodinjstev (izrač. s programom SPSS)

vrednost parametra v ničelni domeni

$$t = \frac{\overline{y} - Mo}{S\,y}\sqrt{n} = \frac{2{,}88 - 3{,}274}{1{,}382}\sqrt{41} = -1{,}835$$

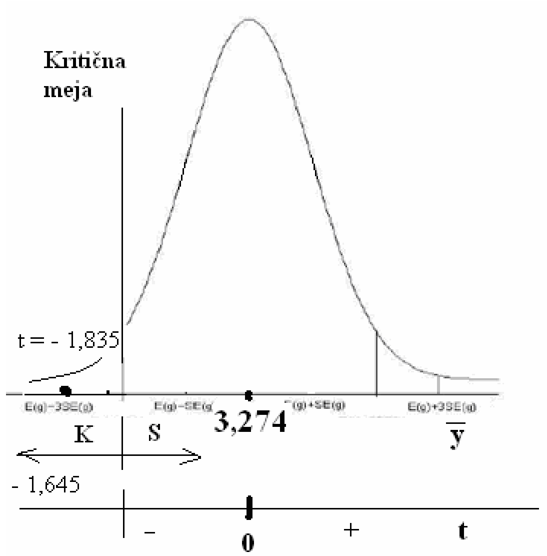
ocena standardnega odklona (standardne deviacije)

Največkrat bodo ti t-ji že dani s pomočjo računalniškega izpisa.

- pomeni **ugotoviti kakšen je t kritični (tc)**

  Tega ne moremo ugotoviti drugače kot da uporabimo tablice. Kot smo pri normalni porazdelitvi, za tisti z dobili vrednost 1,96 iz nekih tablic za normalno porazdelitev, smo tc-ju priredili $\alpha = 0{,}05$ pri **stopinjah prostosti** $m = n - 1$ (41-1=40). Logika vsega je, da **izračunano vrednost primerjamo s kritično**. Takrat dobimo odgovor kje vrednost leži ali na levi ali desni strani kritične meje.

$tc_{\alpha = 0{,}05 \;\; m = n - 1} = -1{,}645$ (odčitamo iz tablic)



Skala na kateri prikazujemo porazdelitev potem, spodaj rišemo vzporedno še drugo skalo na kateri nanašamo vrednosti t statistike.

Tam kjer je vrednost v ničelni domnevi (3,274) ima t statistika vrednost 0. Če bi dobili enako oceno kot je zapisana vrednost bi ta vrednost bila enaka 0. Na levo stran so vrednosti negativne na desno pa pozitivne.

Zapisali smo domnevi, označili smo kje leži naše kritično področje, ugotovili smo kaj pomeni narediti t test (pomeni na eni strani izračunati testno statistiko, na eni strani odkriti kritično vrednost).

Izračun t-ja vzamemo iz Priloge 1

One-Sample Test

	Test Value = 3.274					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Clani	-1,835	40	,074	-,396	-,83	,04

Č. KRITIČNE VREDNOSTI t-PORAZDELITVE pri stopnji tveganja α in stopinjah prostosti m

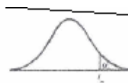
$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	m
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4

Gre za 5 različnih tablic (stolpcev).

Ne glede na to ali delamo enostranski ali dvostranski preizkus so te tablice formulirane, da je α (5 različnih, v vsakem stolpcu ena). Vse se nanašajo vedno samo na α na eni strani. Kar pomeni, če tako kot v našem primeru delamo enostranski preizkus in če je ta preizkus pri stopnji tveganja 0,05, je to za nas ustrezen stolpec, v katerem je t statistika pri stopnji tveganja 0,05.

Če bi izvajali dvostranski preizkus, kjer imamo na vsaki strani α/2, bi bilo potrebno gledati vrednosti v 0,025. To je glede stolpca.

Glede vrstice v skrajnem desnem stolpcu, kjer imamo m (stopinje prostosti). Poiščemo našo ustrezno vrednost. V našem primeru je ta m enak 41-1=40 gospodinjstev. Če se premaknemo čisto do konca navzdol vidimo, da so strukturirane vrednosti samo do 29, potem pa so od 30 naprej vse vrednosti enake in so enake vrednosti 1,645.



tc = 0,05
m = 40

$t_{.100}$	$\boxed{t_{.050}}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	m
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15
1.337	1.746	2.120	2.583	2.921	16
1.333	1.740	2.110	2.567	2.898	17
1.330	1.734	2.101	2.552	2.878	18
1.328	1.729	2.093	2.539	2.861	19
1.325	1.725	2.086	2.528	2.845	20
1.323	1.721	2.080	2.518	2.831	21
1.321	1.717	2.074	2.508	2.819	22
1.319	1.714	2.069	2.500	2.807	23
1.318	1.711	2.064	2.492	2.797	24
1.316	1.708	2.060	2.485	2.787	25
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	$\boxed{1.645}$	1.960	2.326	2.576	∞

Pozor! Ker je ta t-porazdelitev simetrična, so v teh tablicah podane samo absolutne vrednosti. Če smo si prej na sliki označili, da je kritično področje na levi strani in če se zavedamo da so t-ji negativni, pomeni da je tej vrednosti, ki smo jo prepisali iz tablic, potrebno spredaj dodati predznak minus. Naša dejanska kritična vrednost t-ja je -1,645. Če bomo dobili našo izračunano vrednost t statistike manjšo od 1,645 bo to znak, da leži v kritičnem področju, ničelno domnevo bomo zavrnili in sprejeli alternativno.

Če bomo dobili oceno t statistike, ki bi bila večja od kritične vrednosti, ki leži v območju sprejema bomo rekli. Ničelne **domneve ni mogoče zavrniti in se tu ustavimo**.

Z izračunom smo ugotovili da je t = - 1,835. To pomeni da leži v kritičnem področju. Ničelno domnevo lahko zavrnemo in pri stopnji tveganja α sprejmemo alternativno, ki pravi da je povprečno število članov gospodinjstva še manjše kot je bilo že ugotovljeno leta 1993, kar pomeni, da se trend zmanjševanja števila članov gospodinjstev nadaljuje.

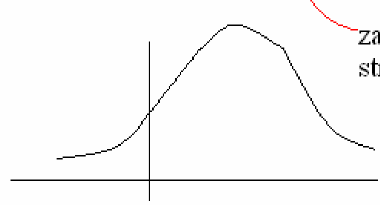
Sklep se glasi: *Na podlagi opazovanega vzorca gospodinjstev lahko pri stopnji tveganja α=0,05 zavrnemo ničelno domnevo in sprejmemo sklep, da je povprečno število članov gospodinjstva manjše od 3,274 oz. da se trend zmanjševanja velikosti gospodinjstev še nadaljuje.*

Kako bi postopali, če bi razpolagali z računalniškim programom. V tem primeru nam en korak odpade. Ravno tako je potrebno znati formulirati ničelno in alternativno domnevo. 2. korak, prikazati to s skico je še vedno priporočljivo. Tretji korak izvedbe t preizkusa nam pa računalnik olajša v toliko, da nam ni treba t-ja računati in da nam ni treba niti ugotoviti kakšen je kritični t, ker nam že vse skupaj pove kar **stopnja značilnosti**. Potrebno je pa paziti stopnjo značilnosti za dvostranski preizkus. (Stopnjo tveganja α, s katero v postopku preskušanja domnev zavrnemo ničelno domnevo, imenujemo **stopnja značilnosti**).

One-Sample Test

$\alpha/2 + \alpha/2$

	Test Value = 3.274					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Clani	-1,835	40	,074	-,396	-,83	,04



zanima nas samo polovica na spodnji strani

stopnja značilnosti enostranskega preizkusa

$$\frac{\text{sig (2 - tailed)}}{2} = \frac{0{,}074}{2} = 0{,}037$$

Če je stopnja značilnosti manjša ali enaka 0,05 pomeni, da Ho zavrnemo in sprejmemo H1. Če pa je stopnja značilnosti, ki jo ugotovimo, večja od 0,05 je ugotovitev, da Ho ne moremo zavrniti. Ne glede ali rešujemo to preko računalniškega izpisa ali ročno, moramo priti do enake ugotovitve.

sig $\leq$ 0,05 $\rightarrow$ Ho zavrnemo in sprejmemo H1

sig $\geq$ 0,05 $\rightarrow$ Ho ne moremo zavrniti

mi dobimo vrednost 0,037

VPRAŠANJI:

Kako vemo ali je enostranska ali dvostranska?

Če bi bilo vprašanje postavljeno samo ali je število članov še vedno 3,274 ali se to število razlikuje pa ne bi bilo pomembno ali se razlikuje navzdol ali navzgor, potem bi iz tega izhajalo, da nas zanima samo to ali je povprečje enako 3,274 ali ni. To bi bil dvostranski preizkus. Ker pa v tem primeru piše, da so želeli ugotoviti ali se trend zmanjševanja števila nadaljuje, pomeni da nas zanima samo, če je morda povprečje manjše kot je že bilo. In čim zapišemo neenačaje je to takoj znak za enostranski preizkus.

Kako pa vemo glede na zapis kje leži kritično področje?

Če bomo ničelno domnevo zavrnili takrat kadar bo vrednost manjša, pomeni da bomo zavrnili ko bo v kritičnem področju, pomeni kritično področje bo na levi strani. Če bi bila formulacija v H_1, da je povprečje večje od 3,274, (če bi bilo vprašanje ali se trend povečuje). Iz neenačaja oz. enačaja veste kje leži kritično področje.

2. del

~~~~~~

Ponovitev:

V prejšnjem primeru smo imeli preizkus domneve v eni sami aritmetični sredini, kjer sta teoretično dva možna preizkusa (z in t). ***V praksi se bolj ali manj uporablja t preizkus****. Je univerzalen ko imamo majhen ali velik vzorec. Drugi razlog je standardni odklon, ki ga za populacijo, kot pri z-preizkusu, običajno nimamo poznanega. Zato pa ga vedno ocenimo na podlagi vzorca.*

~~~~~~

Nadaljujemo z novim preizkusom, ki se nanaša na **primerjavo med dvema aritmetičnima sredinama**. Pri gospodinjstvih je šlo samo zato, da imamo en sam vzorec in na podlagi tega vzorca ugotavljamo kakšna je dejanska aritmetična sredina. V tem primeru pa operiramo z dvema vzorcema in na podlagi enega ocenimo aritmetično sredino na podlagi drugega ocenimo aritmetično sredino in nato s pomočjo nekega preizkusa ugotavljamo kakšna je dejanska vrednost aritmetične sredine v eni populaciji in v drugi.

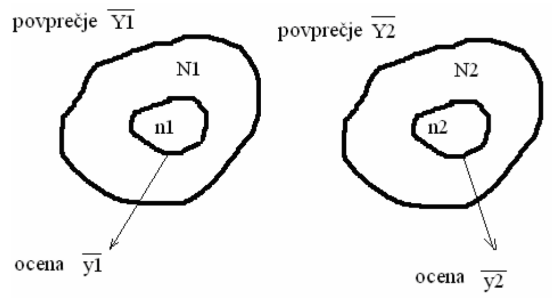
Preskušanje domneve o razliki med aritmetičnima sredinama za neodvisna vzorca

- Predpostavimo: $y_1 = N(M_1, \sigma_1)$ $\quad y_2 = N(M_2, \sigma_2)$

  $\sigma_1^2 = \sigma_2^2$
- Za preskus ničelne domneve $H_0 : M_1 = M_2$ je ustrezen t-preizkus:

  $$t = \frac{(\bar{y}_1 - \bar{y}_2) - (M_1 - M_2)}{se(\bar{y}_1 - \bar{y}_2)}$$
- H_0 zavrnemo, če je s preizkusom izračunana vrednost t večja ali vsaj enaka od teoretične vrednosti za t- porazdelitev pri stopinjah prostosti m=n_1+n_2-2 in $\alpha \leq 0,05$

Imamo populacijo z N1 enotami. Nato populacijo z N2 enot. Za prvo populacijo je povprečje Y_1 črtica, za drugo pa povprečje Y_2 črtica, ki ju ne poznamo. Kar pa poznamo je pri drugi, da izberemo vzorec mali n2 pa dobimo oceno mali y2, pri prvi pa izberemo vzorec mali n1 in na podlagi tega pridemo do ocene mali y1 črtica. Potem z našim preizkusom skušamo na podlagi primerjave teh dveh ocen ugotoviti, kakšna je razlika med dvema dejanskima aritmetičnima sredinama, ki ju sicer ne poznamo.



Tudi v tem primeru uporabimo t-preizkus, ki bo nekoliko drugače izračunan. Tudi tu enako konfiguriramo ali je preizkus enostranski ali dvostranski.

Dvostranski

$$Ho: \overline{Y1} = \overline{Y2} \quad \text{oziroma} \quad Ho: \overline{Y1} - \overline{Y2} = 0$$
$$H1: \overline{Y1} \neq \overline{Y2} \qquad\qquad\quad H1: \overline{Y1} - \overline{Y2} \neq 0$$

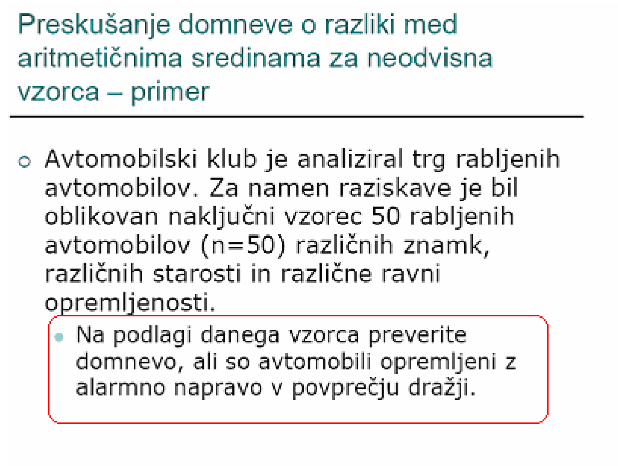
Enostranski

$$Ho: \overline{Y1} \leq \overline{Y2} \quad \text{oziroma} \quad Ho: \overline{Y1} \geq \overline{Y2}$$
$$H1: \overline{Y1} > \overline{Y2} \qquad\qquad\quad H1: \overline{Y1} < \overline{Y2}$$

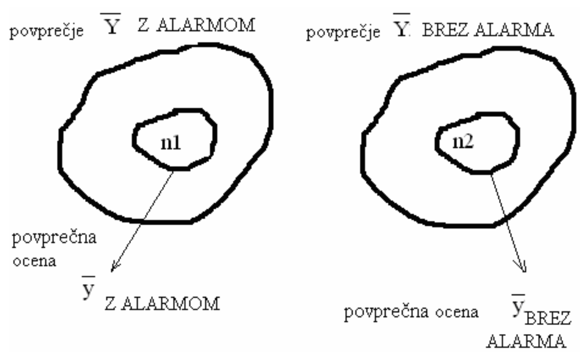
$$Ho: \overline{Y1} \geq \overline{Y2} \quad \text{oziroma} \quad Ho: \overline{Y1} - \overline{Y2} \geq 0$$
$$H1: \overline{Y1} < \overline{Y2} \qquad\qquad\quad H1: \overline{Y1} - \overline{Y2} < 0$$

Katerikoli izmed teh treh testov lahko pride v tem primeru poštev. Katerega bomo izbrali je odvisno od vsebine. Na podlagi besedila, na podlagi vprašanja, na podlagi vsebine, je potrebno razbrati ali gre za primerjavo tipa dvostranski preizkus (zanima nas samo ali sta aritmetični sredini enaki ali ne) ali naj bi bila prva večja ali naj bi bila druga večja.

P R I M E R



Imamo populacijo v kateri so **avtomobili z alarmom** in drugo populacijo v kateri so **avtomobili brez alarma**.



Kako formulirati ničelno in alternativno domnevo? Vprašanje je ali so avtomobili opremljeni z alarmno napravo v povprečju dražji? Po logiki je alarmna naprava dodatni element opreme, zato bi pričakovali da je temu tako.

1. primer

$$Ho : \overline{Y}_{ALARM} \leq \overline{Y}_{BREZ\ ALARMA}$$

$$H1 : \overline{Y}_{ALARM} > \overline{Y}_{BREZ\ ALARMA}$$

2. primer

$$Ho : \overline{Y}_{ALARM} - \overline{Y}_{BREZ\ ALARMA} \leq 0$$

$$H1 : \overline{Y}_{ALARM} - \overline{Y}_{BREZ\ ALARMA} > 0$$

Po konvenciji je vedno tako, da:

- **ničelne domneve** predstavljajo znak enačaj (enostranski ali dvostranski);
- **alternativne domneve** imajo znak različno, večje ali manjše.

1. primer zgoraj

ALTERNATIVNA DOMNEVA

Tisto kar skušamo dokazati, kar je vprašanje, damo v alternativno domnevo. Povprečna cena avtomobila z alarmno napravo je večja od povprečne cene avtomobila brez alarmne naprave.

NIČELNA DOMNEVA

Povprečna cena z alarmno napravo je manjša ali enaka povprečni ceni avtomobila brez alarma.

2. primer zgoraj

NIČELNA DOMNEVA

Povprečna cena z alarmom minus povprečna cena brez alarma je manjša ali enaka 0.

ALTERNATIVNA DOMNEVA

Povprečna cena z alarmom minus povprečna cena brez alarma je večje od 0.

Oba zapisa sta povsem enakovredna.

Spet uporabimo t-preizkus. Imamo izračunani t in kritični tc. Če bo naš izračunani t padel preko kritične meje v kritično področje, ničelno domnevo zavrnemo in sprejmemo alternativno in bomo potrdili našo domnevo da so avtomobili z alarmom v resnici dražji . Če bo pa naša ocena padla v območju sprejema bomo rekli, da na podlagi podatkov žal tega ne moremo potrditi. Ni mogoče trditi da so avtomobili z alarmom dražji kot avtomobili brez.

t-statistika se računa na dva različna načina:

- če sta varianci enaki;
- če varianci nista enaki.

Če sta varianci enaki ali ne to nima nikakršne vsebinske zveze s tem ali je povprečna cena 1 ali 2, enaka, večja ali manjša. Ta enakost varianc je samo tehnični vmesni pogoj, da lahko to nalogo sploh izpeljemo.

Group Statistics

	Alarmna naprava	N	Mean	Std. Deviation	Std. Error Mean
Cena vozila v DEM	0	47	16790,00	7808,391	1138,971
	1	2	21200,00	4666,905	3300,000

povprečna cena z alarmno napravo

povprečna cena brez alarmne naprave

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Cena vozila v DEM	Equal variances assumed	,681	,414	-,788	47	,435	-4410,000	5598,935	-15673,605	6853,605
	Equal variances not assumed			-1,263	1,252	,393	-4410,000	3491,025	-32354,325	23534,325

ta test se nanaša na naslednji preizkus:

stopnja značilnosti

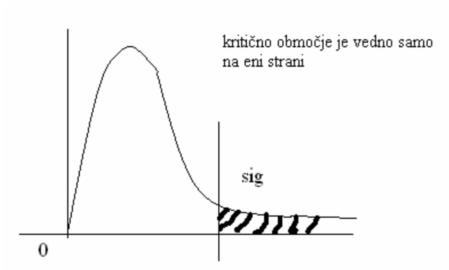
$$Ho : \sigma^2 ALARM = \sigma^2 BREZ\ ALARM$$

$$H1 : \sigma^2 ALARM \neq \sigma^2 BREZ\ ALARM$$

Nekdo, ki bi pogledal ceni bi rekel da je jasno, da so avtomobili z vgrajeno alarmno napravo dražji kot avtomobili brez. Ni pa to nujno. Na podlagi primerjave ocen nikoli ne moremo ugotoviti že v naprej znanega sklepa. Sam test je prikazan v spodnjem delu tabele.

Ta preizkus nima neposredne povezave z vsebinskim vprašanjem. Vse kar je potrebno razumeti je **stopnja značilnosti**. (**STOPNJA ZNAČILNOSTI** - je stopnja tveganja α, s katero, v postopku preskušanja domnev, zavrnemo ničelno domnevo)

t-preizkus se lahko izvede kot enostranski ali dvostranski (zato tudi na SPSS-u na nekaterih mestih piše Sig, (1-tailed), Sig. (2-tailed). V ozadju Levenovega preizkusa je F statistika oz. **F porazdelitev**, ki izgleda asimetrično. **Kritično območje je vedno samo na eni strani**, ker je sama porazdelitev asimetrična, zato lahko zavzame samo pozitivne vrednosti. Pri t-ju ni tako. t lahko zavzame tudi negativne vrednosti.



Če je **sig ≤ 0,05** → Ho zavrnemo

> 0,05 → Ho ne moremo zavrniti

0,414 → predpostavimo da sta varianci enaki

Ali varianci sta enaki ali nista?

Vrednost je **večje od 0**, pomeni da (**Ho ne moremo zavrniti**). Pravimo da sta enaki in ugotavljamo da tega ne moremo negirati. Če tega ne moremo zanikati, kar pomeni da je zelo velika verjetnost da ničelna domneva drži. Predpostavili bomo**, da varianci sta enaki**.

Imamo dve vrstici. Po eni se je izračunal t = - 0,788, po drugi različici bi se pa izračunal

t = - 1,263. Ker smo ugotovili da varianci sta enaki, **nadaljujemo s prvo vrstico v tabeli.**

S tem nismo še vedno nič ugotovili ali so avtomobili z alarmi dražji ali cenejši. Ugotovili smo samo s pomočjo katerega orodja bomo to ugotovili ali so avtomobili dražji ali cenejši. Kaj moramo uporabiti?

~~~~~~

VMESNO VPRAŠANJE PREDAVATELJA

Ali razumemo vsi kaj je povprečje in kaj je varianca?

Lahko rečemo, povprečna plača v Sloveniji je 800 EUR. Vsi vemo, da nimajo vsi 800 EUR. Nekateri imajo 8000 EUR ali še več, nekateri pa morda 200 EUR. ***Koliko so razlike med posamezniki toliko nam kaže standardni odklon ali varianca*** *kot neka mera.* **POVPREČJE NAM KAŽE NEKO SREDNJO VREDNOST ZA POPULACIJO, VARIANCA OZ. STANDARDNI ODKLON NAM PA KAŽE KAKŠNE SO RAZLIKE MED ENOTAMI**. *Lahko vzamemo dve regiji, Osrednjo slovensko in primorsko regijo. Tako za eno kot za drugo regijo ugotovimo recimo, da imajo v povprečju plačo 800 EUR V Osrednji slovenski regiji ugotovimo, da so vse plače nekje na intervalu med 600 in 1500, na primorskem pa ugotovimo, da imamo ljudi z 200 EUR plače pa tudi take ki imajo 5000 EUR plače. Vidimo da so razlike na primorskem bistveno*
~~~~~~

večje kot so v Osrednji slovenski regiji. Eno je kazalnik srednje vrednosti (povprečje), ravni, s čimer bi se primerjali katera regija ima višjo povprečno plačo. Na podlagi tega bi rekli, da je ta regija bolj razvita. Drugi kazalnik, ki bi nam pokazal razlike med nami posamezniki, bi bil pa standardni odklon oz. varianca. Večja kot je varianca, večje so razlike med enotami, manjša kot je varianca, manjše so razlike (Osnove Statistika I).

Pridemo na Statistiko II. Zakaj sploh te variance primerjamo je ena izmed predpostavk.

Statistika I pravi : Normalna porazdelitev je opredeljena s povprečjem (minus ena, minus dva, minus tri standardni odkloni, plus ena, plus dva, plus tri standardni odkloni). Pomeni da nam ***povprečje kaže lego te krivulje, bolj levo ali desno, standardni odklon pa kaže ali je ta krivulja bolj razpotegnjena ali bolj sploščena****. Večji kot je standardni odklon, bolj je sploščena ta krivulja, kar kaže na večje razlike. Bolj kot je ta krivulja koničasta pomeni da so med nami manjše razlike.*

Med seboj primerjamo ti dve porazdelitvi, t.j. v populaciji, kjer so avtomobili z alarmom in avtomobili brez alarma. Naša predpostavka pravi ali sta ti dve krivulji enakih oblik. Če nista enakih oblik, da je ena precej bolj razpotegnjena, je potem prekrivanje bistveno drugačno, kot pa je, če bi krivulja bila bolj koničasta. Če se v obeh populacijah meri "z enakim metrom" je to ena primerjava, če se pa v eni populaciji meri z natančnejšim merilom (kljunastim merilom) v drugi pa z (zidarskim metrom), je to povsem druga primerjava.

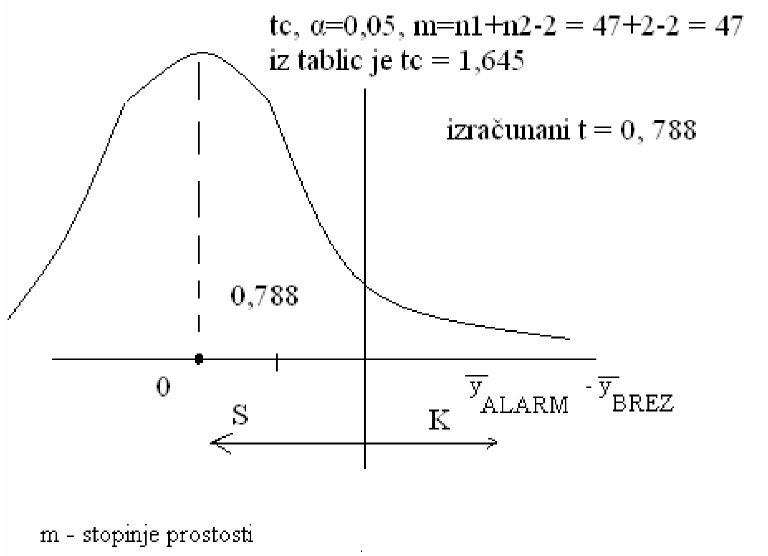
Naš preizkus je temeljil na primerjavi ali je varianca 1 enaka varianci 2 v smislu da se prepričamo ali sta ti dve obliki enaki oz. ali imamo v obeh primerih v rokah enak meter. Ko ugotovimo ali sta metra enaka ali ne, gremo dejansko meriti. Z enim metrom izmerimo ceno avtomobilov brez alarma, z drugim metrom ceno avtomobilov z alarmom. Ti dve vrednosti bi primerjali in potem nadaljevali.

~~~~~~

Gremo na to kaj je ta preizkus pokazal.

Primerjavo imamo z F testom. Kriterij kako računalniške izpise uporabljamo je, da kadar je stopnja značilnosti manjša ali enaka 0,05 bi to pomenilo. Če vzamemo kritično mejo pri α =0,05. To je naša zgornja sprejemljiva stopnja tveganja.

S preizkusom primerjamo eno in drugo varianco. Ta nam pokaže stopnjo značilnosti, ki je znatno večja od sprejemljive. Če bi želeli zavrniti ničelno domnevo in sprejeli alternativno bi to lahko naredili samo ob zelo visokem tveganju. To je druga interpretacija stopnje značilnosti. Ker bi morali tvegati več kot 40%. Ker toliko nismo pripravljeni tvegati, ne moremo zavrniti ničelne domneve.
~~~~~~



Ocenjujemo razliko med povprečno ceno avtomobila z alarmom in med povprečno ceno avtomobila brez alarma. Vrednost v ničelni domnevi je 0, v alternativni domnevi pravimo da naj bi ta razlika bila pozitivna. Kritično področje je na desni strani.

t kritični pri α=0,05, stopinje prostosti m=n1+n2-2=47+2-2=47 (47 avtomobilov iz prve skupine plus 2 avtomobila iz druge skupine minus 2 je 47). Če pogledamo v tablice je tudi tokrat v resnici neskončno in je vrednost enaka (tc=1,645). Izračunani t je bil pa 0,788. Ocena leži v območju sprejema, kar pomeni da ničelne domneve ne moremo zavrniti. Ne moremo potrditi teze, da so avtomobili z alarmom dražji v povprečju z avtomobili brez alarma. To pokaže naš izračun.

Čeprav smo videli, da je bila razlika med obema cenama kar precejšnja (n2 je sicer nekoliko problem. Če bi bil večji bi se potem ta ocena izkazala kot značilna in bi se kot značilna pokazala tudi razlika). Izračunana t-statistika znaša 0,788, kritični t odčitan iz tablic 1,645. Ker je izračunani t manjši od kritičnega pade v območje sprejema, ničelne domneve ni mogoče zavrniti.

Sklep:

Na podlagi primerjave dveh vzorcev avtomobilov ni mogoče zavrniti ničelne domneve pri dovolj nizki stopnji tveganja, zato ne moremo trditi, da so avtomobili z vgrajeno alarmno napravo v povprečju dražji kot avtomobili brez nje.

Kako bi ravnali, če bi upoštevali računalniški izpis

Kadar imamo računalniški izpis, načeloma zadostuje da pogledamo stopnjo značilnosti preizkusa. Mi izvajamo enostranski preizkus, zato je ustrezna stopnja značilnosti za nas:

$$sig_{\text{enostranski}} = \frac{\text{sig (2-tailed)}}{2} = \frac{0{,}435}{2} = 0{,}2175$$

> 0,05 kar sledi da Ho ne moremo zavrniti

Dobimo enak rezultat. Tudi tu bi bil lahko isti sklep: Na podlagi primerjave dveh vzorcev avtomobilov je dana stopnja tveganja za zavrnitev ničelne domneve previsoka.

preizkus za eno aritmetično sredino

Imamo preizkus tipa povprečje je, so stopinje prostost enake **m = n - 1**.

$$\mathbf{Ho} : \overline{Y} =$$

pri primerjavi dveh aritmetičnih sredin

Potem imamo preizkus tipa povprečje ena, povprečje dva. **Stopinje prostosti** so enake

m = n1 + n2 – 2.

$$\mathbf{Ho} : \overline{Y}_1 = \overline{Y}_2$$

m lahko računamo na različne načine pri različnih preizkusih.

Rabimo **α** in rabimo **m**. V tem danem primeru smo rekli da je α=0,05 in vedno začnemo s to zgornjo mejo, ker je ta še najbolj sprejemljiva.

Vsak stolpec v tablicah je podan pri drugačni stopnji tveganja. Skrajno levi stolpec predstavlja α=0,01, sledi α=0,05, α=0,25, α=0,01 in α=0,005. Najprej izberemo ustrezni stolpec 0,05 potem je v zadnjem stolpcu m (1,2,3 vse do 29, potem pa od 30 naprej neskončno ∞).

Za vse vrednosti, ki so nad 30 se vrednost t-ja približa normalni porazdelitvi. V praksi so vrednosti velikosti vzorca od 30 naprej enake kot pri normalni porazdelitvi. Iz stolpca α=0,05 in vrstice m = ∞ (vzorci nad 30), dobimo vrednost 1,645. (Lahko je plus ali minus). V tablicah je zapisano samo absolutno. Če je kritično področje na desni je plus, če je kritično področje na levi, vzamemo isto vrednost in ji damo predznak minus.

PREIZKUŠANJE DOMNEV O RAZLIKI ZA POLJUBNO ŠTEVILO ARITMETIČNIH SREDIN

Doslej smo imeli preizkušanje domnev **ene aritmetične sredine**, primerjavo **dveh aritmetičnih sredin,** to pa še posplošimo na **poljubno število aritmetičnih sredin**. Tej metodi specifično pravimo, da **gre za ANALIZO VARIANCE**. (Pozor!) Kljub temu, da se metoda imenuje analiza variance, je to iz drugega razloga.

TO NI METODA ZA PREIZKUS DOMNEVE O VARIANCAH AMPAK JE METODA ZA PREIZKUS DOMNEVE O ARITMETIČNI SREDINI.

Alternativna domneva pravi: vsaj ena aritmetična sredina $\overline{Yi}$ (Yi črtica) se razlikuje.

V Ho zapišemo da so vse enake, matematično gledano pa lahko to zavrnemo čim ena odstopa (pomeni da enakost ne velja), ničelno zavrnemo, sprejmemo alternativno.

a) $\text{Ho} : \overline{Yit} = \overline{Yde} = \overline{Yfr} = \overline{Yjp} = \overline{Yost}$

H1 : vsaj ena $\overline{Yi}$ (aritmetična sredina) se razlikuje

uporabimo ali F ali Robust Welch statistiko

Če želimo ugotoviti ali so si avtomobili po opremljenosti enaki ali ne, moramo na tak način zapisati alternativno domnevo. Da bomo znali uporabiti ali F ali Robust Welch statistiko, je treba vmesno opraviti še drug preizkus, podobno kot prej.

Kadar bomo ugotovili da variance so enake, bomo (podobno kot pri prejšnjem primeru nalogo nadaljevali ali v prvi ali v drugi vrstici). Tudi kadar bomo ugotovili, da variance so enake bomo za preizkus uporabili F-test. Če bomo ugotovili da variance niso enake, bomo uporabili t.i. **Robust Welch test**.

Na podlagi vsebine formuliramo ničelno in alternativno domnevo, da bi jo znali preveriti moramo vedeti kaj je z variancami v ozadju.

Program se vpraša ali so metri v vseh populacijah enaki, če so enaki potem to domnevo preveri z F-testom. Če skala, v vseh populacijah ni enaka, dovolj da v eni izstopa, potem moramo uporabiti drug test, ki je na skalo neobčutljiv in računa po Robust Welch testu.

Analiza variance – primer

- Avtomobilski klub je analiziral trg rabljenih avtomobilov. Za namen raziskave je bil oblikovan naključni vzorec 50 rabljenih avtomobilov (n=50) različnih znamk, različnih starosti in različne ravni opremljenosti. Ali lahko na podlagi danega vzorca sklepamo, da se povprečna raven opremljenosti razlikuje med avtomobili različnega porekla?
  - Oblikujte ustrezno ničelno in alternativno domnevo;
  - Predlagajte ustrezni preskus;
  - Preverite ali podatki zadoščajo predpostavkam izbranega testa;
  - Komentirajte rezultat.

Besedilo naloge je spet osnovano na avtomobilih, različnih znamk, različne opremljenosti, itn. Ključno vprašanje je ali lahko na podlagi vzorca sklepamo, da se povprečna raven opremljenosti razlikuje med avtomobili različnega porekla. V ozadju imamo pet različnih porekel (italijanski, nemški, francoski, japonske in vse ostale). Glede opremljenosti, na podlagi podatkov o tem ali imajo airbag, servo volan, centralno zaklepanje, električni dvig stekel, itn. bi izračunali nek indeks na intervalu med 0 in 100 in kaže koliko odstotkov od celotnega nabora opreme ima posamezen avtomobil.

Naša naloga je ugotoviti ali so avtomobili, ki prihajajo iz italijanskih tovarn, enako opremljeni kot avtomobili, ki prihajajo iz nemških, …, ki prihajajo iz japonskih, itd.

Ali morda obstajajo razlike in so avtomobili iz ene države bolj opremljeni ali manj v odnosu do druge države?

Pri analizi variance je olajševalna okoliščina, da ni dileme ali je to enostranski ali dvostranski preizkus (v Ho se vedno zapiše, da so vsa povprečja enaka, da pa v H1 ena odstopa).

b) $Ho : \sigma^2_{it} = \sigma^2_{de} = \sigma^2_{fr} = \sigma^2_{jp} = \sigma^2_{ostalo}$

$H1$: vsaj ena varianca se razlikuje

Pod točko b zapišemo v ničelno domnevo (Ho) : povprečna opremljenost avtomobilov italije, je enaka povprečni opremljenosti avtomobilov iz nemčije, …., je enaka povprečni opremljenosti avtomobilov iz japonske, je enaka povprečni opremljenosti avtomobilov iz vseh ostalih držav.

Priloga 3: Analiza variance

Descriptives

Raven opreme v odstotkih

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Italija	4	20,3125	5,98392	2,99196	10,7907	29,8343	12,50	25,00
Nemčija	15	37,0833	23,91235	6,17414	23,8411	50,3256	6,25	93,75
Japonska	13	9,1346	9,75460	2,70544	3,2400	15,0293	,00	25,00
Francija	14	33,4821	21,17452	5,65913	21,2563	45,7079	,00	62,50
Ostalo	4	34,3750	38,69620	19,34810	-27,1993	95,9493	,00	87,50
Total	50	27,2500	23,13723	3,27210	20,6745	33,8255	,00	93,75

V tabeli so prikazana povprečja (Mean) opremljenosti vozil iz vseh petih držav. (Italijansko tržiče 20,3%, Nemčija 37%,…Ostalo 34%. To so zgolj ocene in na podlagi teh ocen ne moremo neposredno ugotoviti kakšno je stanje v teh populacijah v ozadju. Da bi to ugotovili rabimo ustrezen test.

Točka b – da bi ugotovili ali F ali Robust Welch, moramo najprej ugotoviti kaj je z variancami in spet imamo podobno kot prej Levenov test of Homogeneity Variances. Vse kar je pomembno je **stopnja značilnosti**.

Test of Homogeneity of Variances

Raven opreme v odstotkih

Levene Statistic	df1	df2	Sig.
3,031	4	45	,027

Ker je ta stopnja značilnosti sig = 0,027 in ker je to manjše od 0,05, lahko Ho zavrnemo. Sprejmemo sklep, da se vsaj ena razlikuje in če se razlikuje, variance niso enake.

Robust Tests of Equality of Means

Raven opreme v odstotkih

	Statistic[a]	df1	df2	Sig.
Welch	6,305	4	13,093	,005

a. Asymptotically F distributed.

Če variance niso enake je treba uporabiti Robust Welch test.

INTERPRETACIJA

Na podlagi primerjave opremljenosti avtomobilov različnega porekla lahko, pri stopnji tveganja α=0,005, zavrnemo ničelno domnevo in sprejmemo sklep, da med avtomobili z različnim poreklom obstajajo značilne razlike v njihovi opremljenosti.

Regresijska analiza – osnovana na analizi povezanosti med pojavi

Faktorska analiza – statistična analiza za analizo negativnih pojavov

Konec.

~~~~~~

STANDARDNI ODKLON

Stándardni odklòn (tudi **stándardna deviácija**) (**σ**, sigma) je statistični kazalec, največkrat uporabljen za merjenje statistične razpršenosti enot. Z njim je moč izmeriti, kako razpršene so vrednosti, vsebovane v populaciji. Standardni odklon je definiran kot kvadratni koren variance, s čimer je v vsakem primeru dosežena pozitivna vrednost kazalca.

Standardni odklon je lahko računan kot σ (sigma) in sicer kot odklon celotne populacije ali njene naključne spremenljivke, ali pa kot s in sicer kot odklon posameznega vzorca statistične populacije. Za ta različna odklona se formuli razlikujeta.

Merjenje standardnega odklona je v statistiko vpeljal angleški statistik Karl Pearson.

MATEMATIČNA DEFINICIJA

Standardni odklon vseh enot statistične populacije je definiran s formulo:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}}$$

kjer je x_i i-ta enota v statistični populaciji, $\overline{x}$ aritmetična sredine populacije, N pa število vseh enot.

Standardni odklon vzorca statistične populacije je definiran s formulo:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}}$$

Velik standardni odklon σ kaže na veliko razpršenost enot v populaciji, tj. enote so razporejene v velikem obsegu okoli aritmetične sredine. Majhen standardni odklon σ pa nasprotno predstavlja veliko koncentracijo statističnih enot okoli aritmetične sredine.
~~~~~~