Egy faktor szerinti ANOVA

Nevével ellentétben nem szórások, hanem átlagok összehasonlítására szolgál

Több független mintánk van, elemszámuk $p_1, p_2, p_3, ..., p_r$

$$\bar{y}_1, \bar{y}_2, \bar{y}_3, ..., \bar{y}_r; \quad s_1^2, s_2^2, s_3^2, ..., s_r^2$$

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

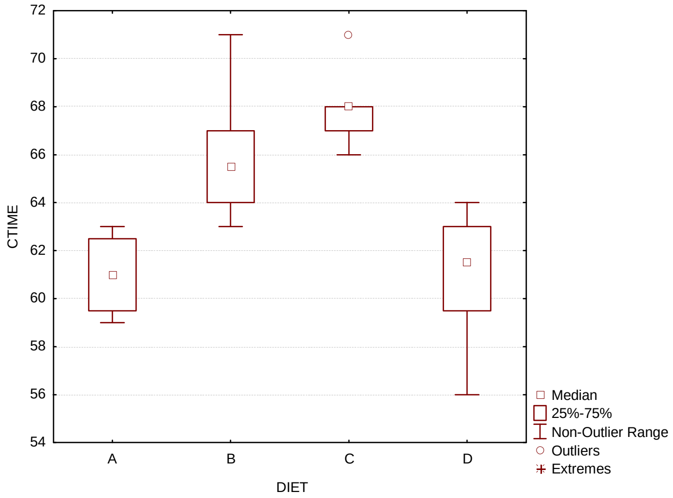
1. példa

(Box-Hunter-Hunter: Statistics for Experimenters, J. Wiley, 1978, p. 165) Véralvadási idő (sec) négyféle diéta esetén

veralv.sta

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

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



Ha nincs különbség a csoportok között, csak a véletlen ingadozás miatt térnek el egymástól az átlagok.

$\hat{\sigma}_y^2$ kétféleképpen adható meg:

- ismétlések közötti eltérés
- átlagok közötti eltérés

s_A^2 és s_R^2 is σ_y^2 körül ingadozik

Az *i*-edik csoporton belüli ingadozás varianciája

$$s_i^2 = \frac{\sum_{j=1}^{p_i} \left(y_{ij} - y_{i\cdot}\right)^2}{p_i - 1}$$

(a csoport-átlagtól való eltérések)

$$y_{i\cdot} = \frac{\sum_{j=1}^{p_i} y_{ij}}{p_i}$$

Az egyesített csoportokon belüli szórásnégyzet (ha σ konstans):

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

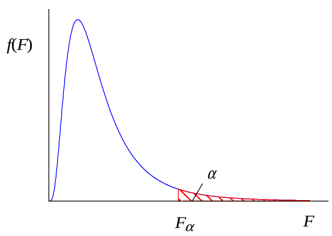
$$s_A^2 = \frac{\sum_{i=1}^{r} p_i \left(y_{i\cdot} - y_{\cdot\cdot}\right)^2}{r - 1} \quad \text{(between)}$$

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

$$s_R^2 = \frac{\sum_i \sum_j \left(y_{ij} - y_{i\cdot}\right)^2}{\sum_i p_i - r} \quad \text{(within)}$$

ha a valóságban van különbség (H_1)

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



ANOVA táblázat

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

S_0 ún. teljes négyzetösszeg. Az A faktor hatása jelentős (elutasítjuk a H_0 hipotézist), ha $s_A^2 / s_R^2 \geq F_{krit}$

Kiegyensúlyozott terv: $p_1 = p_2 = ... = p_r = p$

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

Summary fülön: Descriptive cell statistics

Effect	Level of Factor	N	CTIME Mean	CTIME Std.Dev.	CTIME Std.Err
	Descriptive Statistics (Veralv)				
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

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

Summary fülön: Test all effects

$$\mathrm{H}_0:\ \mu_1 = \mu_2 = \mu_3 = \mu_4$$

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		

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

between

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

within

Feltételezések:

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

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

Modell

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

kísérleti (nemcsak mérési) hiba

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

mért érték

igazi érték
várható érték

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

$$\mathrm{H}_0: \ \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_r$$

átlag-modell

$$\mu_i = \mu + \alpha_i$$

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

μ közös érték; r+1 paraméter

$i=1,\ldots,r$

$$\sum_i^r p_i \alpha_i = 0$$ sum to zero

$$\alpha_r = 0$$ set to zero

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

$$\mathrm{H}_0: \ \alpha_i = 0, \quad i = 1,\ldots,r$$

hatás-modell

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

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

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

$$\sum_i \sum_j y_{ij} = \hat{\mu} \sum_i p_i + \sum_i p_i \hat{\alpha}_i$$

=0

$$\hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{\sum_i p_i} = \frac{\sum_i p_i y_{i\cdot}}{\sum_i p_i} = y_{\cdot\cdot}$$

főátlag

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

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

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

hatás, csak *r*-1 független

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

az *i*-edik csoport átlaga

Konfidencia-intervallum az egyes csoportok várható értékére

Pont-becslés: $\hat{Y}_i = \hat{\mu}_i = y_{i\cdot}$

Intervallum-becslés: $t = \dfrac{y_{i\cdot} - \mu_i}{s_{y_{i\cdot}}}$

$s_{y_{i\cdot}}^2 = \dfrac{s_R^2}{p_i}$ szab. fokszáma: $\sum_i p_i - r$

Az *i*-edik csoport várható értékének konfidencia-intervalluma:

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

Summary fülön: Coefficients

sigma-restricted

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					

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

Mindegyik különböző?

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

Összehasonlítások: tervezett, post hoc

$H_0: \mu_2 = \mu_3$ $t_0 = \frac{y_{2\cdot} - y_{3\cdot}}{s_{y_{2\cdot} - y_{3\cdot}}}$

$Var(y_{2\cdot} - y_{3\cdot}) = \sigma_e^2 \left(\frac{1}{p_2} + \frac{1}{p_3} \right)$ $s^2_{y_{2\cdot} - y_{3\cdot}} = s^2 \left(\frac{1}{p_2} + \frac{1}{p_3} \right)$

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)

$$H_0: \mu_2 = \mu_3$$

Általánosítás:
(*k*-adik nullhipotézis)

$$H_0^k: \sum_i c_{ik}\mu_i = 0$$

$$H_0: \mu_2 - \mu_3 = 0$$

$$\sum_i c_{ik} = 0$$

c_{ik} kontraszt-együtthatók

$$C_k = \sum_i c_{ik} y_{i\cdot}$$

kontraszt

$$H_0^1: \mu_2 - \mu_3 = 0 \qquad C_1 = y_{2\cdot} - y_{3\cdot}$$

c_{11}=0, c_{21}=1, c_{31}=-1, c_{41}=0

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

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

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

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

$$H_0^1: \mu_2 - \mu_3 = 0$$

$$H_0^2: \mu_1 - \mu_4 = 0$$

$$H_0^3: \mu_1 - \mu_2 - \mu_3 + \mu_4 = 0$$

$$\sum_i c_{ik} c_{il} = 0 \quad ?$$

$$\longrightarrow \quad \mu_1 = \mu_2 = \mu_3 = \mu_4$$

		H_0^1	H_0^2	H_0^3			
i	μ_i	c_{i1}	c_{i2}	c_{i3}	$c_{i1}\,c_{i2}$	$c_{i1}\,c_{i3}$	$c_{i2}\,c_{i3}$
1	μ_1	0	1	1	0	0	1
2	μ_2	1	0	–1	0	–1	0
3	μ_3	–1	0	–1	0	1	0
4	μ_4	0	–1	1	0	0	–1
	Σ	0	0	0	0	0	0

$\binom{4}{2} = 6$ összehasonlítás (1-2, 1-3, 1-4, 2-3, 2-4, 3-4)

egy összehasonlításra az elsőfajú hiba valószínűsége α^*
(pl. 0.05)
(individual error rate)

hogy nem követünk el elsőfajú hibát: 1- α^*

hogy r független összehasonlítás egyikénél
sem követünk el elsőfajú hibát: $(1-\alpha^*)^r$

hogy r független összehasonlítás valamelyikénél
elkövetünk elsőfajú hibát: $\alpha = 1-(1-\alpha^*)^r$

(family error rate) pl. $1-(1-0.5)^6 = 0.265$

Nem független összehasonlítások esetén

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

pl. 6 nem független összehasonlításra $6 \cdot 0.05 = 0.3$

Post hoc összehasonlítások

Post-hoc fülön: LSD

	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	

$0.003803 \cdot 6 = 0.022815$

Tervezett összehasonlítások Planned comps fülön: Specify contrasts

$$\mu_1 - \mu_4 = 0$$

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

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		

$$\mu_2 - \mu_3 = 0$$

Cell No.	Between Contrast Coefficients (Veralv) Coefficients for each cell in the selected effect DIET	Cell N	CNTRST1
1	A	4	0
2	B	6	1
3	C	6	-1
4	D	8	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	12.00000	2.142857	0.158776
Error	112.0000	20	5.60000		

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

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

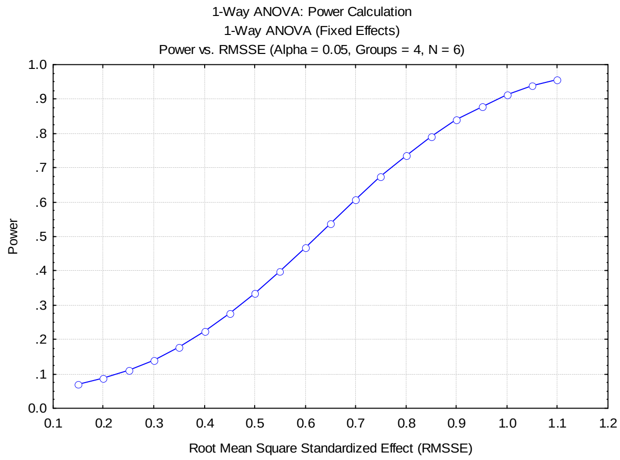
$$\mu_1 - \mu_2 - \mu_3 + \mu_4 = 0$$

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		

Mekkora különbséget tudnánk kimutatni?

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



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

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

Pl.ha α_1=-3, α_2=3, α_3= α_4=0

$$RMSSE = \sqrt{\frac{(-3)^2 + 3^2 + 0^2 + 0^2}{(4-1)\cdot 5.6}} = \sqrt{\frac{18}{3\cdot 5.6}} = 1.035$$

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

More results>Assumptions fülön: 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

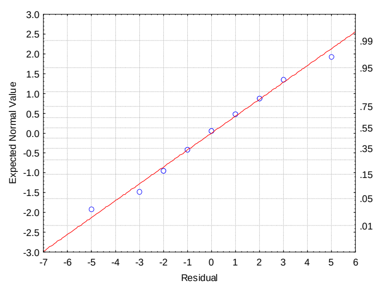
Levene-próba

	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

Residuals 1 fülön

Normality →



Pred & resids

Predicted results (histogram)