KOMBINATORIKA:

n - univerzalna množica
k - število elementov v množici

Permutacije brez ponavljanja (urejamo **VSE** elemente množice v vrsto /n=k/, in **VSI** elementi so si med sabo različni)

$$p_n = n!$$

Permutacije s ponavljanjem (urejamo **VSE** elemente množice v vrsto, elementi se lahko ponavljajo; npr. a,a,b - vedno jih je manj kot brez ponavljanja)

$$p_{n(n_1,n_2,\ldots)} = \frac{n!}{n_1! * n_2! * \ldots}$$

Variacije brez ponavljanja (v vrsto množice (n) urejamo manj (k) elementov, kot je vseh, in vsak lahko nastopi samo enkrat; npr. imamo cifre 1, 2, 3, 4, ki jih razporejamo na dve mesti, npr. 21, 32, 41... Vrstni red **JE** pomemben, torej 21 ni enako 12)

$$V_n^k = \frac{n!}{(n-k)!}$$

Variacije s ponavljanjem (v vrsto množice (n) urejamo manj (k) elementov, kot je vseh, elementi se lahko ponavljajo; npr. imamo cifre 1, 2, 3, 4, ki jih razporejamo na dve mesti, npr. 21, 32, 41, pa tudi 11, 22, 33...)

$$V_n^k = n^k$$

Kombinacije brez ponavljanja (v vrsto množice (n) urejamo manj (k) elementov, kot je vseh, in vsak lahko nastopi samo enkrat; vrstni red **NI** pomemben, torej 21 je enako 12)

$$C_n^k = \frac{n!}{k! * (n-k)!} \quad \textbf{tudi} \quad C_n^k = \binom{n}{k}$$

Kombinacije s ponavljanjem (v vrsto množice (n) urejamo manj (k) elementov, kot je vseh, elementi se lahko tudi ponavljajo; vrstni red **NI** pomemben, torej 21 je enako 12)

$$C_{n+k-1}^k = \frac{(n+k-1)!}{k! * (n-1)!}$$

VERJETNOST VSOTE ZDRUŽLJIVIH DOGODKOV

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D) - P(A \cap B) - P(A \cap C) - P(A \cap D) - P(B \cap C) - P(B \cap D) - P($$

PRIMER: Izberemo število med 1 in 100. Kakšna je verjetnost, da je deljivo s 3 **<u>ali</u>** 5?

P(A) = 33/100 - 33 števil je deljivih s 3
P(B) = 20/100 - 20 števil je deljivih s 5
P(A∩B) = 6/100 - 6 števil je deljivih s 3*5 (15)

P(A∪B) = 33/100 + 20/100 - 6/100 = 47/100 = **0,47**

<u>Popoln sistem dogodkov je množica paroma nezdružljivih dogodkov, katerih vsota je gotov dogodek.</u>

POPOLNA VERJETNOST (poskusi, ki se zgodijo v dveh fazah)

$$P(A) = P(H_1) * P\left(\frac{A}{H_1}\right) + P(H_2) * P\left(\frac{A}{H_2}\right) + \ldots + P(H_n) * P\left(\frac{A}{H_n}\right)$$

PRIMER:
- tri posode s kroglami, v prvi 4 bele 1 rdeča, v drugi 3 bele 2 rdeči, v tretji 1 bela 3 rdeče
- mečemo kocko; 1-3 sežemo v prvo posodo, 4-5 v drugo, 6 v tretjo

<u>Kakšna je verjetnost, da izberemo belo kroglo?</u>

H_1= 3/6 = 1/2 $P(A/H_1)$=4/5
H_2= 2/6 = 1/3 $P(A/H_2)$=3/5
H_3= 1/6 $P(A/H_3)$=1/4

P(A)=1/2*4/5 + 1/3*3/5 + 1/6*1/4 = 0,6417 = **64,17%**

POGOJNA VERJETNOST (Bayesova formula)

$$P\left(\frac{H_k}{A}\right) = \frac{P(H_k) * P\left(\frac{A}{H_k}\right)}{P(H_1) * P\left(\frac{A}{H_1}\right) + P(H_2) * P\left(\frac{A}{H_2}\right) + \ldots + P(H_n) * P\left(\frac{A}{H_n}\right)}$$

PRIMER: če v zgornjem primeru izberemo belo kroglo, kakšna je verjetnost, da je iz tretje posode?

$P(H_k/A)=(1/6*1/4) / 0,6417 = 0,0649 =$ **6,49%**

BINOMSKA ALI BERNOULIJEVA FORMULA:

- zaporedje enakih, neodvisnih preizkusov
- verjetnost prejšnjega poizkusa ne vpliva na izid novega

n - število ponovitev
p - verjetnost poizkusa
k - število ugodnih izidov

$$P(n,p,k)=\binom{n}{k}*p^k*(1-p)^{n-k}$$

PRIMER:

Kolikšna je verjetnost, da v treh metih nepoštenega kovanca, pri katerem je verjetnost padca grba 0,75, vržemo grb natančno dvakrat?

$$P(3,0.75,2)=\binom{3}{2}*0,75^2*(1-0,75)^{3-2}=\frac{3!}{2!*(3-2)!}*0,75^2*0,25^1=3*0,75^2*0,25=0,4219=42,19$$

HIPERGEOMETRIJSKA PORAZDELITEV:

n - število poskusov
k - število ugodnih izidov posamezne serije poskusa
N - število VSEH možnih izidov
M - število VSEH ugodnih izidov

$$P=\frac{\binom{M}{k}*\binom{N-M}{n-K}}{\binom{N}{n}}$$

PRIMER: V posodi imamo 3 bele, 4 zelene in 5 modrih kroglic. Izberemo 6 kroglic, kakšna je verjetnost, da je med njimi točno 1 bela, 2 zeleni in 3 modre?

$$P=\frac{\binom{3}{1}*\binom{4}{2}*\binom{5}{3}}{\binom{12}{6}}=\frac{\frac{\frac{3!}{1!*(3-1)!}*4!}{2!*(4-2)!}*5!}{3!*(5-3)!}}{\frac{12!}{6!*(12-6)!}}=\frac{\frac{\frac{3*24}{4}*120}{6*2}}{\frac{12*11*10*9*8*7}{6*5*4*3*2*1}}=\frac{3*6*10}{11*7*4*3}=\frac{180}{924}=0{,}1948=19{,}48$$

Parameter: značilnost populacije kot celote
Spremenljivka: značilnost posamezne enote
Statistična enota: posamezen pojav statistične množice
Statistična populacija: množica vseh istovrstnih pojavov – statističnih enot, ki so bili izbrani za namen analize; ti pojavi izpolnjujejo opredeljujoče pogoje (časovna, krajevna in stvarna opredelitev)

Trije opredeljujoči pogoji proučevane populacije:

- **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 in dogodki/dogajanja**

Vrste podatkov:

PRIMARNI PODATKI so tisti, ki jih zberemo sami, glede na naše potrebe.

SEKUNDARNI PODATKI, ki jih običajno zbira, na podlagi neke formalne, zakonodajne podlage, uradna institucija (Statistični urad, Banka Slovenije, ministrstva). V te podatke je običajno vključena zelo velika populacija. Ti podatki so bolj verodostojni, bolj popolni in kontrolirani.

Vrste spremenljivk:

Nominalne spremenljivke: ali sta dve vrednosti enaki ali ne (Primer: spol [moški/ženski]. Ne moremo reči povprečen spol, ne moremo dveh spolov seštevati). Edini parameter, ki ga lahko ugotovimo tako da preštejemo kolikokrat se katera vrednost ponovi je modus.

$$y_a \neq y_b$$

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).

$$y_a > y_b$$

Intervalne spremenljivke so številske spremenljivke. 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.

$$y_a - y_b$$

Razmernostne spremenljivke Razmerja lahko računamo za tiste spremenljivke, ki so po značilnostih razmernostne spremenljivke. To so tiste, ki lahko zavzamejo samo pozitivne ali amo 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.

$$y_a / y_b$$

Načina izbire vzorčenja:

1. **Enostavno slučajno vzorčenje**
    - enote imajo enako možnost, da bodo izbrane v vzorec
    - vsi vzorci (ker je več različnih) imajo enako možnost da bodo izbrani
    - objektivnost pri izbiri enot v vzorec --> velika verjetnost, da bo vzorec reprezentativen
    - možno ocenjevanje kakovosti vzorčnih ocen s posebnimi kazalci

2. **Neslučajni vzorci**

    - ni znana verjetnost izbora posamezne enote v vzorec
    - ni mogoče izračunati ustreznih kazalcev kakovosti ocene
    - načeloma ni mogoče sklepati iz vzorca na populacijo

To so 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.

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).

STANDARDNA NAPAKA OCENE nam kaže, koliko se ocene vzorcev med sabo razlikujejo. 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.

Ocenjevanja parametrov (stopnja tveganja α in stopnja tveganja 1-α):

- z mejami oz. intervali zaupanja, pri velikih vzorcih
- s preizkušanjem domnev o vrednostih parametrov, pri majhnih vzorcih
-

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)**

Povprečje je prava vrednost, minus (ali 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 ali: 68% verjetnost je, da dobimo oceno, ki je nekje v intervalu od prave vrednosti navzdol in navzgor za eno standardno napako). Pri +/- 2 standardni napake je teh vzorcev 95% (95% verjetnost), pri +/- 3 pa 99%

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).

1) Dvostranska intervalna ocena:

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

g - točkovna ocena
z - v standardizirani normalni porazdelitvi je vrednost 1,96
G - tisto kar ocenjujemo, populacijska vrednost parametra
SE - standardna napaka

Trije primeri intervalov zaupanja:

- Ocena aritmetične **sredine**

$$\overline{y} - z * se(\overline{y}) < \overline{y} + z * se(\overline{y}); \alpha$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$ (aritmetična sredina = vsota spremenljivk/število enot v vzorcu)

$$SE(\overline{y}) = \frac{\sigma_y}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} \cong \frac{\sigma_y}{\sqrt{n}}$$ (standardna napaka populacije ≈ standardni odklon populacije/koren števila spremenljivk)

$$se(\overline{y}) \cong \frac{s(y)}{\sqrt{n}}$$ (standardna napaka vzorca ≈ standardni odklon vzorca/koren števila spremenljivk)

$$\frac{\sum_{i=1}^{n} (¿¿ i - \overline{y})^2}{n-1}, s_y = \sqrt{s_y^2}$$

$$s_y^2 = ¿$$

- Ocena **vsote** vrednosti

$$Y_{SL} - z * se(Y_{SL}) < Y_{SL} + z * se(Y_{SL}); \alpha$$,

$$Y_{SL} = N * \overline{y}, se(Y_{SL}) = N * se(\overline{y})$$

- Ocena **deleža** enot , ki imajo določeno lastnost, odstotka

$$p - z * se(p) < p < p + z * se(p)$$

$$p=\frac{n_a}{n}, p=100*p=\frac{100*n_a}{n}, N_a=N*p$$

$$se(p)=\sqrt{\frac{p*(1-p)}{n-1}}*\sqrt{\frac{N-n}{N}}$$ (drugi del formule za množenjem v praksi velikokrat spustimo, ker je pri pri velikem številu statističnih enot v populaciji in majhnem vzorcu največkrat skoraj enak 1: če imamo npr. populacijo 200.000 enot, vzorec pa 100, potem je drugi del 199.900/200.000, kar je praktično 1)

$$se(p)\cong\sqrt{\frac{p*(1-p)}{n-1}}\cong 100*se(p)$$

$$se(N_a)=N*se(p)$$

2) Preizkus domneve o aritmetični sredini:

Primer: Ali se trend zmanjševanja velikosti gospodinjstev nadaljuje?

a) **najprej postavimo domnevi**

H_0: $M_y \geq 3274$
H_1: $M_y < 3274$ (v alternativni je tisto, kar skušamo dokazati)

b) **nato naredimo t-preizkus po formuli**

$$t=\frac{\overline{y}-M_y}{se(\overline{y})}=\frac{\overline{y}-M_y}{s_y}*\sqrt{n}$$

ali z SPSS, v tem primeru izpis pokaže:

One-Sample Statistics

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

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

Stopnja tveganja je v tem primeru 0,074 (sig. 2-tailed); Za nas sprejemljiva stopnje je 0,05 razen če je določeno drugače (npr. naročnik zahteva manjšo stopnjo tveganja, npr. α=0,01)

Ker pa moramo v našem primeru narediti enostranski preizkus, ne dvostranskega (ugotavljamo, ali se trend zmanjšuje: dvostranski bi bil, ali je število enako 3274 ali ne), moramo dvostranski sig. deliti na pol:

torej: 0,074/2 = 0,037

c) **podamo sklep**

Ker je stopnja tveganja 0,037 < 0,05, lahko ničelno domnevo zavrnemo in sprejmemo alternativno. Torej:

Na podlagi vzorčnih podatkov lahko pri stopnji tveganja α=0,037 zavrnemo ničelno domnevo in sprejmemo sklep, da se trend zmanjševanja gospodinjstev nadaljuje oz. da se je povprečno število gospodinjstev v primerjevi z letom 1993 zmanjšalo

3) Preizkus domneve o razliki med dvema aritmetičnima sredinama:

Primer: Ali so avtomobili za alarmno napravo v povprečju dražji kot avtomobili brez alarmne naprave?

a) **najprej postavimo domnevi**

H_0: $\overline{Y}$ brez alarma ≥ $\overline{Y}$ z alarmom

H_1: $\overline{Y}$ brez alarma < $\overline{Y}$ z alarmom (v alternativni je tisto, kar skušamo dokazati)

b) **nato preverimo predpostavko o enakosti varianc oz. naredimo preizkus domneve o enakosti varianc**

H_0: σ^2 brez alarma = σ^2 z alarmom

H_1: σ^2 brez alarma ≠ σ^2 z alarmom

To naredimo z Levenovim testom

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

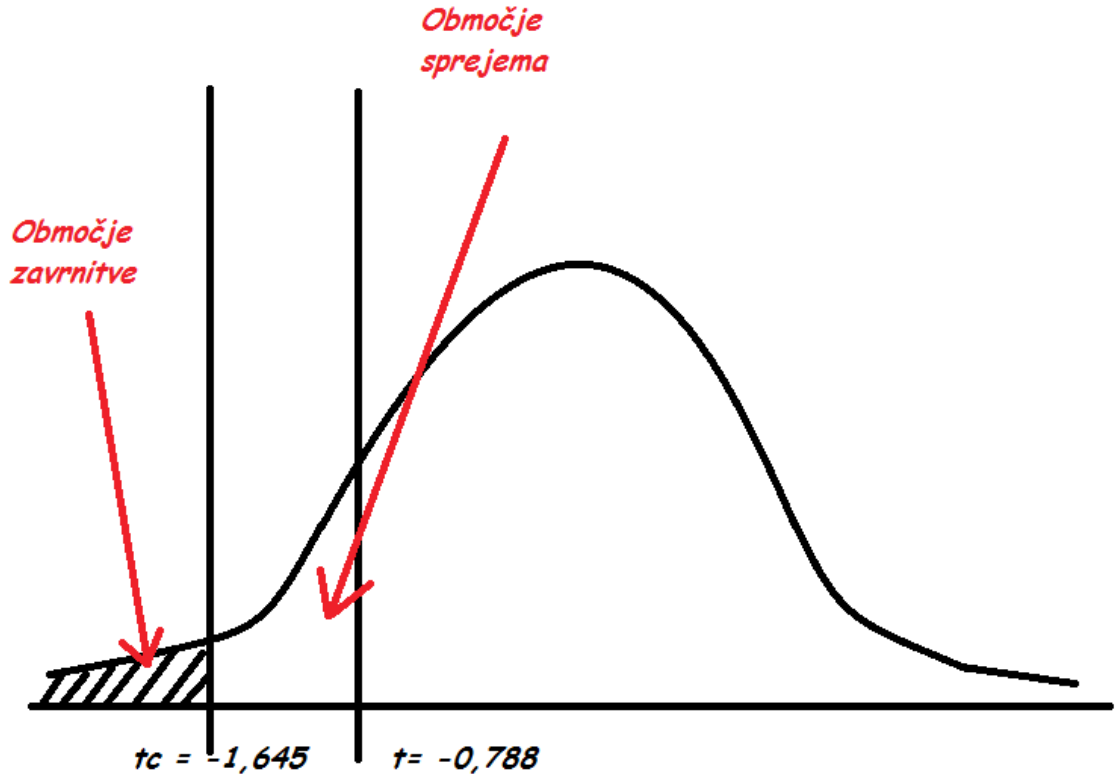
Vidimo, da je stopnja značilnosti (sig = 0,414) previsoka (dovoljenja je do 0,05), zato ničelne domneve **<u>o enakosti varianc</u>** (govorimo samo o enakosti varianc, ne o rešitvi naloge) ne moremo zavrniti.

Podatke bomo zato uporabili iz prve vrstice tabele (equal variances assumed), ne pa iz druge (equal variances not assumed)

c) **izvedba t-preizkusa**

Kot vidimo, je t = -0,788

Ugotoviti je potrebno, ali se t nahaja v območju zavrnitve, ali v območju kritičnega sprejema. Za to potrebujemo t_c, ki ga odčitamo iz tabele; pri α=0,05 in vzorcu, ki je večji od 30, je ta 1,645 oz. -1,645 (odvisno, kako postavimo domnevo). Lažje si bo predstavljati v izrisani obliki:



Kot vidimo, se t nahaja v območju sprejema. To pomeni, da ničelne domneve, ki smo jo postavili na začetku, ne moremo zavrniti.

Na podlagi vzorca rabljenih avtomobilov ni mogoče pri dovolj nizki stopnji tveganja zavrniti ničelne domneve, zaradi česar ne moremo trditi, da so avtomobili z vgrajeno alarmno napravo v povprečju dražji.

4) Preizkus domneve o več aritmetičnih sredinah:

Primer: Ali lahko na podlagi danega vzorca sklepamo, da se povprečna raven opremljenosti razlikuje med avtomobili različnega porekla?

a) **najprej postavimo domnevi**

H$_0$: $\overline{Y}_{it}=\overline{Y}_{de}=\overline{Y}_{fr}=\overline{Y}_{jp}=\overline{Y}_{ostalo}$

H$_1$: vsaj ena $\overline{Y}_t$ se razlikuje

b) **nato preverimo predpostavko o enakosti varianc oz. naredimo preizkus domneve o enakosti varianc**

H_0: $\sigma^2_{it} = \sigma^2_{de} = \sigma^2_{fr} = \sigma^2_{jp} = \sigma^2_{ostalo}$

H_1: vsaj ena σ^2_t se razlikuje

To naredimo z Levenovim testom

Test of Homogeneity of Variances

Raven opreme v odstotkih

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

Iz nizke stopnje značilnosti (sig. 0,027 < 0,05) je razvidno, da ničelno domnevo lahko zavrnemo in sprejmemo alternativno, da se vsaj ena σ^2 razlikuje. **Predpostavimo torej, da variance niso enake. V nadaljevanju uporabimo torej robustni Welch test** (v primeru, da bi bile variance enake, bi uporabili F-preizkus oz. Anovo)

c) **robustni Welch preizkus**

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.

Iz priloženega preizkusa vidimo, da lahko zaradi nizke stopnje tveganja (α=0,005, kar je manjše od sprejemljivega 0,05) ničelno domnevo zavrnemo.

Na podlagi vzorca rabljenih avtomobilov lahko pri stopnji tveganja α=0,005 zavrnemo ničelno domnevo in sprejmemo sklep, da se povprečna raven opremljenosti razlikuje med skupinami avtomobilom različnega porekla.

5) Analiza linearne odvisnosti:

Primer:

n = 33 (število občin)
VIS = število preb. z visoko strokovno izobrazbo
VIS_DOD = število preb. z uni izobrazbo
POD = število preb. s podiplomsko izobrazbo
BREZP = število brezposelnih
DEL_AKT = število delovno aktivnih prebivalcev

a) **Oblikovanje relativnih kazalnikov**

Ker so občine različno velike, moramo analizo delati na relativnih kazalnikih (izločimo vpliv velikosti)

ST_BREZP = BREZP / (DEL_AKT + BREZP) * 100

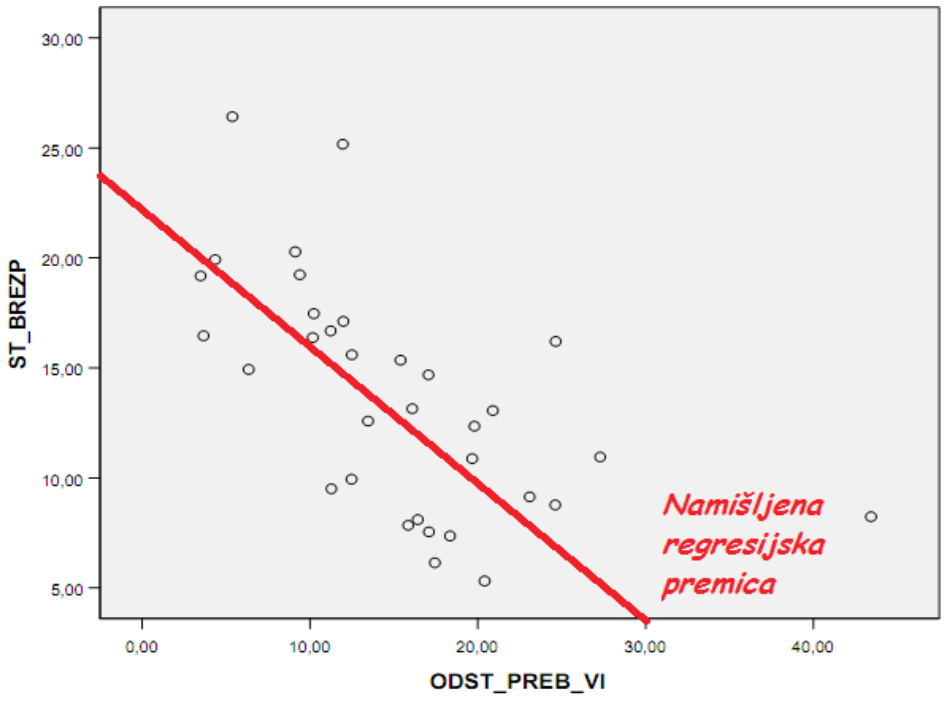
Odstotek brezposelnosti = št. brezposelnih / (delovno aktivnih + brezposelnih) * 100

ODST_PREB_VI = (VIS + VIS_DOD + POD) / (DEL_AKT + BREZP) * 100

Odstotek visoko izobraženih = (visoka strokovna + uni + podiplomska) / (delovno aktivnih + brezposelnih) * 100

b) **Razsevni diagram**

Razsevni diagram iz SPSS nam pokaže naslednjo sliko (dodana je namišljena linearna regresijska premica, čeprav je iz diagrama razvidno, da funkcija ni čisto linearna)



c) **Analiza razsevnega diagrama**

- odvisnost stopnje brezposelnosti od števila prebivalcev z VI izobrazbo je, in je negativna - večji kot je odstotek prebivalstva z VI, manjša je brezposelnost
- Odvisnost ni povsem linearna (regresijska premica je približek; v tem primeru bi bila bolj točna krivulja)
- odvisnost je šibka, ker so točke precej oddaljene od namišljene regresijske premice

d) **Ocena linearnega regresijskega modela**

Splošni bivariantni linearni regresijski model:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Enačba regresijske premice:

y' = α + βx

Na vzorčnem primeru pa:

ST_BREZP'' = a + b * ODST_PREB_VI

e) **Zapišemo enačbo regresijske premice in komentiramo regresijske koeficiente:**

a. Predictors: (Constant), ODST_PREB_VI
b. Dependent Variable: ST_BREZP

a

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	19,707	1,639		12,021	,000
	ODST_PREB_VI	-,393	,095	-,595	-4,127	,000

a. Dependent Variable: ST_BREZP

b

a = 19,707
b = -0,393

Torej: ST_BREZP'' = 19,707 - 0,393 * ODST_PREB_VI

a = konstanta; ni je potrebno razlagati, pove pa nam, kakšen bi bil delež breposelnih, če ne bi bilo nobenega prebivalca z visokošolsko izobrazbo; v praksi tega ni

b:

Na podlagi vzorca občin ocenjujemo, da se stopnja brezposlenosti v povprečju zmanjša za 0,393 odstotne točke, če se odstotek prebivalcev z visokošolsko izobrazbo poveča za 1 odstotno točko.

<u>Pozor: ker smo računali v relativnih kazalnikih (%), se zmanjša za %, ne za število!</u>

f) **Izvedemo preizkus neodvisnosti**

H_0: $\beta = 0$ (spremenljivki sta neodvisni)

H_1: $\beta \neq 0$ (spremenljivki sta odvisni)

Iz izpisa je razvidno, da je t = -4,127, t_c pa razberemo iz tabele; pri vzorcu nad 30 je pri stopnji tveganja α = 0,05 enak -1,645

t se torej nahaja levo od t_c, znotraj območja kritične zavrnitve. Stopnja tveganja je v tem primeru zanemarljiva, saj iz izpisa vidimo, da je sig. enako 0,000

Na podlagi vzorčnih podatkov lahko pri zanemarljivi stopnji tveganja zavrnemo ničelno domnevo in sprejmemo sklep, da je stopnja brezposelnosti odvisna od odstotka prebivalcev z visokošolsko izobrazbo.

Podvprašanje: Ali lahko trdimo, da je v občinah z višjim odstotkom prebivalcev z visokošolsko izobrazbo stopnja brezposelnosti v povprečju nižja?

H0: $\beta \geq 0$
H1: $\beta < 0$

Izračunamo t (t= b/se(b))

Odčitamo t_c pri α=0,05 in vzorcu nad 30 - t_c = -1,645 in ugotovimo, da se nahaja v območju kritične zavrnitve.

Na podlagi vzorca slovenskih občin lahko pri stopnji tveganja α=0,05 zavrnemo ničelno domnevo in sprejmemo sklep, da je v občinah z višjim odstotkom prebivalcev z visokošolsko izobrazbo stopnja brezposelnosti v povprečju nižja.

g) **Komentiramo determinacijski in korelacijski koeficient**

- Korelacijski koeficient $-1 \leq \rho_{yx} \leq 1$
- Do 0,6 šibka povezanost
- Do 0,75 srednje močna
- Do 0,85 mo
- na
- Nad 0,9 zelo močna

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,595[a]	,355	,334	4,32383

a. Predictors: (Constant), ODST_PREB_VI

$$r_{yx}^2 = 0{,}355$$

Na podlagi ocenjenega determinacijskega koeficienta ugotavljamo, da je 35% variance stopnje brezposelnosti pojasnjene z linearnim vplivom odstotka prebivalcev z visokošolsko izobrazbo. Preostalih 65% variance povzročajo neznani, med njimi vsaj slučajni dejavniki.

$$r_{yx} = -0{,}595$$

Na podlagi ocene korelacijskega koeficienta ocenjujemo, da je povezanost med stopnjo brezposelnosti in odstotkom prebivalcev z visokošolsko izobrazbo negativna (zato dodamo minus - vidno iz premice) in šibka (0,565 < 0,6).

6) Multipla linearna odvisnost (multipla regresijska analiza)

Primer:

Analizirali smo odvisnost cene stanovanj od njihove <u>površine</u>, <u>starosti</u> ter <u>oddaljenosti od centra mesta</u>. V vzorec smo zajeli 30 stanovanj.

povm2 = površina stanovanja v m^2
star = starost
odd = oddaljenost

α = regresijska konstanta
β_k = parcialni regresijski koeficient
ε = slučajnostni odkloni

a) Ocena multivariantnega linearnega regresijskega modela

Splošni multi variantni linearni regresijski model:

$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon_i$$

Enačba regresijske hiperravnine:

$$\mathbf{y' = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}$$

Na vzorčnem primeru pa:

cena'' = a + b_1 * povm2 + b_2 * star + b_3 * odd

b) Pearsonova korelacijska matrika

Correlations

		CENA	POVM2	STAR	ODD
Pearson Correlation	CENA	1,000	,925	,196	-,286
	POVM2	,925	1,000	,308	-,077
	STAR	,196	,308	1,000	-,006
	ODD	-,286	-,077	-,006	1,000
Sig. (1-tailed)	CENA	.	,000	,150	,063
	POVM2	,000	.	,049	,343
	STAR	,150	,049	.	,487
	ODD	,063	,343	,487	.
N	CENA	30	30	30	30
	POVM2	30	30	30	30
	STAR	30	30	30	30
	ODD	30	30	30	30

Korelacijski koeficienti so:

med ceno in površino: 0,925
med ceno in starostjo: 0,196
med ceno in oddaljenostjo: -0,286

Korelacijski koeficient, ki kaže odvisnost med spremenljivkama <u>površina</u> stanovanja in <u>oddaljenost</u> stanovanja od centra znaša -0,077, kar pomeni, da je odvisnost med spremenljivkama negativna in zelo šibka.

Izvedemo test neodvisnosti:

H_0: $\rho_{povm2,\ odd} = 0$
H_1: $\rho_{povm2,\ odd} \neq 0$

Ker nas zanima samo, ali sta spremenljivki neodvisni ali ne, gre za dvostranski preizkus (enostranski bi bil večji/manjši). Iz zapisa nato razberemo stopnjo značilnosti, ki pa je enostranska (sig. 1-tailed).

Zato jo moramo množiti z 2: sig. 2-tailed = 0,343*2 = **0,686**

Stopnja tveganja pri zavrnitvi ničelne domneve je torej 0,686, za nas sprejemljiva pa je 0,05.

Na podlagi vzorca stanovanj ni mogoče pri dovolj nizki stopnji tveganja zavrniti ničelne domneve, zato ne moremo trditi, da sta spremenljivki povezani. (težav z multikolinearnostjo ni pričakovati).

c) **Ocena regresijskega modela:**

Vprašanje: Ali vsaj ena izmed spremenljivk vpliva na ceno stanovanja?

Postavimo domnevi:

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
H_1: vsaj en $\beta_i \neq 0$

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	805,587	3	268,529	86,857	,000[a]
	Residual	80,382	26	3,092		
	Total	885,969	29			

a. Predictors: (Constant), ODD, STAR, POVM2

b. Dependent Variable: CENA

Iz izpisa vidimo, da F = 86,857, pri stopnji značilnosti sig. = 0,000; to pomeni:

Na podlagi vzorca stanovanj lahko pri zanemarljivi stopnji tveganja zavrnemo ničelno domnevo in sprejmemo sklep, da je vsaj eden izmed regresijskih koeficientov različen od 0 (vsaj ena izmed vključenih spremenljivk dejansko vpliva na ceno stanovanja).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,954[a]	,909	,899	1,75830

a. Predictors: (Constant), ODD, STAR, POVM2

Iz izpisa vidimo, da je popravljeni determinacijski koeficient enak 0,899:

$$\overline{r^2} = 0{,}899$$

Na podlagi vzorčnih podatkov ocenjujemo, da je (popravljeni) multipli determinacijski koeficient enak 0,899, kar pomeni, da je 89,9% variance cene stanovanj pojasnjeno z linearnim vplivom površine stanovanj, starosti stanovanj in oddaljenosti od centra mesta. Preostalih 10,1 odstotkov povzročajo neznani dejavniki.

Ocenimo še regresijsko funkcijo:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,207	1,078		2,976	,006
	POVM2	,205	,014	,937	15,044	,000
	STAR	-,049	,033	-,094	-1,512	,143
	ODD	-,164	,045	-,214	-3,616	,001

a. Dependent Variable: CENA

CENA'' = 3,207 + 0,205*POVM2 - 0,049*STAR - 0,164*ODD

Ali starost stanovanja vpliva na ceno?

Najprej postavimo domnevi:

H_0: $\beta_2 = 0$
H_1: $\beta_2 \neq 0$

t pri spremenljivki STAR = -1,512, stopnja tveganja (sig) pa = 0,143

Ker je stopnja tveganja večja od sprejemljive 0,05, ničelne domneve ne moremo zavrniti.

Na podlagi vzorčnih podatkov ne moremo zavrniti ničelne domneve, da je regresijski koeficient β_2 enak nič. Ne moremo torej trditi, da starost stanovanj vpliva na njihovo ceno.

Očitno torej na ceno vplivata ostala dva dejavnika: Oceno zato ponovimo, vendar v drugačni obliki - brez starosti.

cena'' = a + b_1*povm2 + b_2*odd

Dobimo naslednji izpis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,949[a]	,901	,894	1,79966

a. Predictors: (Constant), ODD, POVM2

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	798,522	2	399,261	123,276	,000[a]
	Residual	87,447	27	3,239		
	Total	885,969	29			

a. Predictors: (Constant), ODD, POVM2

b. Dependent Variable: CENA

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,796	1,067		2,620	,014
	POVM2	,199	,013	,908	14,973	,000
	ODD	-,165	,046	-,216	-3,561	,001

a. Dependent Variable: CENA

Popravljeni determinacijski koeficient se je zmanjšal za 0,005 in znaša 0,894 (namesto 0,899).

Na podlagi F-statistike (Anova) lahko še pri manjši stopnji tveganja (0,000) zavrnemo ničelno domnevo in sprejmemo sklep, da je vsaj en regresijski koeficient različen od 0 (vsaj ena spremenljivka vpliva na ceno).

CENA = 2,796 +0,199*POVM2 - 0,165*ODD; Torej:

Na podlagi vzorčnih podatkov ocenjujemo, da se cena stanovanja v povprečju poveča za 199.000 SIT, če se površina stanovanja poveča za kvadratni meter, oddaljenost od centra mesta pa ostane nespremenjena.

Na podlagi vzorčnih podatkov ocenjujemo, da se cena stanovanja v povprečju zmanjša za 165.000 SIT, če se oddaljenost od centra mesta poveča za en kilometer, površina stanovanja pa ostane nespremenjena.

Obe spremenljivki statistično značilno vplivata na ceno stanovanj.

7) Faktorska analiza:

Uporaba na področju pojavov, ki jih ni mogoče točno izmeriti; miselni procesi, npr. zadovoljstvo, vzdušje oz. org. klima; lahko se preverja npr. z anketo, vendar so to samo posredni kazalniki. Analizo teh kazalnikom pa omogoča faktorska analiza.

Pri faktorski analizi opravljamo analizo medsebojnih korelacij spremenljivk (npr. dvajsetih); ugotavljanje skupnih faktorjev med posameznimi skupinami spremenljivk (npr. x1, 2, 4 imajo skupni faktor, x6, 7 in 10, pa x3 in x5…). Spremenljivke lahko ostanejo tudi »same« oz. niso povezane z ostalimi (vpliv specifičnih faktorjev).

Potrebno je izmeriti, koliko posamezna »utež« vpliva na spremenljivke. Merimo samo tiste, pri katerih je faktorska utež dovolj velika; Faktorji naj bi bili med seboj neodvisni. Variabilnost, ki jo povzročajo drugi dejavniki, naj bi bila majhna.

Komunaliteta: delež variance, pojasnjen s skupnimi faktorji (h_i^2)

Primer:

Ugotoviti bomo poskušali, ali obstajajo določeni skupni dejavniki (faktorji) s pomočjo katerih je mogoče pojasniti uspešnost dijakov. Glede na dejstvo, da uspešnosti ne moremo neposredno meriti smo za indikatorje uspešnosti izbrali ocene pri določenih predmetih. V vzorec je bilo vključenih 45 dijakov.

a) **Prikaz porazdelitev dijakov glede na ocene posameznega predmeta**

Preverimo ali so porazdelitve dijakov, glede na ocene, podobne normalni porazdelitvi, ki jo predpostavlja metoda največjega verjetja, ali pa so asimetrične.

b) **Oblikovanje korelacijske matrike**

Ocenimo, kako močne so odvisnosti med ocenami posameznih predmetov.

Correlations

		ANG	MAT	GEO	SJK	FIZ	ZGO	GV
ANG	Pearson Correlation	1	,107	,138	,316*	,164	,021	,129
	Sig. (2-tailed)		,483	,367	,034	,280	,890	,398
	N	45	45	45	45	45	45	45
MAT	Pearson Correlation	,107	1	,043	,076	,524**	-,072	,079
	Sig. (2-tailed)	,483		,778	,621	,000	,640	,605
	N	45	45	45	45	45	45	45
GEO	Pearson Correlation	,138	,043	1	,067	,012	,179	,161
	Sig. (2-tailed)	,367	,778		,661	,939	,239	,292
	N	45	45	45	45	45	45	45
SJK	Pearson Correlation	,316*	,076	,067	1	-,169	,269	,001
	Sig. (2-tailed)	,034	,621	,661		,266	,074	,997
	N	45	45	45	45	45	45	45
FIZ	Pearson Correlation	,164	,524**	,012	-,169	1	-,210	,108
	Sig. (2-tailed)	,280	,000	,939	,266		,165	,481
	N	45	45	45	45	45	45	45
ZGO	Pearson Correlation	,021	-,072	,179	,269	-,210	1	,149
	Sig. (2-tailed)	,890	,640	,239	,074	,165		,330
	N	45	45	45	45	45	45	45
GV	Pearson Correlation	,129	,079	,161	,001	,108	,149	1
	Sig. (2-tailed)	,398	,605	,292	,997	,481	,330	
	N	45	45	45	45	45	45	45

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Povezanost med posameznimi spremenljivkami je relativno nizka, razen pri (SJK - ANG = 0,316 in MAT - FIZ = 0,524) Očitno je, da na ocene pri posameznem predmetu vplivajo v veliki meri specifični dejavniki, oziroma lahko pričakujemo relativno nizek vpliv skupnih dejavnikov.

c) **Ocena faktorskega modela s pomočjo metode glavnih osi (principal axis factoring)**

Glede na vrednosti v korelacijskih matriki oziroma glede na majhno število spremenljivk je smiselno predpostaviti, da je **linearno povezanost** med ocenami predmetov **mogoče pojasniti z dvema oziroma kvečjemu tremi skupnimi faktorji** uspešnosti dijakov.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,487
Bartlett's Test of Sphericity	Approx. Chi-Square	32,126
	df	21
	Sig.	,057

Na podlagi KMO statistike (vrednost 0,487), ki je manjša od 0,5 ocenjujemo, da podatki niso povsem primerni za faktorsko analizo (to smo načeloma že ugotovili ob analizi korelacijske matrike, iz katere smo razbrali relativno velik vpliv specifičnih dejavnikov).

S pomočjo metode glavnih osi ne pridemo do ocen faktorskih uteži, niti do ocen komunalitet. Postopek ne skonvergira (slabost metode). Zato poskusimo z drugo oceno.

d) **Ocena faktorskega modela s pomočjo metode največjega verjetja (maximum likelihood) – brez rotacije faktorjev**

Communalities[a]

	Initial	Extraction
ANG	,176	,216
MAT	,305	,311
GEO	,066	,025
SJK	,233	,614
FIZ	,368	,999
ZGO	,152	,133
GV	,071	,021

Extraction Method: Maximum Likelihood.

a. One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Pri ocenah FIZ, SJK in MAT lahko večji del ocene pojasnimo s predpostavljenima skupnima faktorjema uspešnosti (99,9%, 61,4% in 31,1%)

Pri ocenah GV in GEO (2,1% in 2,5%) pa nasprotno opazimo, da na ocene skoraj v celoti vplivajo specifični dejavniki.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1,671	23,876	23,876	1,385	19,792	19,792
2	1,593	22,758	46,634	,933	13,325	33,116
3	1,086	15,518	62,152			
4	,859	12,275	74,427			
5	,831	11,870	86,297			
6	,577	8,246	94,542			
7	,382	5,458	100,000			

Extraction Method: Maximum Likelihood.

Iz tabele lahko razberemo, da lahko z dvema skupnima faktorjema pojasnimo približno 47% celotne variabilnosti izpitnih ocen.

Iz naslednje tabele matrike faktorskih uteži so za prvi faktor razvidne visoke uteži pri predmetu MAT in FIZ medtem, ko kaže drugi faktor močan vpliv na ocene SJK in ANG. Ti dve skupini predmetov oziroma povezanost med njimi je bila razvidna tudi že iz korelacijske matrike.

Factor Matrix[a]

	Factor	
	1	2
ANG	,164	,434
MAT	,524	,190
GEO	,012	,157
SJK	-,170	,765
FIZ	,999	,000
ZGO	-,210	,298
GV	,108	,098

Extraction Method: Maximum Likelihood.

a. 2 factors extracted. 11 iterations required.

Glede na sorodnost med omenjenima dvojicama predmetov bi lahko rekli, da prvi faktor predstavlja nadarjenost dijakov za naravoslovne vsebine, drugi faktor pa njihovo nadarjenost za jezike.

e) **Ocena faktorskega modela s pomočjo metode največjega verjetja (maximum likelihood) – <u>poševna rotacija faktorjev</u>**

V tabeli so predstavljene "pattern" uteži, ki predstavljajo regresijske koeficiente.

Pattern Matrix[a]

	Factor	
	1	2
ANG	,236	,372
MAT	,552	,033
GEO	,038	,148
SJK	-,041	,788
FIZ	,993	-,287
ZGO	-,160	,348
GV	,123	,064

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 7 iterations.

V spodnji tabeli pa vidimo strukturne uteži, ki predstavljajo korelacijske koeficiente med posamezno spremenljivko in faktorjem.

Structure Matrix

	Factor	
	1	2
ANG	,281	,401
MAT	,556	,101
GEO	,056	,152
SJK	,055	,783
FIZ	,958	-,165
ZGO	-,117	,329
GV	,131	,079

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.

V obeh primerih lahko vidimo, da so na prvem faktorju najbolj izrazite uteži pri predmetih MAT in FIZ, na drugem faktorju pa pri predmetih SJK, ANG in recimo še ZGO.

Na podlagi rotirane rešitve se kaže, da drugi dejavnik najverjetneje predstavlja nadarjenost za družboslovne vsebine (nekoliko širše gledano) in ne nujno ožje opredeljeno samo za področje jezikov. Sicer pa se razlaga celotne strukture faktorskih uteži bistveno ne spremeni niti po rotaciji.

Factor Correlation Matrix

Factor	1	2
1	1,000	,123
2	,123	1,000

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.

Glede na šibko korelacijo med faktorjema prikazano v tabeli Factor Correlation Matrix bi bilo smiselno namesto poševnokotne rotacije opraviti pravokotno rotacijo faktorjev.

f) **Ocena faktorskega modela s pomočjo metode največjega verjetja (maximum likelihood) – <u>pravokotna rotacija faktorjev</u>**

Rotated Factor Matrix[a]

	Factor	
	1	2
ANG	,263	,382
MAT	,554	,059
GEO	,049	,149
SJK	,018	,783
FIZ	,971	-,238
ZGO	-,133	,340
GV	,128	,069

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Ponovno lahko vidimo, da so na prvem faktorju izrazite uteži pri predmetih MAT in FIZ, na drugem faktorju pa SJK, ANG in ZGO.

SKLEP:

Ocene faktorskih uteži so ne glede na uporabljeno rotacijo bolj ali manj enake na podlagi česar lahko sklepamo na stabilnost ocen.

Zaključimo lahko naslednje:

- pri predmetih MAT in FIZ so izrazite uteži na prvem faktorju in manjše uteži na drugem faktorju
- pri predmetih SJK, ANG in načeloma ZGO so relativno visoke uteži na drugem faktorju (>0,4), predmeta GV in GEO pa nimata izrazitih uteži na nobenem faktorju kar pomeni, da so ocene pri teh predmetih večinoma pojasnjene s specifičnimi dejavniki.

Ker je uspešnost pri predmetih GV in GEO pojasnjena večinoma s specifičnimi dejavniki, se odločimo, da ju bomo izločili iz modela.

g) **Ocena faktorskega modela s pomočjo metode največjega verjetja (maximum likelihood) – poševna rotacija faktorjev (ponovno, tokrat brez GV in GEO)**

Še vedno lahko vidimo največje deleže pojasnjene variance s skupnima dejavnikoma pri predmetih SJK (pri tem predmetu se je delež pojasnjene variance povečal približno za polovico) in FIZ.

Communalities[a]

	Initial	Extraction
ANG	,152	,149
MAT	,304	,303
SJK	,229	,999
FIZ	,364	,999
ZGO	,101	,100

Extraction Method: Maximum Likelihood.

a. One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	1,651	33,013	33,013	1,362	27,249	27,249	1,209
2	1,440	28,802	61,815	1,187	23,749	50,998	1,342
3	,884	17,678	79,493				
4	,642	12,832	92,325				
5	,384	7,675	100,000				

Extraction Method: Maximum Likelihood.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Vidimo, da lahko sedaj z dvema skupnima dejavnikoma pojasnimo že skoraj 62% skupne variabilnosti v vzorcu.

Factor Matrix[a]

	Factor	
	1	2
ANG	,099	,373
MAT	-,293	,465
SJK	,764	,644
FIZ	-,765	,644
ZGO	,313	,045

Extraction Method: Maximum Likelihood.

a. 2 factors extracted. 9 iterations required.

Prvi faktor je bipolaren saj kaže hkrati močan pozitiven vpliv na ocene SJK ter negativen vpliv na ocene pri FIZ.

Glede na to, da so uteži na drugem faktorju visoke skoraj pri vseh predmetih lahko označimo ta faktor kot splošni faktor.

Običajno bipolarne in splošne faktorje težje smiselno pojasnimo zato smo se odločili za rotacijo faktorjev (želimo priti do enostavnejše faktorske strukture).

Po uporabljeni rotaciji se struktura faktorskih uteži približa že ocenjeni strukturi v modelu z vključenimi vsemi ocenami.

Factor Correlation Matrix

Factor	1	2
1	1,000	-,003
2	-,003	1,000

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.

Glede na to, da sta faktorja skoraj neodvisna (korelac. koeficient znaša -0,003) bi bilo smiselno uporabiti še pravokotno rotacijo in primerjati strukturo uteži.

h) **Sklepne ugotovitve**

S pomočjo faktorske analize smo skušali opredeliti morebitne skupne dejavnike (faktorje), s katerimi je mogoče pojasniti razlike v uspešnosti dijakov (uspešnost dijakov smo merili z ocenami izbranih predmetov).

Na podlagi **korelacijske matrike** smo ugotovili, da je **povezanost** med ocenami različnih predmetov **relativno nizka**, kar pomeni, da **na ocene** pri posameznem predmetu **vplivajo** v veliki meri **specifični dejavniki**. Največja je sicer odvisnost med ocenami pri predmetu ANG in SJK, oziroma MAT in FIZ.

Glede na vrednosti v korelacijskih matriki oziroma glede na število spremenljivk smo **predpostavili faktorski model z dvema skupnima faktorjema** s katerima lahko pojasnimo nekaj manj kot polovico variabilnosti učnega uspeha.

Za oceno komunalitet smo uporabil metodo glavnih osi ter metodo največjega verjetja. Prva metoda ni dala ustrezne rešitve (ni skonvergirala).

Na podlagi ocen komunalitet je bilo mogoče ugotoviti najvišje deleže pojasnjene variance s skupnima faktorjema pri ocenah predmetov FIZ (99,9% variabilnosti ocen pojasnjene s skupnima faktorjema uspešnosti), SJK (61% variabilnosti pojasnjene s skupnima faktorjema) in MAT (31% variabilnosti pojasnjene s skupnima faktorjema). Izredno nizek delež variance pojasnjene s skupnima faktorjema pa smo ugotovil pri predmetih GEO in GV. Na ocene pri slednjih skoraj v celoti vplivajo specifični dejavniki.

Faktorske uteži smo skušali oceniti s **poševnokotno** in **pravokotno rotacijo faktorjev**. Ocene uteži ob uporabljeni rotaciji faktorjev se niso bistveno razlikovale od ocen uteži brez uporabljene rotacije.

Do manjših sprememb je prišlo le pri drugem skupnem faktorju. Na prvem skupnem faktorju so bile najvišje uteži pri predmetih FIZ in MAT, na drugem pa pri predmetih SJK in ANG oziroma ZGO (razlike glede na uporabljeno metodo oziroma rotacijo).

Predmeta MAT in FIZ sta bolj naravoslovne narave, predmeti SJK, ANG in ZGO pa bolj družboslovne narave. **Glede na to lahko prvi faktor označimo kot nadarjenost oziroma uspešnost pri naravoslovnih vsebinah, drugi faktor pa nadarjenost za družboslovne vsebine**.