Statistik II - Exercise session 7
21.1.2015

Info

- Classroom: SPA1 203
- Time: Wednesdays, 16:15 - 17:45
- in English
- Assignments on webpage (lvb>staff>PB)

Contact: Petra Burdejova
petra.burdejova@hu-berlin.de
Office: SPA1 R400 (upon agreement)

Schedule:

Date	Week	Exercises
15.10.14	E1	4-11, 4-18, 4-19
22.10.14	E1	4-11, 4-18, 4-19
29.10.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
5.11.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
12.11.14	E3	5-16, 5-21, 5-23, 6-1
19.11.14	E3	5-16, 5-21, 5-23, 6-1
26.11.14	E4	6-3, 6-9, 6-13, 7-1
03.12.14	E4	6-3, 6-9, 6-13, 7-1
10.12.14	E5	7-3, 7-5, 7-26, 7-45
17.12.14	E5	7-3, 7-5, 7-26, 7-45
07.01.15	E6	(7-45), 8-1, 8-4, 8-7
14.01.15	E6	(7-45), 8-1, 8-4, 8-7
21.01.15	E7	8-11, 9-3,9-5,9-7
28.01.15	–	–
04.02.15	E8	Review for exam
11.02.15	E8	Review for exam

Review

- week 11 & week 12
- Slides:
  - Testing of Hypotheses (82-100/162)
  - Regression

χ^2**-test** X discrete, with unknown dist. $F(x)$
Observed frequencies h_i for values x_i
Hypot. distribution assumes $p_i = P(X = x_i | F_0(x))$

$$V = \sum_i \frac{(h_i - n.p_i)^2}{np_i} \sim \chi^2_{I-k-1}$$

k - number of unknown parameters, I - number of values

General regression

$$\begin{aligned} Y_i &= f(x_{1i}, x_{2i}, ..., x_{mi}) + U_i \\ &= E(Y_i) + U_i \qquad \text{with } E(U_i) = 0 \end{aligned}$$

Simple linear regression function

	$E(Y_i) = \beta_0 + \beta_1 \cdot x_i$
Model	$Y_i = E(Y_i) + U_i = \beta_0 + \beta_1 \cdot x_i + U_i$
Error term	$U_i = Y_i - E(Y_i)$
	with $E(U_i) = 0, Var(U_i) = \sigma_u^2, Cov(U_i U_j) = 0$ for $i \neq j$
	and $U_i \sim N(0; \sigma_u^2)$
Fitted values	$\hat{y}_i = b_0 + b_1 \cdot x_i$
	$y_i = \hat{y}_i + \hat{u}_i = b_0 + b_1 \cdot x_i + \hat{u}_i$
Residuals	$\hat{u}_i = y_i - \hat{y}_i$

Least-Squares estimator for β_0, β_1 minimize $\sum \hat{u}_i$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{s_{xy}}{s_x^2} = r_{xy} \cdot \frac{s_y}{s_x}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i \cdot y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i} = \overline{y} - b_1 \cdot \overline{x}$$

Coefficient of determination

$$R^2_{yx} = R^2_{xy} = \frac{s^2_{yx}}{s^2_y \cdot s^2_x} = r^2_{yx} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Exercises

Exercise 8-11 - Coins

Three distinct coins are thrown 240 times and each time was the number of ”head” is observed. The results are summarized as follows:

0x head: 24

1x head: 108

2x head: 85

3x head: 23

Test the hypothesis that three coins are ideal. (α = 0,05)! (In an ideal coin heads and tails with are thrown with equal probability.)

Exercise 9-3 - Cross-sectional analysis of 11 companies

In a cross-sectional analysis of 11 companies in each sector we examine the dependence of Y - sales (in mil. EUR) from investments X1 (in 1 000 EUR), the expenses for research and development X2 (in 1 000 EUR) and advertising expenses X3 (in 1 000 EUR) for a given period of time. The values of explanatory variables X1, X2, X3 and Y are stated in the following table:

i	y_i	x_{i1}	x_{i2}	x_{i3}
1	12,6	117,0	84,5	3,1
2	13,1	126,3	89,7	3,6
3	15,1	134,4	96,2	2,3
4	15,1	137,5	99,1	2,3
5	14,9	141,7	103,2	0,9
6	16,1	149,4	107,5	2,1
7	17,9	158,4	114,1	1,5
8	21,0	166,5	120,4	3,8
9	22,3	177,1	126,8	3,6
10	21,9	179,8	127,2	4,1
11	21,0	183,8	128,7	1,9

a) Determine the simple linear regression function of sales with respect to investments or expenditures for research and development and advertising expenses, as well as the associated coefficient of determination.

b) Calculate all simple linear correlation coefficients between these characteristics.

Exercise 9-5 - Consumption expenditure

In March 1992 the total available income of 8 households was 30 880 EUR. In the same month actuated every 8 households consumption expenditure of 26 800 EUR. Per EUR of additional income of these households in average 0.813 EUR was spent on consumption.

a) State the (economically viable) linear regression function.

b) Which consumer spending are expected on average for level of disposable income of 2 800 EUR ?

Exercise 9-7 - Consumption expenditure

The variables X and Y were observed in 9 statistical units, however, not the individual pairs of values (x_i, y_i), but instead the sums

$$\sum_{i=1}^{9} x_i = 34, \sum_{i=1}^{9} y_i = 60, \sum_{i=1}^{9} x_i^2 = 144, \sum_{i=1}^{9} y_i^2 = 422, \sum_{i=1}^{9} x_i y_i = 244.$$

Subsequently it turns out that the pair of values (x_{10}, y_{10}) = (6; 10) needs to be taken into account. Which of the following regression line $y = b_0 + b_1 x$ is correct by the method of least squares for the 10 pairs of values?

a) $y = 0,2 + 1,7x$ b) $y = 0,6 + 1,6x$ c) $y = 1,2 + 1,5x$ d) $y = 1,4 + 1,4x$
e) $y = 1,8 + 1,3x$ f) $y = 2,0 + 1,8x$ g) $y = 2,2 + 1,2x$ h) $y = 2,8 + 1,0x$