

VARIANCIAANALÍZIS (ANOVA)

Nevével ellentétben nem varianciák, hanem a várható értékek összehasonlítására szolgál.

Emlékeztető: kétmintás t -próba

Két független minta: $n_1, n_2; s_1^2, s_2^2; \bar{x}_1, \bar{x}_2$

Nullhipotézis: $H_0 : \mu_1 = \mu_2$

Feltételezzük a két sokaság varianciájának egyenlőségét: $\sigma_1^2 = \sigma_2^2$

Próbastatisztika:
$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \nu = n_1 + n_2 - 2$$

$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Példa: kétmintás t -próba

Két minta statisztikái:

	n	átlag	szórás
A	8	10,5	1,24
B	8	10,1	1,08

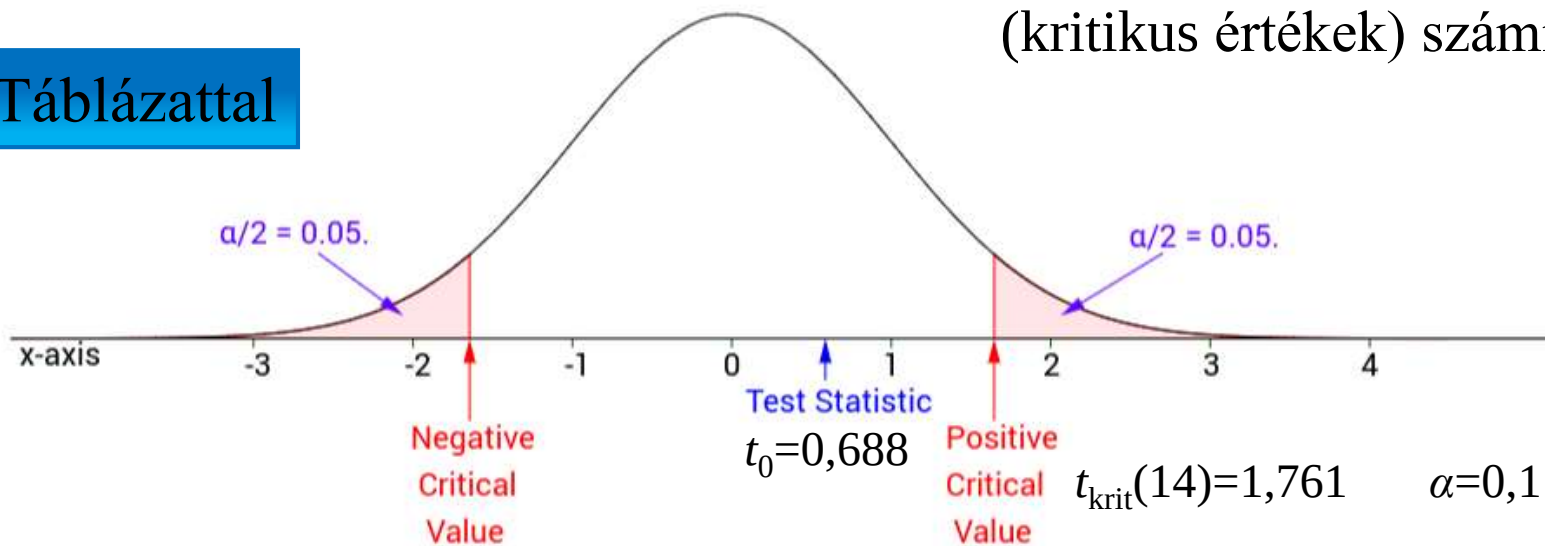
$$t_0 = \frac{\bar{x}_A - \bar{x}_B}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{10,5 - 10,1}{\sqrt{1,352 * \left(\frac{1}{8} + \frac{1}{8} \right)}} = 0,688$$

$$s^2 = \frac{s_A^2(n_A - 1) + s_B^2(n_B - 1)}{n_A + n_B - 2} = \frac{1,24^2(8 - 1) + 1,08^2(8 - 1)}{14} = 1,352$$

Döntés:

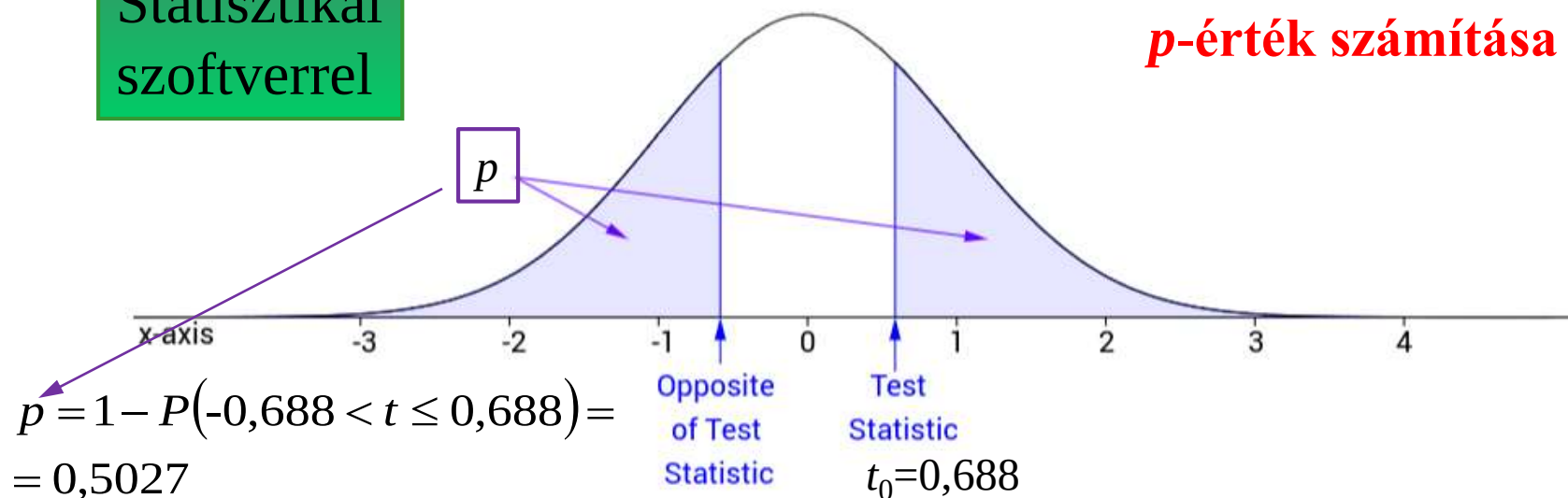
Táblázattal

Elfogadási tartomány
(kritikus értékek) számítása



Statisztikai
szoftverrel

p-érték számítása



p -érték

Definíció:

p annak valószínűsége, hogy a próbastatisztika a talált vagy annál szélsőségesebb értéket vegyen fel, ha H_0 igaz

Másként: p annak valószínűsége, hogy a mintában kapott vagy még annál is szélsőségesebb adatok adódjanak, miközben H_0 igaz.

Döntés a p -érték alapján:

ha $p > \alpha$, elfogadjuk a nullhipotézist

ha $p < \alpha$, elutasítjuk a nullhipotézist

Egy faktor szerinti ANOVA

Több független mintánk van

- elemszámuk: $p_1, p_2, p_3, \dots, p_r$
- csoportátlagok: $\bar{y}_1, \bar{y}_2, \bar{y}_3, \dots, \bar{y}_r$
- csoporton belüli szórásnégyzetek: $s_1^2, s_2^2, s_3^2, \dots, s_r^2$

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

1. példa (veralv.sta)

Véralvadási idő (sec) négyféle diéta esetén

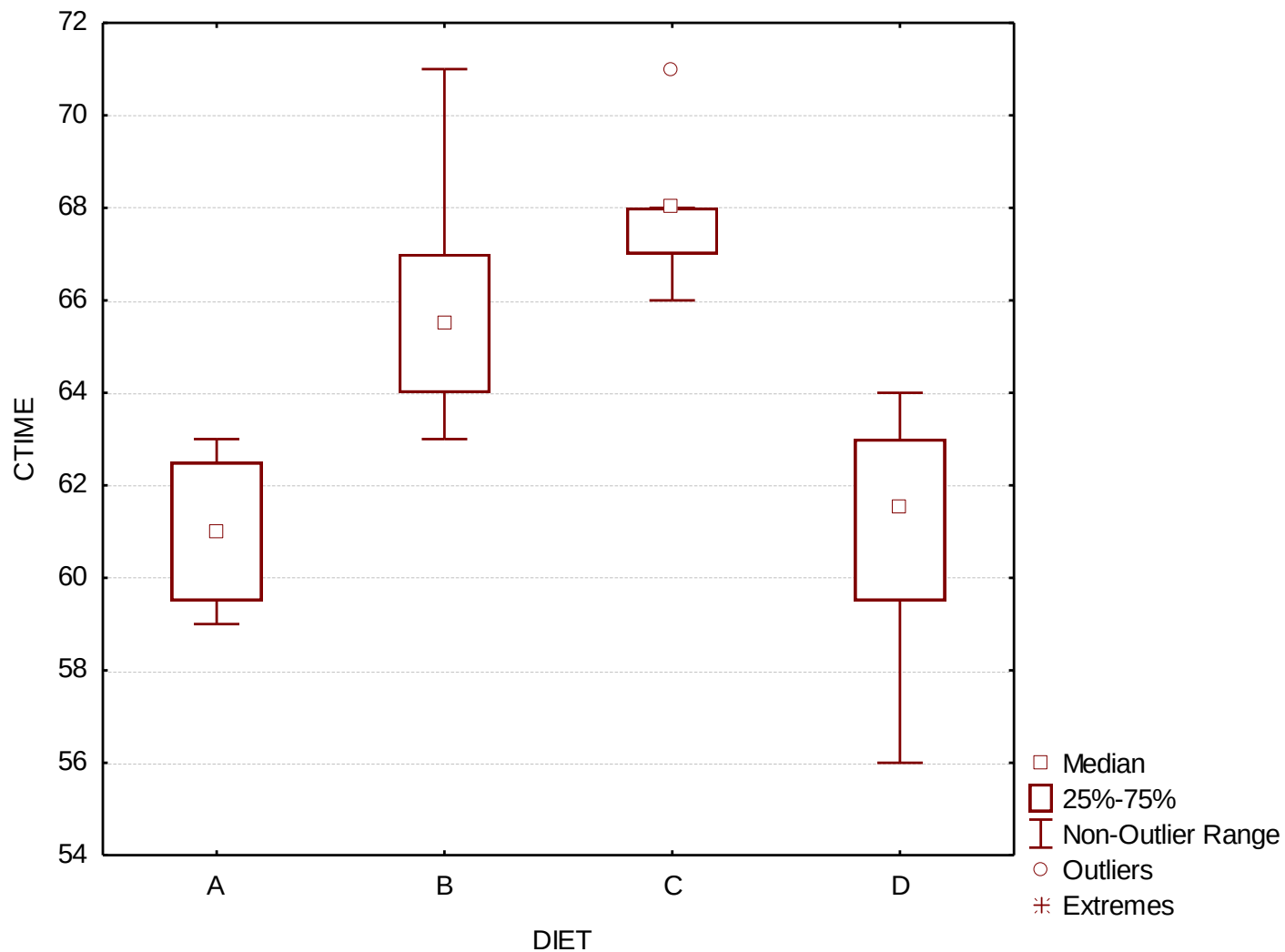
(Box-Hunter-Hunter: Statistics for Experimenters, J. Wiley, 1978, p. 165)

Diéta			
A	B	C	D
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Adatok ábrázolása: Box-plot (dobozos ábra)



Jelölések

- csoportok száma: $i = 1 \dots r$ itt $r=4$
- csoporton belüli
ismétlések száma: $j = 1 \dots p_i$ itt pl. $p_1=4$

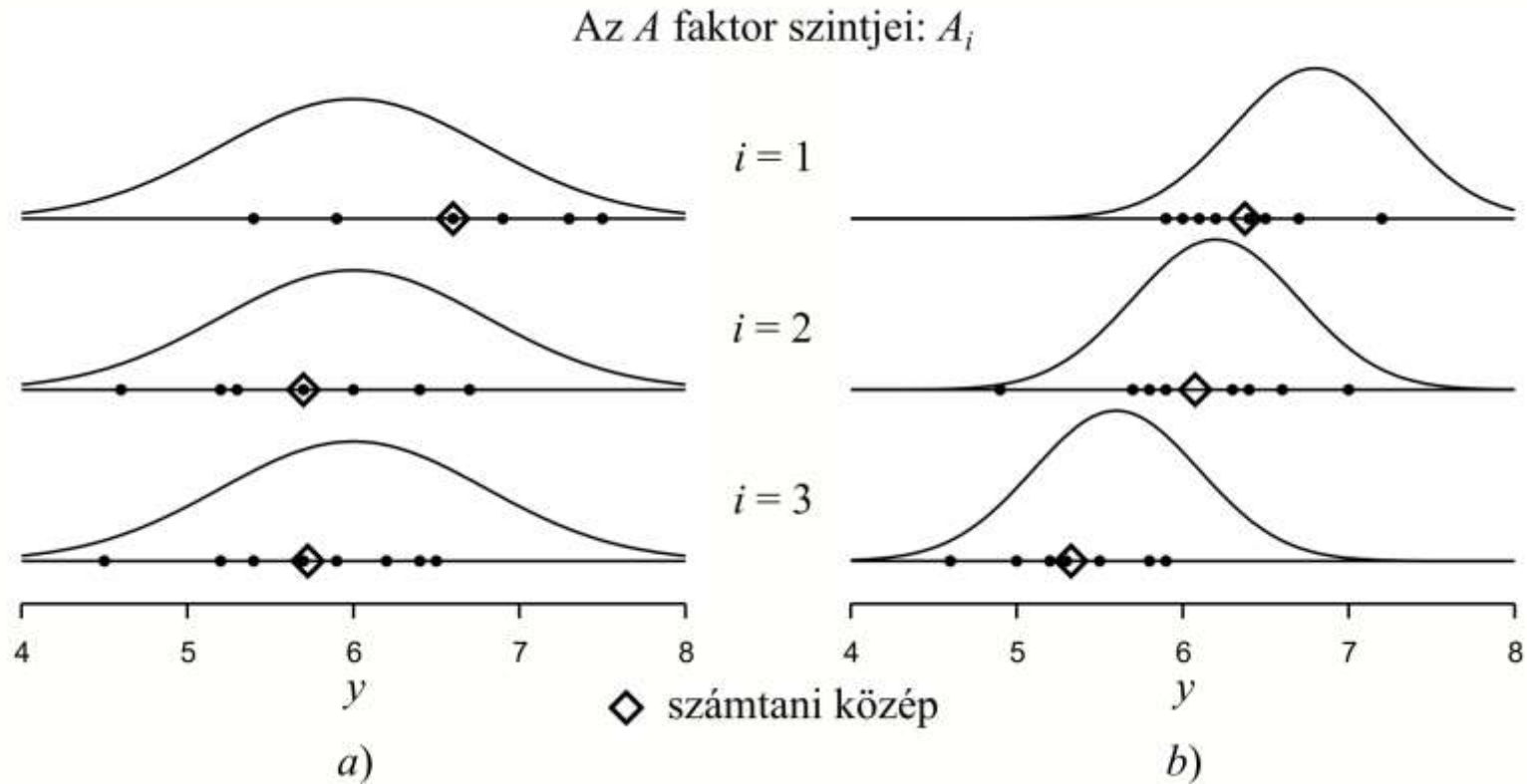
- egyedi mérési adat: y_{ij}

- csoportátlag: $y_{i.} = \frac{\sum_{j=1}^{p_i} y_{ij}}{p_i}$

- összes mérés
átlaga: $y_{..} = \frac{\sum_{i=1}^r \sum_{j=1}^{p_i} y_{ij}}{\sum_{i=1}^r p_i}$

Diéta			
A	B	C	D
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59

Szemléltető ábra az ANOVA alapgondolatához



csoportok: $i = 1 \dots r$
ismétlések: $j = 1 \dots p_i$

Az ANOVA alapelve

Csoporton belüli ingadozás varianciája

Az ***i*-edik csoporton belüli** ingadozás
varianciájának becslése:
(a csoport-átlagtól való eltérések)

$$\hat{\sigma}_i^2 = s_i^2 = \frac{\sum_{j=1}^{p_i} (y_{ij} - y_{i.})^2}{p_i - 1}$$

Az **egyesített csoportokon belüli** szórásnégyszet (ha konstans):

$$\hat{\sigma}_y^2 = \boxed{s_R^2} = \frac{\sum_i \sum_j (y_{ij} - y_{i.})^2}{\sum_i p_i - r} = \frac{\sum_i s_i^2 (p_i - 1)}{\sum_i p_i - r}$$

csoporton belüli
(within groups),
maradék vagy
error szórásnégyszet

csoportok: $i=1\dots r$
ismétlések: $j=1\dots p_i$

Az ANOVA alapelve

Csoportok közötti ingadozás varianciája

A csoport-átlagoknak a közös átlagtól való eltéréséből becsülve (ha $p=\text{konst}$):

$$\hat{\sigma}_{\bar{y}}^2 = s_{\bar{y}}^2 = \frac{\sum_{i=1}^r (y_{i\cdot} - y_{..})^2}{r-1}$$

Emlék: Az egyedi érték ingadozásának varianciája becsülhető a csoport-átlagok varianciájából:

$$\hat{\sigma}_y^2 = p^* \sigma_{\bar{y}}^2 \longrightarrow \hat{\sigma}_y^2 = \boxed{s_A^2} = p s_{\bar{y}}^2 = \frac{p \sum_{i=1}^r (y_{i\cdot} - y_{..})^2}{r-1}$$

csoporton közötti (between groups),
vagy az A faktor szórásnégyzete

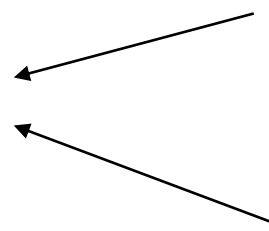
$$s_A^2 = \frac{\sum_{i=1}^r p_i (y_{i\cdot} - y_{..})^2}{r-1}$$

általánosan
(ha $p \neq \text{konst}$)

csoportok: $i = 1 \dots r$
ismétlések: $j = 1 \dots p_i$

Az ANOVA alapelve

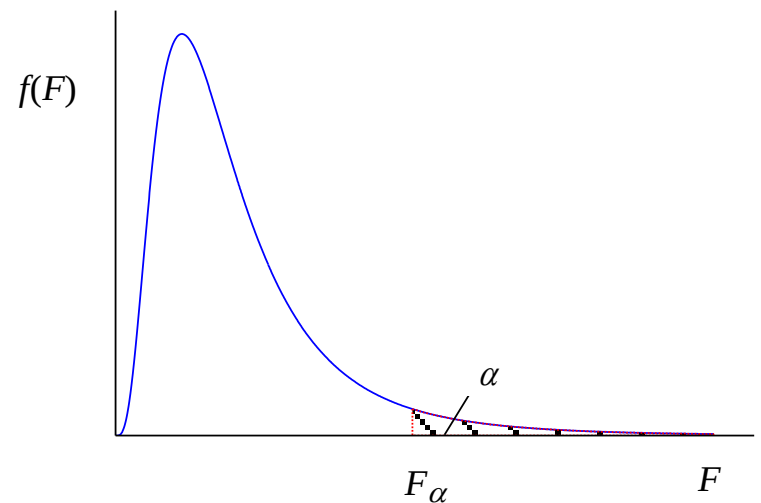
Ha nincs különbség a csoportok között (azaz H_0 igaz),
a csoport-átlagok is csak a véletlen ingadozás miatt térnek
el egymástól,
tehát s_A^2 és s_R^2 is ugyanúgy σ_y^2 körül ingadozik,
hányadosuk pedig F -eloszlású lesz

$$F_0 = \frac{s_A^2}{s_R^2}$$

$$s_A^2 = \frac{\sum_{i=1}^r p_i (y_{i.} - y_{..})^2}{r - 1} \quad \text{(csoportok közötti)}$$
$$s_R^2 = \frac{\sum_i \sum_j (y_{ij} - y_{i.})^2}{\sum_i p_i - r} \quad \text{(csoporton belüli)}$$

Az ANOVA alapelve

Ha van különbség a csoportok között (azaz H_1 igaz),
akkor s_A^2 nem σ_y^2 ingadozik, hanem sokkal nagyobb lesz s_R^2 -nél.

$$F_0 = \frac{s_A^2}{s_R^2} > F_\alpha \quad \text{egyoldali próba!}$$



ANOVA táblázat

Az eltérés forrása	eltérés-négyzetösszeg	szabadsági foksám	szórás-négyzet	F
A hatása (csoportok közötti)	$S_A = \sum_i p_i (y_{i.} - y_{..})^2$	$r-1$	$s_A^2 = \frac{S_A}{r-1}$	$\frac{s_A^2}{s_R^2}$
Ismétlések (csoportokon belüli)	$S_R = \sum_i \sum_j (y_{ij} - y_{i.})^2$	$\sum_i p_i - r$	$s_R^2 = \frac{S_R}{\sum_i p_i - r}$	
Teljes	$S_0 = \sum_i \sum_j (y_{ij} - y_{..})^2$	$\sum_i p_i - 1$		

A faktor hatása jelentős (elutasítjuk a H_0 hipotézist), ha:

$$s_A^2 / s_R^2 \geq F_{krit}$$

Kiegyensúlyozott terv: $p_1=p_2=\dots=p_r=p$

Az eltérés forrása	eltérés- -négyzetösszeg	szabadsági fok	szórásnégyzet	F
A hatása (csoportok közötti)	$S_A = p \sum_i (y_{i\cdot} - y_{..})^2$	$r-1$	$s_A^2 = \frac{S_A}{r-1}$	$\frac{s_A^2}{s_R^2}$
Ismétlések (csoportokon belüli)	$S_R = \sum_i \sum_j (y_{ij} - y_{i\cdot})^2$	$r(p-1)$	$s_R^2 = \frac{S_R}{r(p-1)}$	
Teljes	$S_0 = \sum_i \sum_j (y_{ij} - y_{..})^2$	$rp-1$		

1. példa eredménytáblázatai STATISTICA programmal / 1

Summary fülön: Descriptive cell statistics

Descriptive Statistics (Veralv)					
Effect	Level of Factor	N	CTIME Mean	CTIME Std.Dev.	CTIME Std.Err
Total		24	64.00000	3.844816	0.784820
DIET	A	4	61.00000	1.825742	0.912871
DIET	B	6	66.00000	2.828427	1.154701
DIET	C	6	68.00000	1.673320	0.683130
DIET	D	8	61.00000	2.618615	0.925820

$$\hat{\sigma}_{\bar{y}_i} = \hat{\sigma}_{y_{ij}} / \sqrt{p_i}$$

Summary fülön: Test all effects → **ANOVA tábla**

Univariate Tests of Significance for CTIME (Veralv) Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	92521.41	1	92521.41	16521.68	0.000000
DIET	228.00	3	76.00	13.57	0.000047
Error	112.00	20	5.60		

$$s_A^2 = \frac{\sum_{i=1}^r p_i (y_{i.} - y_{..})^2}{r - 1}$$

← between

within

$$s_R^2 = \frac{\sum_i s_i^2 (p_i - 1)}{\sum_i p_i - r}$$

KISTERV2_ANOVA_1

ANOVA modell

$$y_{ij} = Y_i + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}$$

kísérleti (nemcsak mérési) hiba

mért érték

igazi érték
várható érték

The diagram illustrates the ANOVA model equation $y_{ij} = Y_i + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}$. Arrows point from descriptive text to the components of the equation: 'mért érték' (measured value) points to y_{ij} ; 'igazi érték' (true value) and 'várható érték' (expected value) both point to μ_i ; and 'kísérleti (nemcsak mérési) hiba' (experimental (not only measurement) error) points to ε_{ij} . A downward arrow from 'mért érték' points towards the explanatory text at the bottom.

a faktor i -edik szintje (i -edik diéta)
az i -edik csoporton belüli j -edik ismételés

1. átlag-modell: $y_{ij} = \mu_i + \varepsilon_{ij}$

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_r$$

2. hatás-modell: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ $\mu_i = \mu + \alpha_i$ felbontással

α_i a faktor i -edik szintjének (i -edik diéta) hatása

μ közös érték; $r+1$ paraméter

$$\sum_i^r p_i \alpha_i = 0 \quad \text{sum to zero}$$

$$\alpha_r = 0 \quad \text{set to zero}$$

$$H_0 : \alpha_i = 0, \quad i = 1, \dots, r$$

A modell paramétereinek becslése

$$\phi = \sum_i^r \sum_j^{p_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \min$$

$$\hat{Y}_i = \hat{\mu} + \hat{\alpha}_i$$

$$\frac{\partial \phi}{\partial \mu} = -2 \sum_i^r \sum_j^{p_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0$$

$$\sum_i^r \sum_j^{p_i} y_{ij} = \hat{\mu} \sum_i^r p_i + \sum_i^r p_i \hat{\alpha}_i \quad \leftarrow = 0 \quad \text{sum to zero}$$

$$\hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{\sum_i p_i} = \frac{\sum_i p_i y_{i\cdot}}{\sum_i p_i} = y_{..} \quad \text{a várható érték becslése a főátlag}$$

A modell paramétereinek becslése

$$\phi = \sum_i^r \sum_j^{p_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \min$$

$$\frac{\partial \phi}{\partial \hat{\alpha}_i} = -2 \sum_j^{p_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0$$

$$p_i y_{i.} = \sum_j^{p_i} y_{ij} = \hat{\mu} p_i + p_i \hat{\alpha}_i$$

$$\hat{\alpha}_i = y_{i.} - y_{..}$$

$$\hat{Y}_i = \hat{\mu}_i = y_{i.}$$

hatások becslése a megfelelő átlagokkal
csak $r-1$ független!

az i -edik csoport várható értékének becslése
az i -edik csoport átlaga

1. példa eredménytáblázatai STATISTICA programmal / 2

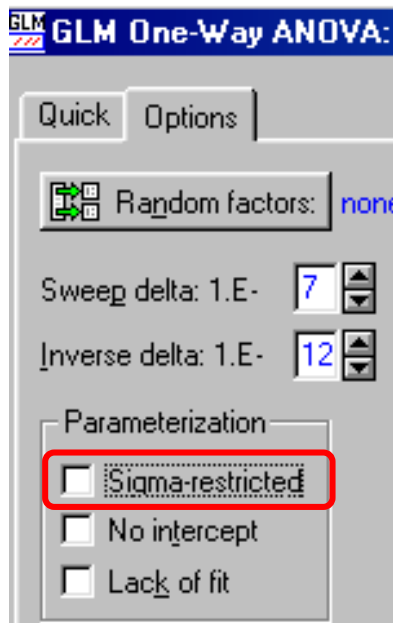
Summary fülön: *Coefficients* → a modell paramétereinek becslése

Parameter Estimates (Veralv) Sigma-restricted parameterization						
Effect	Level of Effect	Column	CTIME Param.	CTIME Std.Err	CTIME t	CTIME p
Intercept		1	64.00000	0.497912	128.5367	0.000000
DIET	A	2	-3.00000	0.973610	-3.0813	0.005889
DIET	B	3	2.00000	0.845330	2.3659	0.028195
DIET	C	4	4.00000	0.845330	4.7319	0.000128

$$\sum_i^r p_i \alpha_i = 0$$

sum to zero

sigma-restricted



Parameter Estimates (Veralv) (*Zeroed predictors failed tolerance check) Over-parameterized model					
Effect	Level of Effect	CTIME Param.	CTIME Std.Err	CTIME t	CTIME p
Intercept		61.00000	0.836660	72.90895	0.000000
DIET	A	0.00000	1.449138	0.00000	1.000000
DIET	B	5.00000	1.278019	3.91230	0.000864
DIET	C	7.00000	1.278019	5.47723	0.000023
DIET	D	0.00000			

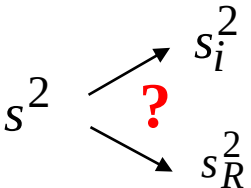
$$\alpha_4 = 0 \quad \text{set to zero}$$

Konfidencia-intervallum az egyes csoportok várható értékére

Pontbecslés: $\hat{Y}_i = \hat{\mu}_i = y_i.$

Intervallumbecslés: $t = \frac{y_{i\cdot} - \mu_i}{s_{y_{i\cdot}}}$ $s_{y_{i\cdot}}^2 = \frac{s^2}{p_i}$

szab. fok:

s^2

 s_i^2 egyes csoportokon belüli szórásnégyzet $p_i - 1$
 s_R^2 egyesített szórásnégyzet (jobb becslés) $\sum_i p_i - r$

Az i -edik csoport várható értékének konfidencia-intervalluma:

$$y_{i\cdot} - t_{\alpha/2} s_R / \sqrt{p_i} < \mu_i \leq y_{i\cdot} + t_{\alpha/2} s_R / \sqrt{p_i}$$

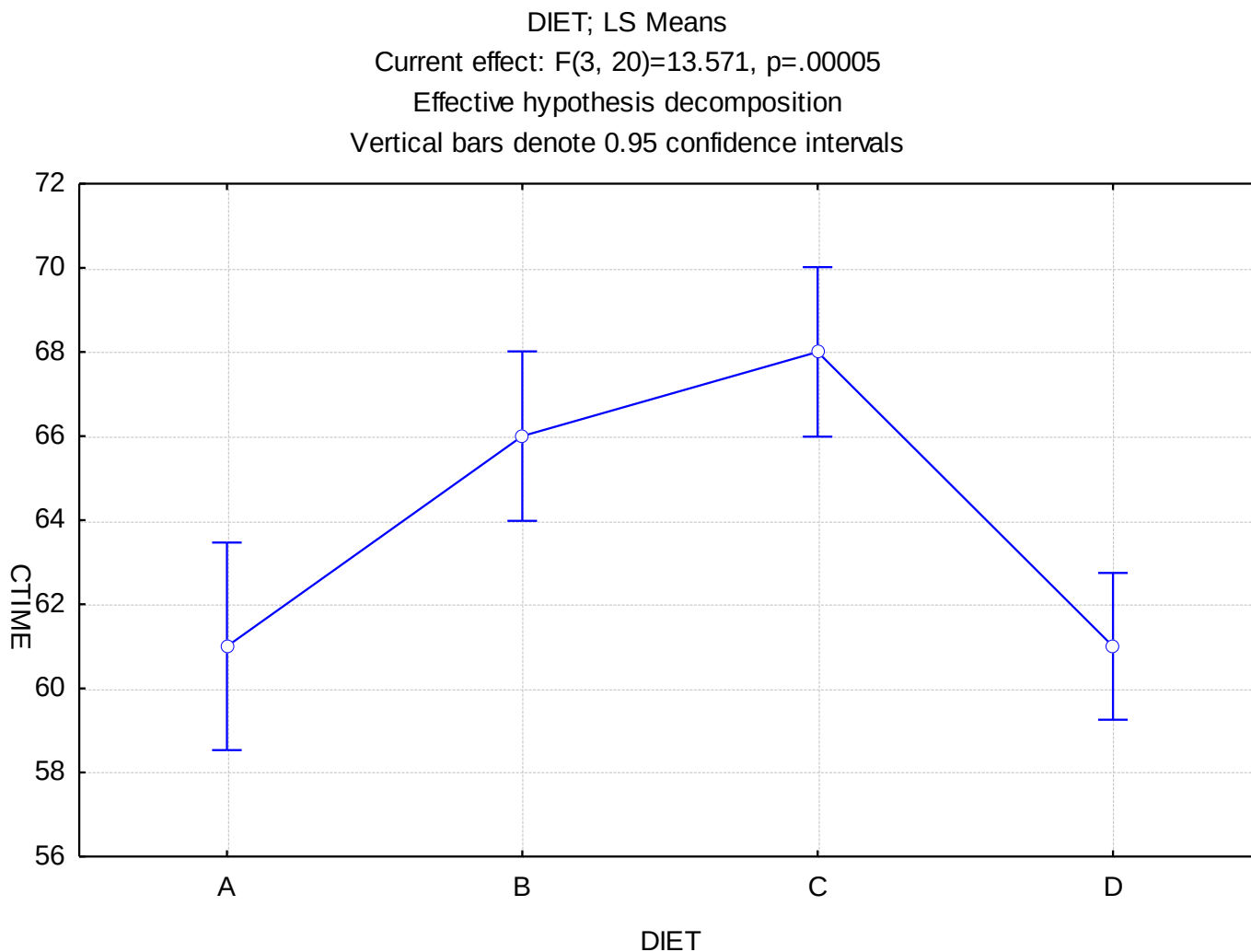
1. példa eredménytáblázatai STATISTICA programmal / 3

Summary fülön: Coefficients → **konfidencia-intervallum a modell paramétereire**

Parameter Estimates (Veralv)								
Sigma-restricted parameterization								
Effect	Level of Effect	Column	CTIME Param.	CTIME Std.Err	CTIME t	CTIME p	-95.00% Cnf.Lmt	+95.00% Cnf.Lmt
Intercept		1	64.00000	0.497912	128.5367	0.000000	62.96137	65.03863
DIET	A	2	-3.00000	0.973610	-3.0813	0.005889	-5.03092	-0.96908
DIET	B	3	2.00000	0.845330	2.3659	0.028195	0.23667	3.76333
DIET	C	4	4.00000	0.845330	4.7319	0.000128	2.23667	5.76333

Parameter Estimates (Veralv)									
(*Zeroed predictors failed tolerance check)									
Over-parameterized model									
Effect	Level of Effect	Column	Comment (B/Z/P)	CTIME Param.	CTIME Std.Err	CTIME t	CTIME p	-95.00% Cnf.Lmt	+95.00% Cnf.Lmt
Intercept		1		61.00000	0.836660	72.90895	0.000000	59.25476	62.74524
DIET	A	2	Biased	0.00000	1.449138	0.00000	1.000000	-3.02285	3.02285
DIET	B	3	Biased	5.00000	1.278019	3.91230	0.000864	2.33410	7.66590
DIET	C	4	Biased	7.00000	1.278019	5.47723	0.000023	4.33410	9.66590
DIET	D	5	Zeroed*	0.00000					

Az egyes csoportok várható értékére vonatkozó konfidencia-intervallumok ábrázolása



ANOVA feltételezései

Modell:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

- az ε_{ij} „hibák” várható értéke zérus
- σ_e^2 varianciájuk konstans (homoszkedaszticitás)
- az ε_{ij} „hibák” csoportokon belül és csoportok között is függetlenek egymástól
- az ε_{ij} „hibák” normális eloszlásúak (nem az y_{ij} adatok!)

Ellenőrizni kell!

ANOVA feltételezéseinek ellenőrzése

1. Homoszkedaszticitás: $\sigma_e^2 = \text{konst}$?

a) Statisztikai próbákkal

Assumptions fölön: Homogeneity of variances ...

Bartlett-próba

Tests of Homogeneity of Variances (Veralv) Effect: DIET					
	Hartley F-max	Cochran C	Bartlett Chi-Sqr.	df	p
CTIME	2.857143	0.381125	1.667956	3	0.644081

érzékeny a
normális eloszlás
feltételezésére

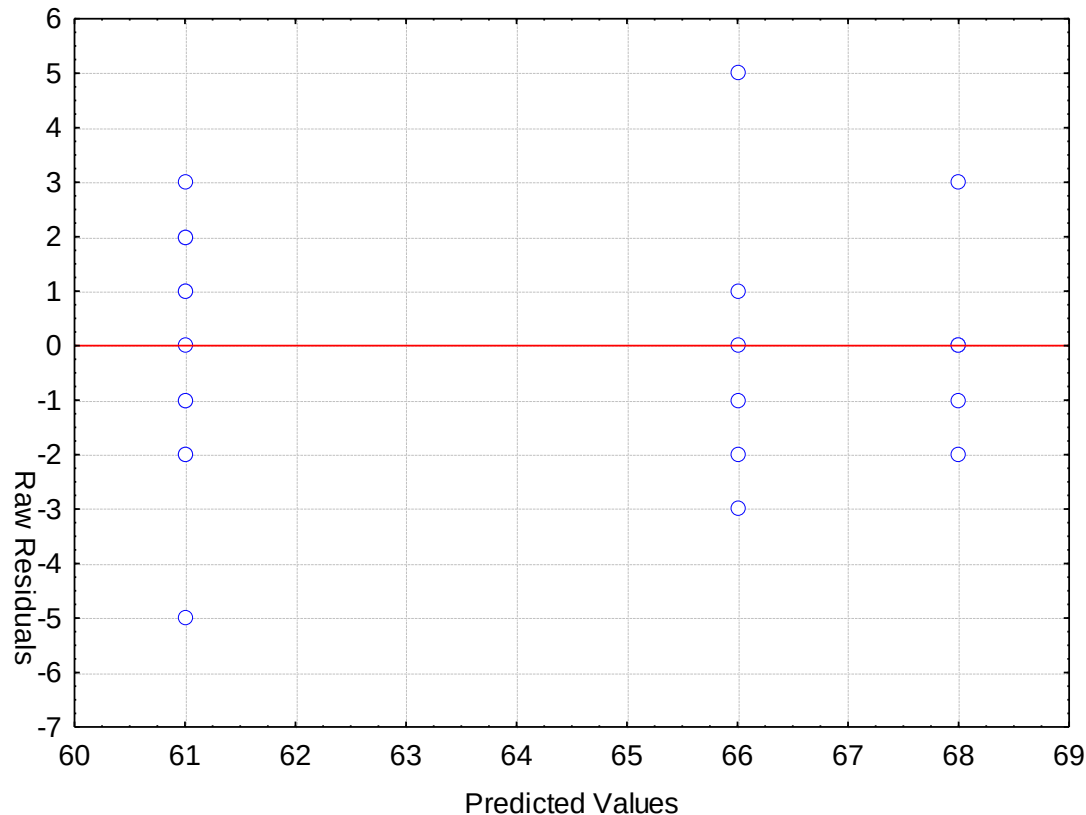
Levene-próba

Levene's Test for Homogeneity of Variances (Veralv) Effect: DIET Degrees of freedom for all F's: 3, 20				
	MS Effect	MS Error	F	p
CTIME	1.444444	2.050000	0.704607	0.560414

A feltételezések ellenőrzése a reziduumok vizsgálatával

1. Homoszkedaszticitás: $\sigma_e^2 = \text{konst}$?

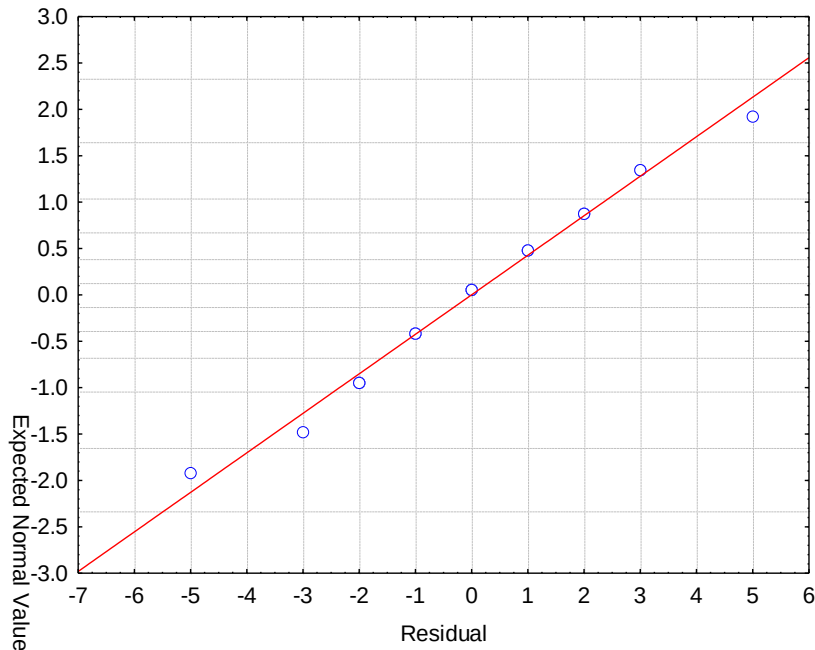
b) Becsült értékek – reziduumok (*Pred & resids*) ábrával



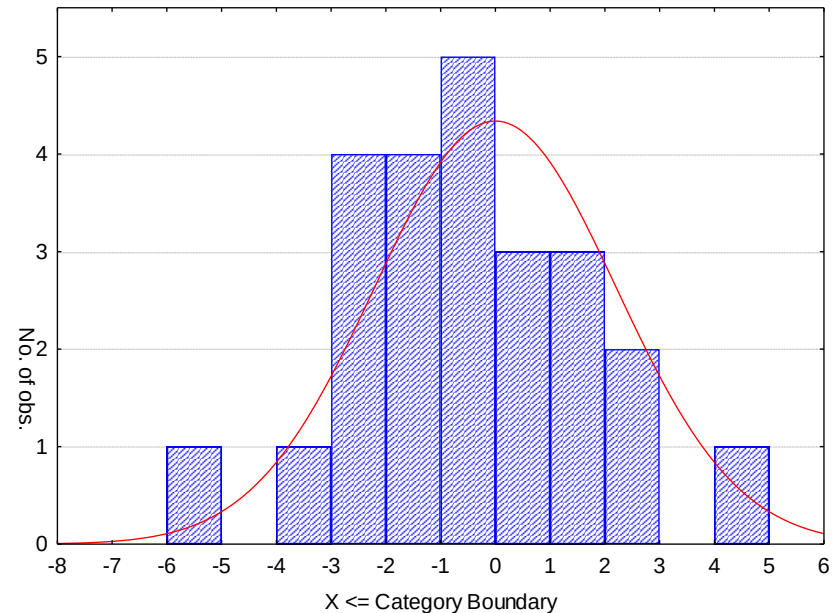
A feltételezések ellenőrzése a reziduumok vizsgálatával

2. az ε_{ij} „hibák” normális eloszlásúak-e?

Normal Probability Plot
vagy Gauss-háló



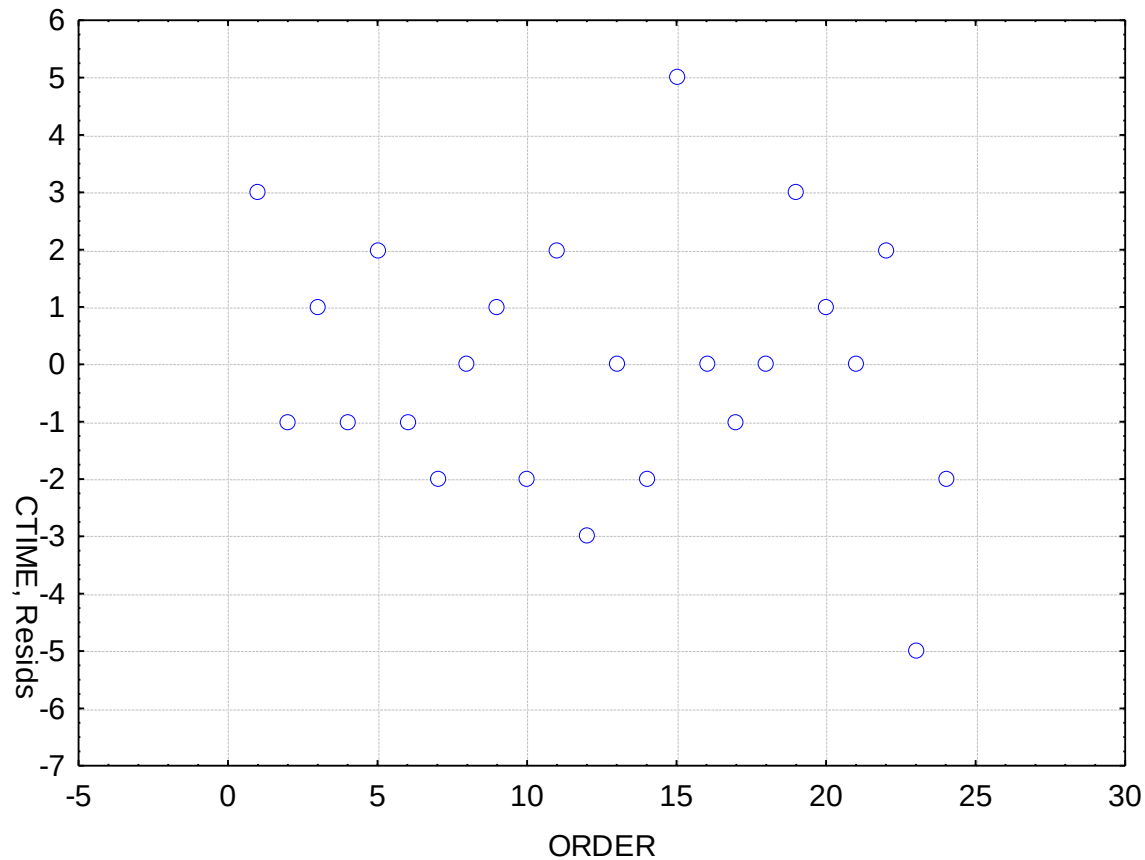
reziduumokra hisztogram



A feltételezések ellenőrzése a reziduumok vizsgálatával

3. az ε_{ij} „hibák” függetlenek-e?

Reziduumok ábrázolása pl. a mérési sorrend függvényében



Többszörös összehasonlítások

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ elutasítva → **Hogyan tovább?**

Mindegyik különböző vagy csak egy különbözik a többitől?

Példák az összehasonlításokra:

$$\mu_1 = \mu_4 \qquad \mu_2 = \mu_3 \qquad \frac{\mu_1 + \mu_4}{2} = \frac{\mu_2 + \mu_3}{2}$$

Összehasonlítások típusai: tervezett vagy post hoc

→ szemléletbeli különbség

→ veszélyei

Hasonlítsuk össze a 2. és a 3. diétát!

$$H_0 : \mu_2 = \mu_3$$

$$t_0 = \frac{y_{2\cdot} - y_{3\cdot}}{s_{y_{2\cdot} - y_{3\cdot}}}$$
$$s_{y_{2\cdot} - y_{3\cdot}}^2 = \overset{?}{s^2} \left(\frac{1}{p_2} + \frac{1}{p_3} \right)$$

s_2^2 és s_3^2 egyesítésével a szabadsági fok $n_2 + n_3 - 2 = 6 + 6 - 2 = 10$ lenne,

s_R^2 szabadsági foka $\sum_i p_i - r = 24 - 4 = 20$

$$t_0 = \frac{y_{2\cdot} - y_{3\cdot}}{s_R \sqrt{\frac{1}{p_2} + \frac{1}{p_3}}}$$

LSD-próba
(Least Significant Difference)

Kontrasztok

Az összehasonlítások általánosítása:

$$H_0^k : \sum_i c_{ik} \mu_i = 0 \quad k\text{-adik nullhipotézis}$$

$$\sum_i c_{ik} = 0 \quad c_{ik} \text{ kontraszt-együtthatók}$$

$$C_k = \sum_i c_{ik} y_i. \quad C_k \text{ kontraszt}$$

$$\text{Pl. } H_0^1 : \mu_2 - \mu_3 = 0$$

$$C_1 = y_{2\cdot} - y_{3\cdot} \quad c_{11}=0, c_{21}=1, c_{31}=-1, c_{41}=0$$

$$E(C_k) = \sum_i c_{ik} \mu_i \qquad \text{Var}(C_k) = \sigma_e^2 \sum_i \frac{c_{ik}^2}{p_i}$$

$$H_0^k : E(C_k) = \sum_i c_{ik} \mu_i = 0$$

$$t_0 = \frac{\sum_i c_{ik} y_i}{s_R \sqrt{\sum_i c_{ik}^2 / p_i}}$$

ortogonálisak a kontrasztok, ha minden $k \neq l$ -re $\sum_i c_{ik} c_{il} = 0$
 ekkor függetlenek az összehasonlítások

→ Példák függetlenek és nem független összehasonlításokra

Többszörös összehasonlítások kockázata

Ha az összehasonlítások függetlenek,
még nem jelenti azt, hogy a statisztikai próbák is függetlenek!

- egy összehasonlításra az elsőfajú hiba valószínűsége α^* (pl. 0,05)
(individual error rate):
- hogy nem követünk el elsőfajú hibát: $1 - \alpha^*$
- hogy r független összehasonlítás egyikénél
sem követünk el elsőfajú hibát: $(1 - \alpha^*)^r$
- hogy r független összehasonlítás valamelyikénél
elkövetünk elsőfajú hibát
(family error rate) $\alpha = 1 - (1 - \alpha^*)^r$

$$\text{pl. } k=6\text{-nál: } 1 - (1 - 0,05)^6 = 0,265$$

Nem független összehasonlítások esetén

$$\alpha \leq k\alpha^* \quad \text{Bonferroni-egyenlőtlenség}$$

pl. 6 nem független összehasonlításra $\alpha = 6 \cdot 0,05 = 0,3$

Többszörös összehasonlítások a Statistica programban / 1

Post-hoc fülön: LSD teszt

Az összes párra elvégzi az összehasonlítást

LSD test; variable CTIME (Veralv) Probabilities for Post Hoc Tests Error: Between MS = 5.6000, df = 20.000					
Cell No.	DIET	{1} 61.000	{2} 66.000	{3} 68.000	{4} 61.000
1	A		0.003803	0.000181	1.000000
2	B	0.003803		0.158776	0.000864
3	C	0.000181	0.158776		0.000023
4	D	1.000000	0.000864	0.000023	

Post-hoc fülön: Bonferroni

Bonferroni test; variable CTIME (Veralv) Probabilities for Post Hoc Tests Error: Between MS = 5.6000, df = 20.000					
Cell No.	DIET	{1} 61.000	{2} 66.000	{3} 68.000	{4} 61.000
1	A		0.022815	0.001083	1.000000
2	B	0.022815		0.952656	0.005182
3	C	0.001083	0.952656		0.000139
4	D	1.000000	0.005182	0.000139	

$$\alpha \leq k\alpha^*$$

$$0,003803 \cdot 6 = 0,022815$$



egy összehasonlításra vonatkozik (α^*)

Többszörös összehasonlítások a Statistica programban / 2

Planned comps fülön: Specify contrasts

Between Contrast Coefficients (Veralv) Coefficients for each cell in the selected effect			
Cell No.	DIET	Cell N	CNTRST1
1	A	4	1
2	B	6	0
3	C	6	0
4	D	8	-1

$$\mu_1 - \mu_4 = 0$$

Univariate Test of Significance for Planned Comparison (Veralv) Dependent variable: CTIME					
Source	Sum of Squares	Degr. of Freedom	Mean Square	F	p
Effect	0.0000	1	0.000000	0.000000	1.000000
Error	112.0000	20	5.600000		

Between Contrast Coefficients (Veralv) Coefficients for each cell in the selected effect			
Cell No.	DIET	Cell N	CNTRST1
1	A	4	0
2	B	6	1
3	C	6	-1
4	D	8	0

$$\mu_2 - \mu_3 = 0$$

Univariate Test of Significance for Planned Comparison (Veralv) Dependent variable: CTIME					
Source	Sum of Squares	Degr. of Freedom	Mean Square	F	p
Effect	12.0000	1	12.000000	2.142857	0.158776
Error	112.0000	20	5.600000		

Többszörös összehasonlítások a Statistica programban / 3

Between Contrast Coefficients (Veralv) Coefficients for each cell in the selected effect			
Cell No.	DIET	Cell N	CNTRST1
1	A	4	1
2	B	6	-1
3	C	6	-1
4	D	8	1

$$\frac{\mu_1 + \mu_4}{2} = \frac{\mu_2 + \mu_3}{2}$$

$$\mu_1 - \mu_2 - \mu_3 + \mu_4 = 0$$

Univariate Test of Significance for Planned Comparison (Veralv) Dependent variable: CTIME					
Source	Sum of Squares	Degr. of Freedom	Mean Square	F	p
Effect	203.2941	1	203.2941	36.30252	0.000007
Error	112.0000	20	5.6000		

Kimutatható különbség és a próba ereje

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ elfogadva $\rightarrow$ Hogyan tovább?

„Emlékek”:

Mitől függ a másodfajú hiba valószínűsége?

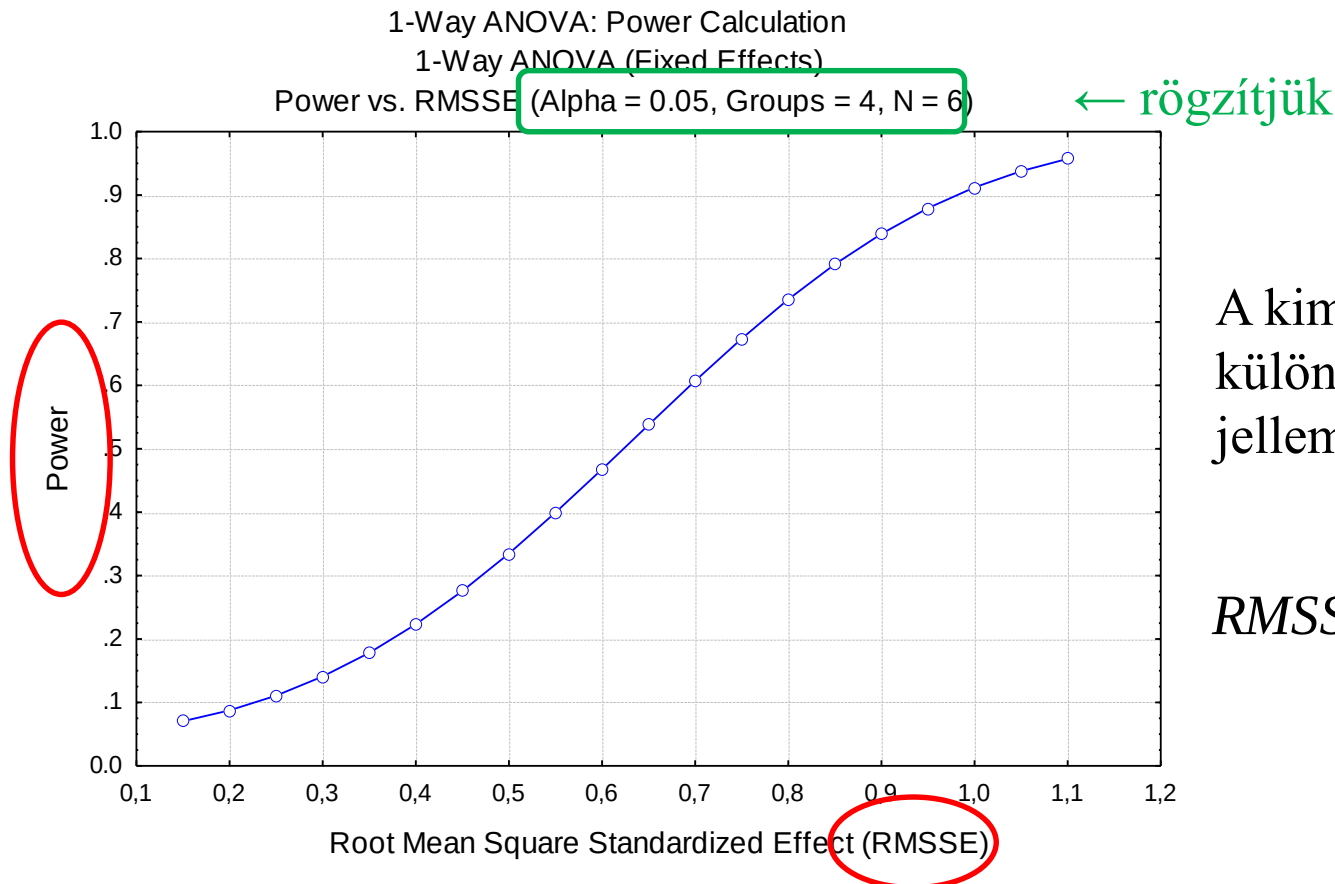
- ✓ Elsőfajú hiba valószínűségétől (α)
- ✓ Kimutatandó különbségtől (Δ)
- ✓ Varianciától (σ^2)
- ✓ Mintaelem-számtól (n)

Próba ereje:

annak valószínűsége, hogy észreveszem a különbséget

1. Milyen valószínűséggel mutatható ki a csoportok között egy adott különbség?
2. Mekkora különbség mutatható ki adott valószínűséggel?

Statistics > Power Analysis > Several Means, ANOVA 1-Way



A kimutatandó
különbséget
jellemző mérőszám:

$$RMSSE = \sqrt{\frac{\sum_i \alpha_i^2}{(r-1)\sigma_e^2}}$$

$$RMSSE = \sqrt{\frac{\sum_i \alpha_i^2}{(r-1)\sigma_e^2}}$$

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Pl. ha $\alpha_1 = -3$, $\alpha_2 = 3$, $\alpha_3 = \alpha_4 = 0$ és $\sigma_e^2 = 5,6$

$$RMSSE = \sqrt{\frac{(-3)^2 + 3^2 + 0^2 + 0^2}{(4-1) \cdot 5,6}} = \sqrt{\frac{18}{3 \cdot 5,6}} = 1,035$$

Az előző diagramról leolvasva:

ehhez a próba ereje $= 1 - \beta \approx 0,95$

De ha $\alpha_1 = -1$, $\alpha_2 = 1$, $\alpha_3 = \alpha_4 = 0$,

akkor RMSSE értéke már csak 0,345;

a próba ereje pedig már csak kb. 0,19.

Sample Size Calculation ANOVA, 1-Way Fixed Effects

Value

Number of Groups	4.0000
RMSSE	1.0350
Noncentrality Parameter (Delta)	32.1368
Type I Error Rate (Alpha)	0.0500
Power Goal	0.9000
Actual Power for Required N	0.9308
Required Sample Size (N)	6.0000

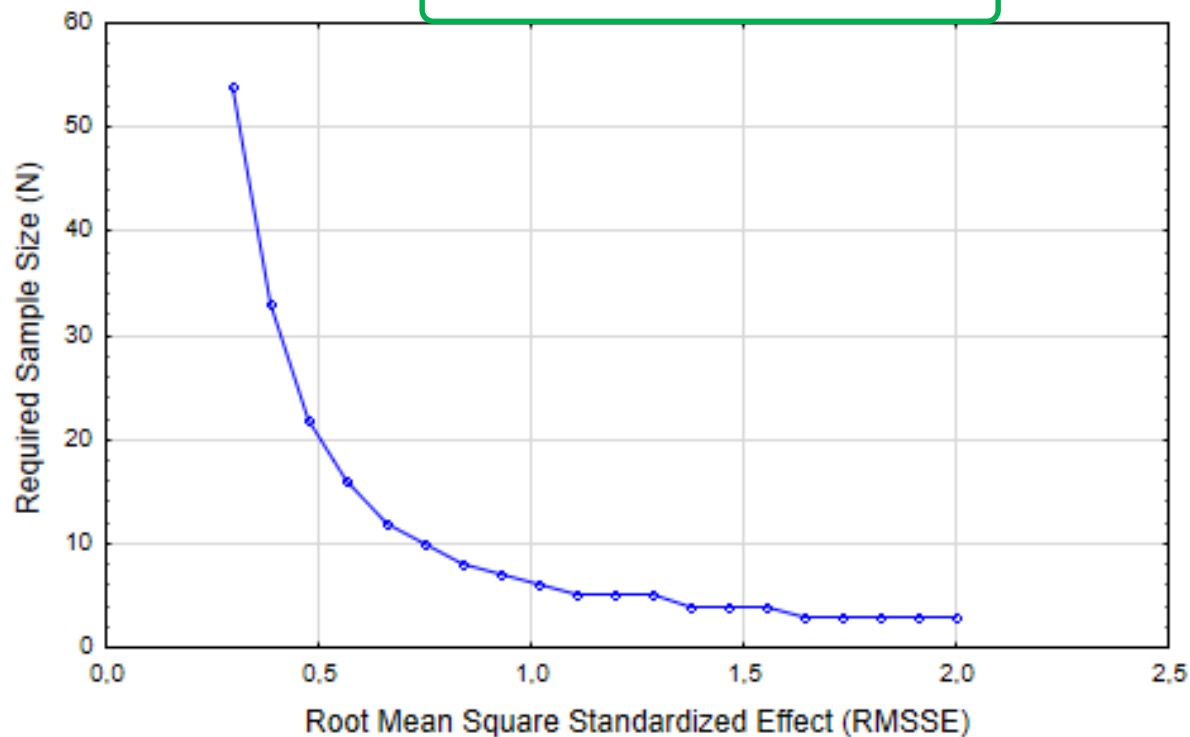
← rögzítjük

Milyen ismétlés-
számmal tudunk
adott különbséget
kimutatni?

1-Way ANOVA: Sample Size Calculation

1-Way ANOVA (Fixed Effects)

N vs. RMSSE (Alpha = 0,05, Groups = 4, Power = 0,9)



Kimutatható különbség számítása táblázattal

	df numerator								
df denom.	1	2	3	4	5	6	10	20	50
1	20.96	23.25	24.16	24.65	24.95	25.15	25.57	25.89	26.08
2	6.796	6.710	6.682	6.668	6.659	6.653	6.642	6.633	6.628
3	5.014	4.630	4.475	4.390	4.336	4.299	4.221	4.159	4.121
4	4.396	3.900	3.692	3.576	3.502	3.450	3.339	3.250	3.193
5	4.092	3.538	3.301	3.166	3.079	3.018	2.886	2.777	2.707
6	3.913	3.324	3.068	2.921	2.825	2.757	2.609	2.486	2.405
7	3.795	3.183	2.914	2.759	2.656	2.583	2.423	2.287	2.197
8	3.712	3.084	2.805	2.643	2.535	2.458	2.288	2.142	2.044
10	3.604	2.953	2.661	2.489	2.375	2.292	2.107	1.944	1.832
12	3.536	2.871	2.570	2.392	2.272	2.186	1.989	1.814	1.690
14	3.489	2.815	2.508	2.325	2.202	2.112	1.907	1.721	1.588
16	3.455	2.774	2.463	2.276	2.150	2.058	1.846	1.652	1.510
18	3.429	2.743	2.428	2.239	2.111	2.017	1.800	1.598	1.449
20	3.409	2.718	2.401	2.210	2.080	1.984	1.762	1.554	1.399
22	3.393	2.698	2.379	2.186	2.054	1.957	1.732	1.519	1.357
24	3.380	2.682	2.361	2.166	2.033	1.935	1.707	1.489	1.322
26	3.368	2.669	2.346	2.150	2.016	1.917	1.686	1.464	1.292
28	3.359	2.657	2.333	2.136	2.001	1.901	1.667	1.442	1.265
30	3.351	2.647	2.322	2.124	1.988	1.888	1.652	1.423	1.242
40	3.322	2.613	2.283	2.082	1.944	1.841	1.597	1.355	1.159
60	3.295	2.580	2.246	2.042	1.900	1.794	1.542	1.287	1.070
80	3.281	2.563	2.227	2.022	1.878	1.772	1.515	1.252	1.022
100	3.273	2.554	2.216	2.010	1.866	1.758	1.498	1.231	.9926
200	3.257	2.534	2.195	1.986	1.840	1.731	1.466	1.187	.9298
500	3.248	2.523	2.182	1.972	1.825	1.715	1.446	1.161	.8894
1000	3.245	2.519	2.178	1.967	1.820	1.709	1.439	1.152	.8754
∞	3.242	2.515	2.173	1.962	1.815	1.704	1.433	1.143	.8610

(Több) rögzített
faktorokra

$$\alpha = 0,05$$

$$\beta = 0,1$$

F-próba
számlálójának
és nevezőjének
szabadsági foka

$$\sqrt{C \frac{\Phi(\text{hatás})}{EMS(\text{nevező})}}$$

A táblázatban $\sqrt{C \frac{\Phi(\text{hatás})}{EMS(\text{nevező})}}$ értékei vannak

C = az összes kísérletek száma / a hatás szintjeinek száma
(egyfaktoros ANOVA-nál: $C = p$)

$$\Phi(A) = \frac{\sum_i^r \alpha_i^2}{r-1}$$

$EMS(\text{nevező})$ = az F -próba nevezőjében szereplő
szórásnégyzet várható értéke
(egyfaktoros ANOVA-nál: σ_e^2)