Gyakorló feladatok
a Kísérletek tervezése és értékelése c. tárgyból
Lineáris regresszió, ismétlés nélküli mérések

1. példa

Az alábbi táblázat egy kalibrációs egyenes felvételekor mért adatokat tartalmazza:

x	*y*
2	1.82
3	3.22
5	4.98
6	6.3

(a) Adjon becslést a kalibrációs egyenes paramétereire!

(b) Vizsgálja meg 5%-os szignifikanciaszinten, hogy az igazi egyenes átmegy-e az origón?

(c) Adjon 90%-os konfidencia-intervallumot az egyenes meredekségére!

(d) Kimondhatjuk-e 90% valószínűséggel, hogy az igazi egyenes az x=4 helyen Y=4 fölött halad?

(e) Milyen intervallumban lenne egy új mérés eredménye az x=4 helyen 90% valószínűséggel?

Megoldás

(a) $a = \hat{\alpha} = 4.08;\ b = \hat{\beta} = 1.072;\ a' = \widehat{\alpha'} = -0.208$

$\hat{Y} = -0.208 + 1.072x$

(b)

Statisztikai próbával megoldva:

$H_0\colon\ Y_{x=0} = 0 \qquad H_1\colon\ Y_{x=0} \neq 0$

$s_y^2 = s_r^2 = \frac{\sum(y_i-\hat{Y}_i)^2}{n-2} = 0.0469; \qquad s^2_{\hat{Y}(x=0)} = s_y^2 * \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] = 0.0868$

$t_0 = \frac{\hat{Y}_{x=0}}{s_{\hat{Y}(x=0)}} = \frac{-0.208}{0.2946} = -0.706; \qquad t_{krit} = t_{0.025}(2) = 4.303$

Elfogadási tartomány: $-4.303 < t_0 < 4.303$

→5%-os szignifikanciaszinten elfogadjuk, hogy az igazi egyenes átmegy az origón.

Konfidencia-intervallummal megoldva:

Kérdésünk ekkor az, hogy az $x = 0$ helyen az igazi egyenes milyen értéket vesz fel 95%-os valószínűséggel.

$P(\hat{Y}_{x=0} - t_{\alpha/2} * s_{\hat{Y}(x=0)} < Y_{x=0} < \hat{Y}_{x=0} + t_{\alpha/2} * s_{\hat{Y}(x=0)}) = 0.95$

$P(-1.475 < Y_{x=0} < 1.059) = 0.95$

→ az intervallum tartalmazza a 0-t, tehát 5%-os szignifikanciaszinten elfogadjuk, hogy az igazi egyenes átmegy az origón.

(c) $P(\hat{\beta} - t_{\alpha/2} * s_b < \beta < \hat{\beta} + t_{\alpha/2} * s_b) = 0.90$

$s_b^2 = \frac{s_y^2}{\sum(x_i - \bar{x})^2}$ =0.00469; s_b=0.0685; $t_{krit} = t_{0.05}(2) = 2.92$

$P(0.872 < \beta < 1.272) = 0.90$

(d) $\hat{Y}_{x=4} = 4.08$

Statisztikai próbával megoldva:

H_0: $Y_{x=4} \leq 4$ H_1: $Y_{x=4} > 4$

$t_0 = \frac{\hat{Y}_{x=4} - (\hat{Y}_{x=4}|H_0)}{s_{\hat{Y}(x=4)}} = \frac{4.08-4}{0.1083} = 0.739$ $t_{krit} = t_{0.1}(2) = 1.886$ (egyoldali!)

Elfogadási tartomány: $(-\infty, t_{krit}]$

→ H_0-t elfogadjuk. Az adatok nem bizonyítják 10 %-os szignifikanciaszinten, hogy az igazi egyenes az *x*=4 helyen *Y*=4 fölött halad (de nem is mondanak ellent ennek).

Konfidencia-intervallummal megoldva:

$P(\hat{Y}_{x=4} - t_\alpha * s_{\hat{Y}(x=4)} < Y_{x=4}) = P(4.08 - 1.86 * 0.1083 < Y_{x=4}) =$

$= P(3.88 < Y_{x=4}) = 0.9$

→ 4 alatti értékek is beleesnek az intervallumba, tehát az is elképzelhető, hogy *x*=4 helyen *Y*=4 alatt halad az egyenes. Azaz nem lehetünk biztosak benne 10%-os szignifikanciaszinten, hogy az igazi egyenes az *x*=4 helyen *Y*=4 fölött halad.

(e) Jóslási intervallum kell.

$P(\hat{Y}_{x=4} - t_{\alpha/2} * s_{y^*(x=4)} < y^*_{x=4} < \hat{Y}_{x=4} + t_{\alpha/2} * s_{y^*(x=4)}) = 0.9$

$s^2_{y^*(x=4)} = s_r^2 * \left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i - \bar{x})^2}\right] = 0.0586;$ $s_{y^*(x=4)} = 0.242$

$P(3.373 < y^*_{x=4} < 4.787) = 0.9$

2. példa

Egy kalibráció során mért adatokat (és az azokból számolt mennyiségeket) tartalmazza az alábbi táblázat. *x* jelöli a koncentrációt, *y* az analitikai jelet.

	x	y	y(x-x_átlag)	(x-x_átlag)^2	y_becsült	(y-y_becsült)^2
	0.109	1482	-286.86	0.0375	1449	1075
	0.152	1923	-289.53	0.0227	1980	3302
	0.196	2532				
	0.254	3264				
	0.302	3804	-2.13	0	3834	879
	0.355	4520	237.03	0.0028	4488	997
	0.398	5012	478.35	0.0091	5020	59
	0.455	5751	876.68	0.0232	5724	735
	0.502	6277	1251.88	0.0398	6305	759
szumma	2.723	34565				

(a) Adja meg a becsült egyenes egyenletét!
(b) Adjon 95%-os konfidencia-intervallumot a meredekségre!

(c) Hihető-e az az állítás 1%-os szignifikanciaszinten, hogy az igazi egyenes meredeksége 12100?
(d) Adjon 95%-os konfidencia-intervallumot az egyenes tengelymetszetére!
(e) Hihető-e az az állítás 1%-os szignifikanciaszinten, hogy az igazi egyenes átmegy az origón?
(f) Adjon 90%-os konfidencia-intervallumot az analitikai jel várható értékére x=0.3-nál!
(g) Alátámasztják-e az adatok azt az állítást 5%-os szignifikanciaszinten, hogy az igazi egyenes x=0.3-nál Y=3830 alatt halad?
(h) Milyen intervallumban várható egy új mérés eredménye x=0.3-nál 90%-os valószínűséggel?
(i) Egy új mérés során x=0.3-nál 3920-at mértek. A mért érték egy analitikus szerint túlzottan különbözik attól, amit a becsült kalibrációs egyenes alapján várnánk, emiatt azt gyanítja, hogy elállítódott a készülék. Vizsgálja a kérdést statisztikai módszerrel, legyen a szignifikanciaszint 0.1!
(j) Az alábbi táblázat a fenti adatsorra a STATISTICA szoftverrel végzett lineáris regresszió eredményeit tartalmazza. Válaszolja meg az (a), (b) és (c) kérdéseket az itt közölt eredmények alapján! (A numerikus eltérések a kerekítésnek köszönhetőek.)

	Parameter Estimates (Spreadsheet14) Sigma-restricted parameterization					
Effect	y Param.	y Std.Err	y t	y p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	102,82	29,55379	3,4791	0,010279	32,94	172,70
x	12353,88	89,90255	137,4142	0,000000	12141,30	12566,47

Megoldás:

(a) A táblázat a hiányzó cellákkal (sárga színnel kiemelve):

	x	**y**	**y(x-x_átlag)**	**(x-x_átlag)^2**	**y_becsült**	**(y-y_becsült)^2**
	0.109	1482	-286.86	0.0375	1449	1075
	0.152	1923	-289.53	0.0227	1980	3302
	0.196	2532	-269.81	0.0114	2524	63
	0.254	3264	-158.50	0.0024	3241	546
	0.302	3804	-2.13	0	3834	879
	0.355	4520	237.03	0.0028	4488	997
	0.398	5012	478.35	0.0091	5020	59
	0.455	5751	876.68	0.0232	5724	735
	0.502	6277	1251.88	0.0398	6305	759
szumma	2.723	34565	1837.12	0.1487	34565	8414

$$\bar{x} = \frac{2.373}{9} = 0.30256$$

$$a = \frac{\sum_i y_i}{n} = \frac{34565}{9} = 3840.56$$

$$b = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{1837.12}{0.1487} = 12354.54$$

$$\hat{Y} = 3840.56 + 12354.54(x - 0.30256) = 102.57 + 12354.54x$$

(b) $s_y^2 = s_r^2 = \dfrac{\sum\left(y_i - \hat{Y}_i\right)^2}{n-2} = \dfrac{8414}{7} = 1202$ $\quad s_b^2 = \dfrac{s_y^2}{\sum\left(x_i - \bar{x}\right)^2} = \dfrac{1202}{0.1487} = 8083.39$

$t_{\alpha/2} = t_{0.025}(7) = 2.365$

$P(\hat{\beta} - \mathrm{t}_{\alpha/2} * s_b < \beta < \hat{\beta} + \mathrm{t}_{\alpha/2} * s_b) =$

$= P(12354.54 - 2.365 * \sqrt{8083.39} < \beta < 12354.54 + 2.365 * \sqrt{8083.39}) = 0.95$

$P(12141.91 < \beta < 12567.17) = 0.95$

(c) $H_0 : \beta = 12100$ $\quad H_1 : \beta \neq 12100$

$t_0 = \frac{b-12100}{s_b} = \frac{12354.54-12100}{\sqrt{8083.39}} = 2.831 \quad t_{\alpha/2} = t_{0.005}(7) = 3.499$

$-3.499 < 2.831 < 3.499$ → tehát 1%-os szignifikanciaszinten az adatok nem mondanak ellent annak az állításnak, hogy a meredekség 12100.

(d) $s_{\alpha'}^2 = s_{\hat{Y}(x=0)}^2 = s_y^2 * \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] = 1202 * \left[\frac{1}{9} + \frac{(0-0.30256)^2}{0.1487}\right] = 873.53$

$t_{\alpha/2} = t_{0.025}(7) = 2.365$

$P(32.67 < \alpha' < 172.47) = 0.95$

(e) H_0: $Y_{x=0} = 0$ $\qquad H_1$: $Y_{x=0} \neq 0$

$t_0 = \frac{\hat{Y}_{x=0}}{s_{\hat{Y}(x=0)}} = \frac{102.57}{\sqrt{873.53}} = 3.470; \qquad t_{\alpha/2} = t_{0.005}(7) = 3.499$

Mivel a 3.470 a (-3.499,3.499) intervallumon még éppen belül van → 1%-os szignifikanciaszinten elfogadjuk azt a feltételezést, hogy az egyenes keresztülmegy az origón.

(f) $\hat{Y}_{x=0.3} = 102.57 + 12354.54 * 0.3 = 3808.93$

$s_{\hat{Y}(x=3)}^2 = s_y^2 * \left[\dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] = 1202 * \left[\dfrac{1}{9} + \dfrac{(0.3-0.30256)^2}{0.1487}\right] = 133.56$

$t_{\alpha/2} = t_{0.05}(7) = 1.895$

$P(3787.03 < Y_{x=0.3} < 3830.83) = 0.90$

(g) $H_0 : Y_{x=0.3} \geq 3830$ $\quad H_1 : Y_{x=0.3} < 3830$

$t_0 = \frac{\hat{Y}_{x=3} - \hat{Y}_{x=3}|H_0}{s_{\hat{Y}(x=3)}} = \frac{3808.93-3830}{\sqrt{133.56}} = -1.823; \qquad -t_{0.05}(7) = -1.895$

Az elfogadási tartomány: $[-1.895, \infty)$ → azaz elfogadjuk a nullhipotézist. Az adatok 5 %-os szignifikanciaszinten nem bizonyítják, hogy az egyenes x=0.3-nál 3830 alatt halad.

(h) Jóslási intervallummal kell számolni.

$$s^2_{y^*-\hat{Y}} = s^2_y * \left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] = 1202 * \left[1 + \frac{1}{9} + \frac{(0.3-0.30256)^2}{0.1487}\right] = 1335.61$$

$$P(3739.7 < y^*_{x=3} < 3878.2) = 0.9$$

(i) Mivel a 90%-os jóslási intervallum egy új mérés értékére (ld. (h) alkérdés) nem tartalmazza a mért 3920-as értéket, jogos a gyanú 10%-os szignifikanciaszinten.

(Természetesen a kérdés statisztikai próbával is vizsgálható.

H_0: $E[y^*_{x=0.3}] = \hat{Y}_{x=0.3}$; $\quad t_0 = \frac{y^*-\hat{Y}_{x=0.3}}{s_{\hat{Y}(x=0.3)}} = \frac{3920-3808.93}{\sqrt{1335.61}} = 3.04$; $\quad \pm t_{0.05}(7) = \pm 1.895$

→ Elutasítjuk a nullhipotézist.)

(j)

Effect	Parameter Estimates (Spreadsheet14) Sigma-restricted parameterization y Param.	y Std.Err	y t	y p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	102,82	29,55379	3,4791	0,010279	32,94	172,70
x	12353,88	89,90255	137,4142	0,000000	12141,30	12566,47

A STATISTICA szoftverrel kapott eredménytáblázat első oszlopának *(y Param.)* *Intercept* sorában az egyenes tengelymetszetének, *X* sorában pedig a meredekségének becsült értékét láthatjuk. Ez alapján a becsült egyenes egyenlete:

$$\hat{Y} = 102.82 + 12353.88x$$

(A számolt és a szoftverrel kapott értékek közti numerikus eltérések a kerekítésnek köszönhetőek.)

Az eredménytáblázat ötödik és hatodik oszlopának (*-/+95,00% Cnf.Lmt*) *X* sora mutatja a meredekség 95%-os konfidencia-intervallumát:

$$P(12141.30 < \beta < 12566.47) = 0.95$$

A c) kérdésbeli hipotézisvizsgálathoz számítandó próbastatisztika: $t_0 = \frac{b-12100}{s_b}$

Ehhez *b* értékét már tudjuk (*y Param.* oszlop *X* sora), míg s_b értékét a második oszlop (*y Std.Err*) *X* sora mutatja. Eszerint:

$t_0 = \frac{12353.88-12100}{89.9026} = 2.824 \quad t_{\alpha/2} = t_{0.005}(7) = 3.499$

Mivel $-3.499 < 2.824 < 3.499$, 1%-os szignifikanciaszinten az adatok nem mondanak ellent annak az állításnak, hogy a meredekség 12100.

3. példa

Az alábbi táblázat egy kalibrációs egyenes felvételekor mért adatokat tartalmazza:

x	y
1	2.35
2	4.6
4	7.32
6	10.93
7	13.84

Rendelkezésre állnak az alábbi értékek:

$\sum (x_i - \bar{x})^2 = 26$

$\sum y_i (x_i - \bar{x}) = 47.13$

(a) Adja meg a becsült egyenes egyenletét!

(b) Mennyi a reziduális szórásnégyzet értéke?

(c) Elfogadná-e 5%-os szignifikanciaszinten azt az állítást, hogy az igazi egyenes meredeksége 2?

(d) Adjon 95%-os konfidencia-intervallumot az igazi egyenes tengelymetszetére!

(e) Milyen intervallumban van 90% valószínűséggel az analitikai jel (*y*) várhatóértéke az *x*=5 helyen?

(f) Milyen intervallumban lesz 90% valószínűséggel egy új mérés eredménye az *x*=5 helyen?

Megoldás:

(a) $\hat{Y} = 0.556 + 1.813x$

(b) $s_r^2 = 0.3396$

(c) $s_b^2 = 0.0131$ $\quad t_0 = -1.64;$ $\quad t_{krit} = t_{0.025}(3) = 3.182$

→ 5%-os szignifikanciaszinten elfogadjuk, hogy az igazi egyenes meredeksége 2.

(d) $\hat{Y}_{x=0} = 0.556;$ $\quad s^2_{\hat{Y}(x=0)} = 0.2769;$ $\quad P(-1.118 < Y_{x=0} < 2.232) = 0.95$

(e) $\hat{Y}_{x=5} = 9.62;$ $\quad s^2_{\hat{Y}(x=5)} = 0.081;$ $\quad P(8.95 < Y_{x=5} < 10.29) = 0.90$

(f) $s^2_{y^*} = 0.4206;$ $\quad P(8.094 < y_{x=5} < 11.147) = 0.90$

4. példa

Egy kalibrációs egyenes felvételekor az alábbi adatokat kapták:

x	0.5	0.9	1.1	2.0	2.7	3.3	4.0	4.5
y	2.62	4.48	7.12	12.09	18.05	24.50	28.20	33.60

A fenti adatsorra a STATISTICA szoftverrel elvégezték a lineáris regressziót, melynek eredményét az alábbi táblázat mutatja:

	Parameter Estimates (Linreg_gyak_4 pelda) Sigma-restricted parameterization					
Effect	y Param.	y Std.Err	y t	y p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	-2,061	0,6913	-2,981	0,0246	-3,752	-0,369
x	7,745	0,2509	30,861	0,0000	7,131	8,359

Effect	Univariate Tests of Significance for y (Linreg_gyak_4 pelda) Sigma-restricted parameterization Effective hypothesis decomposition; Std. Error of Estimate: ,9903767				
	SS	Degr. of Freedom	MS	F	p
Intercept	8,7179	1	8,7179	8,8882	0,024595
x	934,1603	1	934,1603	952,4025	0,000000
Error	5,8851	6	0,9808		

Felhasználva a táblázat adatait, válaszoljon az alábbi kérdésekre!

(a) Adja meg a becsült egyenes egyenletét!

(b) Adjon 90%-os konfidencia-intervallumot az egyenes meredekségére!

(c) Elfogadható-e 5%-os szignifikanciaszinten, hogy a tengelymetszet értéke -1?

(d) Milyen x értéknél lesz a legkisebb az illesztett egyenes bizonytalansága? Miért? Mennyi lesz az értéke?

(e) Egy új mérés eredménye x=1-nél 8-nak adódott. Lát-e ebben bármi rendelleneset? Véleményét számolással támassza alá, a szignifikanciaszint legyen 5%.

Megoldás:

(a) $\hat{Y} = -2.061 + 7.745x$

(b) $s_b = 0.2509$; $t_{\alpha/2} = t_{0.05}(6) = 1.943$ $P(7.258 < \beta < 8.232) = 0.90$

(c) H_0: $Y_{x=0} = -1$; H_1: $Y_{x=0} \neq -1$;

$s^2_{\alpha'} = s^2_{\hat{Y}(x=0)} = 0.4779$; $t_0 = -1.53$; $t_{\alpha/2} = t_{0.025}(6) = 2.447$

(d) $\bar{x} = 2.375$

$$s^2_{\hat{Y}(x=\bar{x})} = s^2_y * \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] = 0.9808 * \left[\frac{1}{8} + 0\right] = 0.1226$$

(e) $\hat{Y}_{x=1} = 5.684$; $s^2_{y^*-\hat{Y}} = 1.222$; $t_{\alpha/2} = t_{0.025}(6) = 2.447$

$P(2.98 < y^*_{x=1} < 8.39) = 0.95$, azaz előfordulhat, hogy x=1-nél y-ra 8-t kapunk, mivel x=1-nél számolt 95%-os kétoldali jóslási intervallum tartalmazza a 8-t.