Segédlet az **SPSS** használatához

Illetve képletek magyarázatai azokhoz a feladatokhoz melyeket nem lehet megoldani SPSS segítségével.

Az egy mintás z-próbát akkor alkalmazzuk, amikor:

- adott a szórás
- várt értékre írjuk fel a hipotézist
- egy mintánk van

$$z = \frac{\bar{x} - m_0}{\sigma} \times \sqrt{n}$$

Ahol $\bar{x}$ a minta átlaga, m_0 a várt érték, σ a szórás n pedig az elemek száma.

Miután megkaptuk a *z*-t azután a táblázatot használva kikeressük a feladatban megadott konfidencia intervallumhoz tartozó értéket és kiszámítjuk az alsó és felső határt.

$$z_{1-\frac{\propto}{2}} < z < z_{1+\frac{\propto}{2}}$$

Ha *z* a kapott értékek közé esik akkor elfogadjuk a nullhipotézist, ellenkező esetben elutasítjuk.

Az egy mintás *t*-próbát akkor alkalmazzuk, amikor:

- a szórás nem ismert
- egy mintánk van
- várt értékre írjuk fel a hipotézist

Ezt SPSS-ben a következő módon tudjuk kivitelezni:

1. Analyze -> Compare Means -> One Sample T-Test…
2. A Test Variable(s)-hez hozzáadjuk a tesztelni kívánt oszlopot
3. A Test Value-hoz beírjuk a várt értéket
4. Az Options… fülön beállítjuk a kívánt szintet (az esetek többségében ez 95%)
5. Majd Ok-kal lefuttatjuk a T-Test-et

One-Sample Test

	Test Value = 60					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
ertekek	3.202	9	.011	2.5400	.745	4.335

Mivel a Sig. (2-tailed) kisebb mint 0.05 (α) (amit úgy kapunk meg, hogy a 100%-ot 1-nek vesszük majd kivonjuk ebből a (95%*1) -et), ezért 95%-os szinten elutasítjuk a nullhipotézist.

Az χ^2-próbát **szórásra** akkor alkalmazzuk, amikor

- egy mintánk van
- a szórás nem ismert
- szórásra írjuk fel a hipotézist (azaz a szórását keressük)

$$\chi^2 = \frac{(n-1){s^*}^2}{\sigma^2}$$

Ahol s^{*2}a becsült szórás, σ^2 a tesztben várt szórás négyzete n pedig az elemek száma.

$${s^*}^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

Ahol n az elemek száma, a $\left(\sum x_i^2 - n\bar{x}^2\right)$ részben pedig az összes elemből kivonjuk az |elemszámot n szorozva, a minta átlagának a négyzetével $\bar{x}^2$| kapott értéket, majd összeadjuk azokat.

Ezután a táblázat alapján megállapítjuk a konfidencia intervallumot. Amely részeként a χ^2 táblázatból kikeressük a(z) $(n-1)$ szabadsági fokkal rendelkező sorban az adott szinten (ez általában 95% vagy 90%).

$$\chi^2 \leq \chi^2_{df}$$

Ahol, $df = (n-1)$ a szabadságfok.

Ebben az esetben csak egyoldalú lehet a nullhipotézis!

Az χ^2-próbát **illeszkedésvizsgálatra** akkor alkalmazzuk, amikor

- egy mintánk van
- meg vannak adva ilyen adatok mint $P_{H_0}(A_0) = valamilyen\ 0.xxx\ érték$
- illetve meg van nekünk adva nekünk a $P_{H_0}(A_0)$ adatokkal egyenlő számú minta adat

Ekkor a megadott n számú értékeken végig haladva, és a vele páros $P_{H_0}(A_n)$ adatot használva egyenként kiszámítjuk az alábbi egyenletet majd az eredményeket összeadjuk, és ez lesz a χ^2-t.

$$\frac{(50 - (0.135 \times 100))^2}{(0.135 \times 100)}$$

Példaként legyen az első adat, azaz a 0. elem értéke 50, és $P_{H_0}(A_0)$ legyen 0.135. Kiszámítjuk a fenti egyenletet n = 0 esetén majd folytatjuk tovább, és a végeredményeket összeadva megkapjuk a χ^2-t.

Ezután az adott szint mellett, megkeressük a χ^2 táblázatban a (n - 1) szabadságfok mellett.

Erre a feladatra van egy példafeladat Pecsora Sándor *Összefoglalás* fóliájában. Az jobban megmutatja.

Az χ^2-próbát **függetlenségvizsgálatra**akkor alkalmazzuk, amikor

- azt vizsgáljuk, hogy egy sokaság, két ismérve független-e egymástól

Ebben az esetben táblázatot készítünk:

	Ismérv csoport A1	Ismérv csoport A2	
Ismérv csoport B1	A1B1	A2B1	$\sum_{i=1}^{n} AiB1$
Ismérv csoport B2	A1B2	A2B2	$\sum_{i=1}^{n} AiB2$
	$\sum_{i=1}^{n} A1Bi$	$\sum_{i=1}^{n} A2Bi$	$\sum_{i=1}^{n}\sum_{j=1}^{n} AiBj$

És minden egyes cellán végig megyünk es elvégezzük a következő műveletet , majd az eredményeket összeadjuk.

$$\frac{\left(A1B1 - \frac{(\sum_{i=1}^{n} AiB1 \times \sum_{i=1}^{n} A1Bi)}{\sum_{i=1}^{n}\sum_{j=1}^{n} AiBj}\right)^2}{\frac{(\sum_{i=1}^{n} AiB1 \times \sum_{i=1}^{n} A1Bi)}{\sum_{i=1}^{n}\sum_{j=1}^{n} AiBj}}$$

Ez így eléggé bonyolultnak látszik, ezt le egyszerűsítendő minden egyes cellán végig haladunk majd először a cellában lévő értékből (ami a képletben $A1B1$) kivonjuk a $\frac{a\ sorban\ található\ számok\ összege \times az\ oszlopban\ található\ számok\ összege}{az\ táblázatban\ található\ számok\ össze}$ egyenlet eredményét majd azt négyzetre emeljük és osztjuk a $\frac{a\ sorban\ található\ számok\ összege \times az\ oszlopban\ található\ számok\ összege}{az\ táblázatban\ található\ számok\ össze}$ egyenlet eredményével.

Az χ^2-próbát **homogenitásvizsgálatra**akkor alkalmazzuk, amikor

- amikor két eloszlás egyezőségére kérdeznek rá a feladatban
- két minta van, több ismérvvel

Ezt is segédtáblázattal érdemes megoldani.

$$\chi^2 = n_x n_y \sum_{i=1}^{k} \frac{1}{n_{x_i} + n_{y_i}} \left(\frac{n_{x_i}}{n_x} - \frac{n_{y_i}}{n_y} \right)^2$$

Ahol n_x az egyik mintában szereplő elemek darabszáma, n_y az másik mintában szereplő elemek darabszáma.

A ∑ szimbólumban szereplő adatok közül k az ismérvek száma, n_{x_i}az x-el jelült minta i. ismérve, ez ugyanúgy megy y-al jelölt esetében is.

A ∑ szimbólumban szereplő számítások lényege az, hogy ismérvenként haladunk előre, de ennek le egyszerűsítésére használhatunk táblázatot.

Ismérv	2020	2000	Összesen	$\frac{1}{n_{x_i} + n_{y_i}} \left(\frac{\boldsymbol{n_{x_i}}}{\boldsymbol{n_x}} - \frac{\boldsymbol{n_{y_i}}}{\boldsymbol{n_y}} \right)^2$
1	20	450	470	$9.323 \cdot 10^{-4}$
2	150	150	300	$5.33 \cdot 10^{-6}$
3	167	50	217	$8.3986 \cdot 10^{-5}$
4+	456	5	461	$6.\ 9737\ \cdot 10^{-4}$
	793	655	1448	0.001719

$$\mathrm{X}^2 = 793 \cdot 655 \cdot 0.001719 = 892.8743$$

Majd kiszámítjuk a megadott szinten a konfidencia intervallumot úgy, hogy ennél a feladattípusnál a szabadságfok $df = (k-1)$.

A két mintás z-próbát akkor alkalmazzuk, amikor

- két mintánk van
- a minták függetlenek egymástól
- a minták szórása ismert
- általában összehasonlítás (magasabb-e az átlaguk)

$H_0: \mu_Y - \mu_X = 0 \; vagy \; H_0: \mu_Y = \mu_X$ a nullhipotézis az, hogy a két minta nem különbözik

$H_1: \mu_Y - \mu_X < 0 \; vagy \; H_1: \mu_Y < \mu_X$ az ellenhipotézis az, hogy a két minta különbözik

$$z = \frac{\bar{y} - \bar{x} - \delta_0}{\sqrt{\frac{\sigma_Y^2}{n_Y} + \frac{\sigma_X^2}{n_X}}}$$

Ahol $\bar{y}$ az y-al jelölt minta átlaga, δ_0 a várt különbség , σ_Y^2 y-al jelölt minta szórásnégyzete, n_Y pedig az y-al jelölt minta darabszáma.

Miután megkaptuk a z értékét, azután kiszámoljuk a kritikus tartományt:

1. Kikeressük a feladatban megadott szinten a z értékét
2. $z \geq z_\alpha$ akkor elfogadjuk a nullhipotézist, azaz nincs különbség a két minta között
3. $z < z_\alpha$ akkor elutasítjuk a nullhipotézist, azaz van különbség a két vizsgált minta között

A két mintás t-próbát (szórások megegyeznek) akkor alkalmazzuk, amikor

- a szórások megegyeznek, de nem ismertek
- két mintánk van
- a minták függetlenek egymástól

$$H_0: \mu_X - \mu_Y = 0 \; vagy \; H_0: \mu_X = \mu_Y \qquad H_1: \mu_X - \mu_Y < 0 \; vagy \; H_0: \mu_X < \mu_Y$$

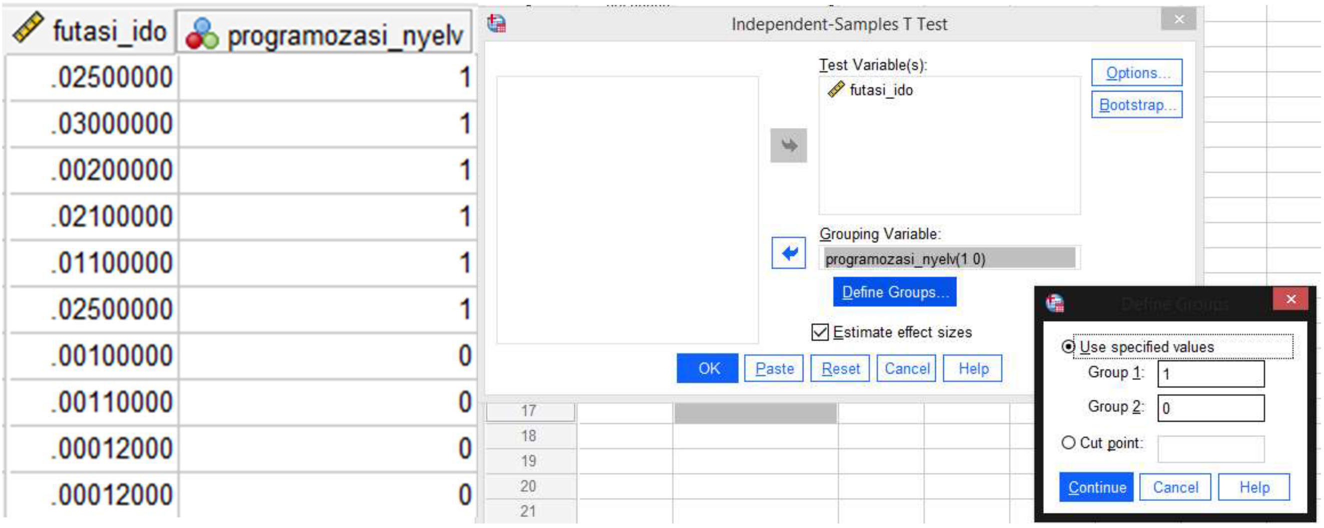
A két mintás t-próbát az SPSS-ben a következő módon valósítjuk meg:

1. Két oszlopba bevisszük az adatokat úgy, hogy az egyik oszlopban az értékek vannak, a másik oszlopban pedig az 1-essel és 0-ssal eljelölt csoportok.

2. Analyze -> Compare Means -> Independent Samples T Test…

3. Ezután a Test Variable(s) hez hozzáadjuk az értékeket a Grouping Variable – hez pedig hozzáadjuk a csoportosító változót, esetünkben az 1-essel és 0-ssal feltöltött oszlopot, majd definiáljuk a csoportokat, ahogyan az majd az ábrán látható lesz.

4. Ez után kikeressük a táblázatból a Sig(2-tailed) értéket és ha az nagyobb vagy egyenlő a döntési szinttel, akkor elfogadjuk ellenkező esetben elutasítjuk a nullhipotézist.



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
futasi_ido	Equal variances assumed	8.915	.017	3.438	8	.009	.0184150000	.0053564112	.0060630937	.0307669063
	Equal variances not assumed			4.292	5.039	.008	.0184150000	.0042902010	.0074126279	.0294173721

Mint ahogyan az a táblázatból is látszik, az első sorra van szükségünk mivel az állítja fel arra a nullhipotézist, hogy a szórások megegyeznek.

A két mintás t-próbát (az ismeretlen szórások különböznek) akkor alkalmazzuk, amikor

- a szórások különböznek, de nem ismertek
- két mintánk van
- a minták függetlenek egymástól

$$H_0: \mu_X - \mu_Y = 0 \; vagy \; H_0: \mu_X = \mu_Y \qquad H_1: \mu_X - \mu_Y < 0 \; vagy \; H_0: \mu_X < \mu_Y$$

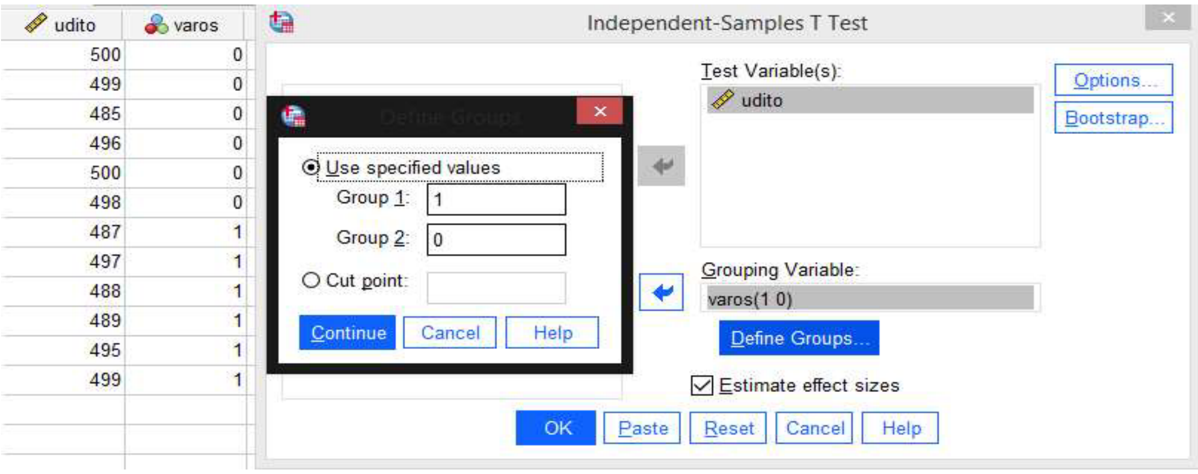
A két mintás t-próbát az SPSS-ben a következő módon valósítjuk meg:

1. Két oszlopba bevisszük az adatokat úgy, hogy az egyik oszlopban az értékek vannak, a másik oszlopban pedig az 1-essel és 0-ssal eljelölt csoportok.

2. Analyze -> Compare Means -> Independent Samples T Test...

3. Ezután a Test Variable(s) hez hozzáadjuk az értékeket a Grouping Variable – hez pedig hozzáadjuk a csoportosító változót, esetünkben az 1-essel és 0-ssal feltöltött oszlopot, majd definiáljuk a csoportokat, ahogyan az majd az ábrán látható lesz.

4. Ez után kikeressük a táblázatból a Sig(2-tailed) értéket és ha az nagyobb vagy egyenlő a döntési szinttel, akkor elfogadjuk ellenkező esetben elutasítjuk a nullhipotézist.



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
udito	Equal variances assumed	.132	.723	-1.219	10	.251	-3.833	3.146	-10.842	3.175
	Equal variances not assumed			-1.219	9.872	.251	-3.833	3.146	-10.854	3.188

Ebben az esetben viszont a második sort nézzük a táblázatban, mivel a szórások nem egyeznek meg. Az ábrán látható esetben elfogadjuk a nullhipotézist, mivel a Sig. (2-tailed) nagyobb, mint a döntésiszint azaz az 5%-nál.

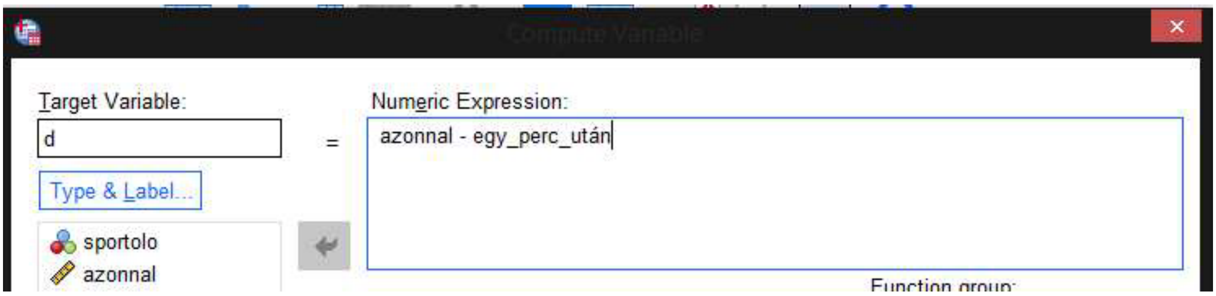
A páros mintás t-próbát (az ismeretlen szórások különböznek) akkor alkalmazzuk, amikor

- két mintánk van
- a minták függenek egymástól
- a szórás nem ismert

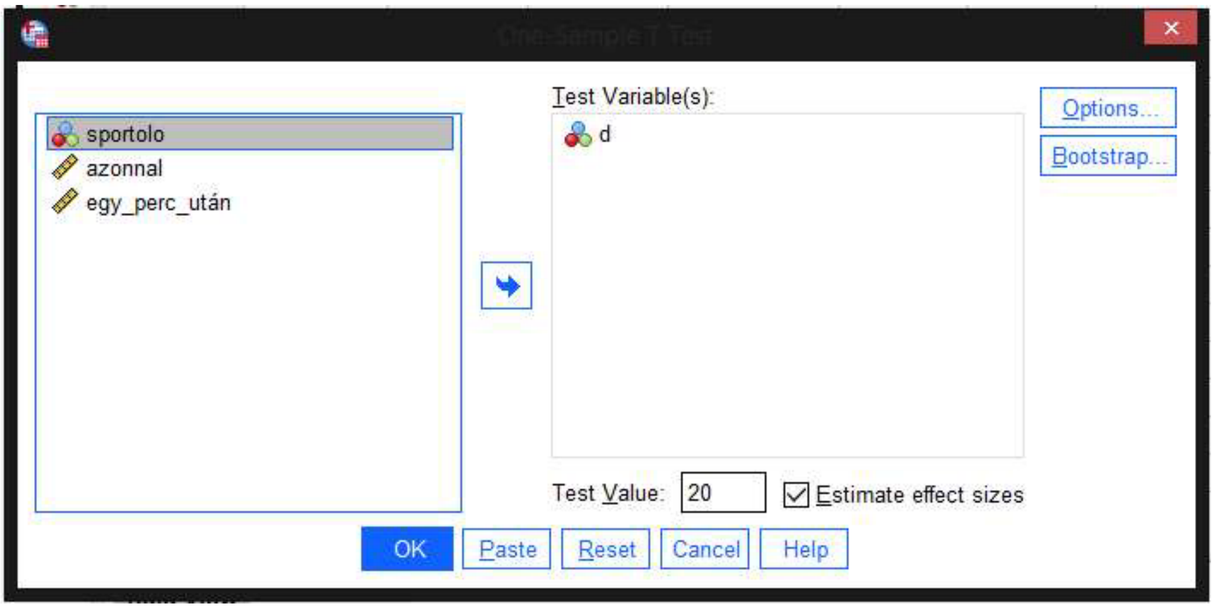
$$H_0: \mu_X - \mu_Y = 0 \; vagy \; H_0: \mu_X = \mu_Y \qquad H_1: \mu_X - \mu_Y < 0 \; vagy \; H_0: \mu_X < \mu_Y$$

A két mintás t-próbát az SPSS-ben a következő módon valósítjuk meg:

1. Két oszlopba bevisszük az adatokat. Majd kiszámoljuk az adatok különbségét egy külön oszlopba: Transform -> Compute Variable...



2. Analyze -> Compare Means -> One Sample T Test...
3. És lényegében az új változó használatával, egy egymintás t-próbát végzünk el a feladatban megadott értékre. (a példa esetében ez 20 lesz)



4. Ez után kikeressük a táblázatból a Sig(2-tailed) értéket és ha az nagyobb vagy egyenlő a döntési szinttel, akkor elfogadjuk ellenkező esetben elutasítjuk a nullhipotézist.

One-Sample Test

	Test Value = 20					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
d	-.067	11	.948	-.33333	-11.2825	10.6159

Mivel példánkban a Sig. (2-tailed) nagyobb, mint a szint, ezért a nullhipotézist elfogadjuk.

Az F-próbát a szórások egyenlőségére akkor alkalmazzuk, amikor

- a minták egymástól függetlenek
- a kérdés a szórások egyenlőségére vonatkozik
- adott egy mintára mért szórás
- adott egy-egy darabszám a két mintához

$$F^* = \frac{s_Y^2}{s_X^2}$$

Ahol s_Y^2 az y-al jelölt minta esetében megadott szórás négyzete, s_X^2 pedig az x-el jelölt minta esetében megadott szórás négyzete.

Miután megkaptuk az F-teszt eredményét, azután kikeressük a táblázatból az F($n_y - 1, n_x - 1$) tartozó értéket.

Ha $F^* \geq F(n_y - 1, n_x - 1)$ akkor elfogadjuk a nullhipotézist, azaz a két minta szórása megegyezik az adott szinten. Ellenkező esetben elutasítjuk a nullhipotézist azaz a két minta szórása nem egyezik meg.

Sokasági arányra vonatkozó kétmintás próbát akkor alkalmazzuk, amikor

- adott egy minta darabszám
- a minták azonos eloszlásúak
- sokasági arányok adottak (például: a választók 35%-a)

Ha az a kérdés nőtt-e a rokonszenv akkor ε = 0 és így írjuk fel a nullhipotézist.

$H_0: P_Y = P_X$

$H_1:\ P_Y < P_X$

Ha az a kérdés, hogy valamennyi %-kal nagyobb-e? Akkor ε = %.

$H_0: P_Y - P_X = 0.1$

$H_1:\ P_Y - P_X$ > 0.1

$$z_0\ \coloneqq\ \frac{p_Y\ - p_X}{\sqrt{\bar{p}(1-\bar{p}) * \left(\frac{1}{n_Y} + \frac{1}{n_X}\right)}}$$

Ahol p_Y az y-al eljelölt mintához tartozó sokasági arány, $\bar{p}$ a két sokasági arány átlaga.

$$\bar{p} = \frac{n_y \times p_y +\ n_x \times p_x}{n_y +\ n_x}$$

$z_0 < -z_{1-\propto}$ akkor elfogadjuk a nullhipotézist

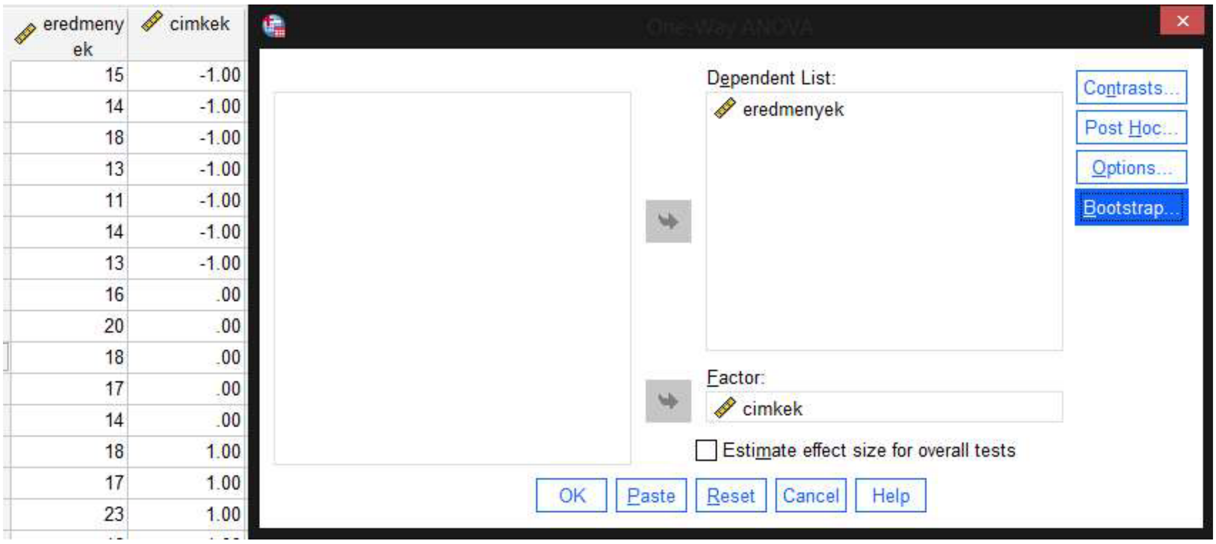
Ha $\varepsilon\ \neq 0 - al$ akkor a próbafüggvény így néz ki:

$$z_0 = \frac{p_Y\ - p_X\ - \varepsilon_0}{\sqrt{\frac{p_Y(1-p_Y)}{n_Y} + \ \frac{p_X(1-p_X)}{n_X}}}$$

$z_0 > z_{1-\propto}$ akkor elfogadjuk a nullhipotézist

Szórásfelbontó táblázat SPSS-ben

1. Bevisszük a minta adatokat egy oszlopba majd a csoportosító címkéket a másik oszlopba, ahogyan az az ábrán is látszik.
2. Analyze -> Compare Means -> One-Way ANOVA…



3. Majd az OK-ra kattintva visszakapjuk a szórásfelbontó táblázatot.

ANOVA

eredmenyek

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	SSK 134.550	M-1 2	S_K^2 67.275	9.859	.001
Within Groups	SSB 116.000	n - M 17	S_b^2 6.824		
Total	SST 250.550	n - 1 19			

A Bartlett próba próbafüggvénye:

$$B^2 \coloneqq \frac{1}{c}\left(v \ln s_b^2 - \sum_{j=1}^{M} v_j \ln s_j^2\right),$$

Ahol $c \coloneqq 1 + \frac{1}{3(M-1)}\left(\sum_{j=1}^{M}\frac{1}{v_j} - \frac{1}{v}\right)$

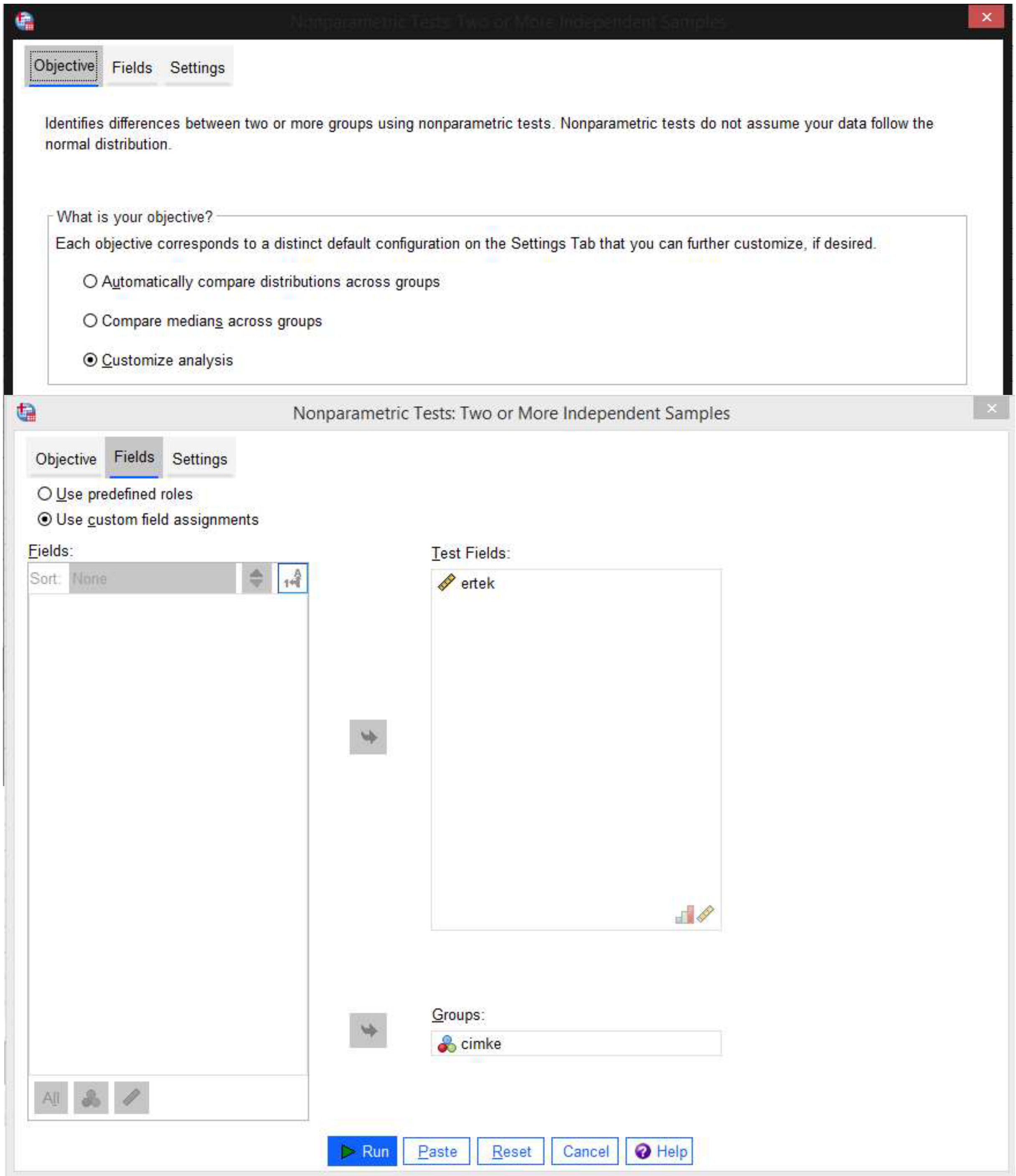
$$s_b^2 \coloneqq \frac{1}{n-M}\sum_{j=1}^{M}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2$$

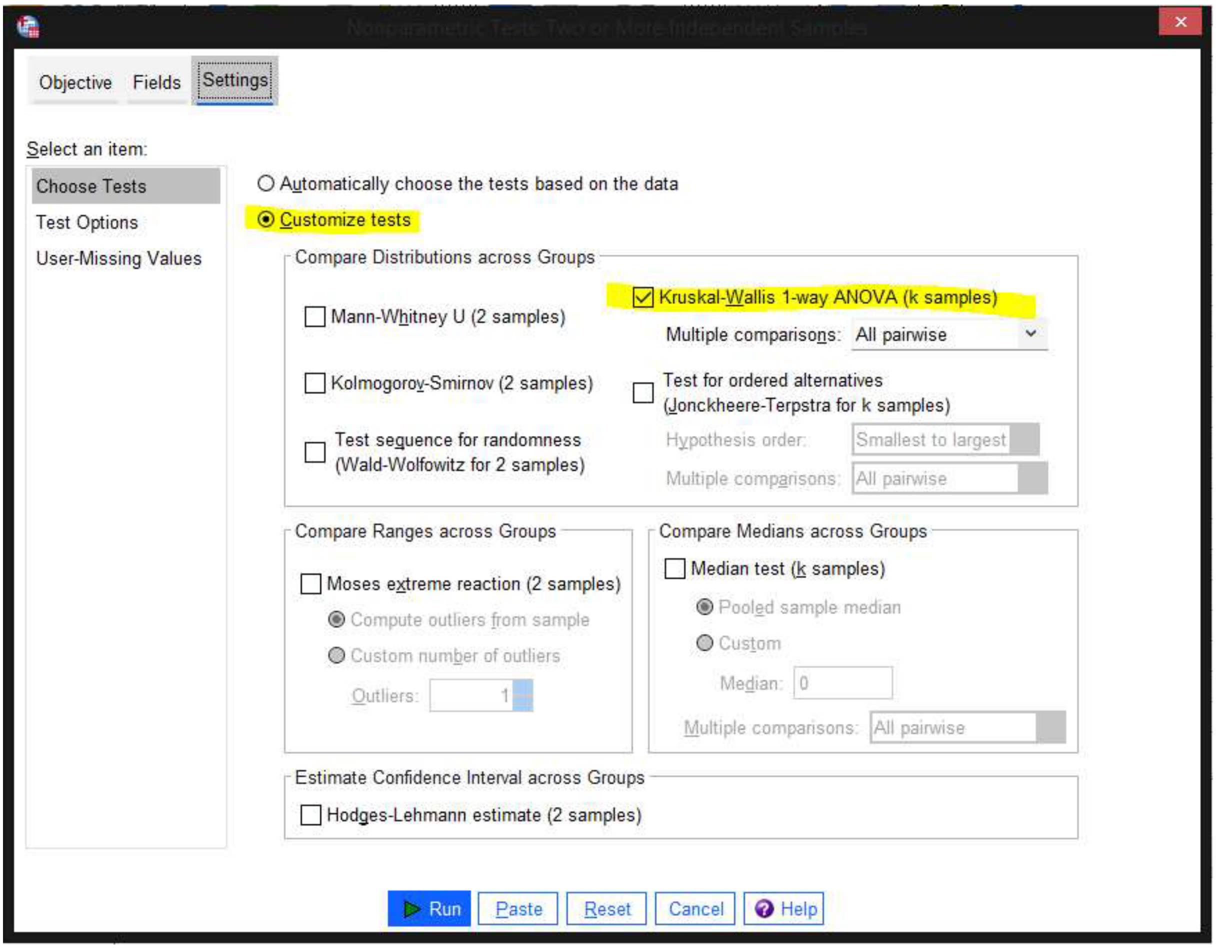
$$s_j^2 \coloneqq \frac{1}{n_j-1}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2$$

$$v \coloneqq n - M, \quad v_j = n_j - 1, \quad j = 1, \dots, M$$

Kruskal-Walis próba SPSS-ben

1. Bevisszük az adatokat az első oszlopba, majd létrehozunk egy másik oszlopot a csoportosító változóknak.
2. Analyze -> Nonparametric Tests -> Independet Samples…





3. Majd lefuttatjuk a Run gomb segítségével

Independent-Samples Kruskal-Wallis Test Summary

Total N	24
Test Statistic	7.290[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.026

a. The test statistic is adjusted for ties.

4. Mivel a példa feladatban a 95%-os szint volt megadva, így elutasítjuk a nullhipotézist hiszen az Asymptotic Sig. (2-sided test) értéke nem éri el a 0.05-ös szintet.