MATHEMATICS 201-510-LW
Business Statistics
Martin Huard
Fall 2008

XXVI – Inferences for Regression
SOLUTIONS

1. Some students claim they can tell the cost of a textbook just by looking at its thickness. To test this claim they picked four hardbound books of the same height and width at random. The cost and thickness for each book was:

x (Thickness in cm)	1.3	3.5	2.6	4.1
y (Cost in \$)	72	88	65	105

a) Find the equation of the least-squares line.

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 37.51 - \frac{11.5^2}{4} = 4.4475$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 1001.1 - \frac{11.5 \cdot 330}{4} = 52.35$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 28178 - \frac{330^2}{4} = 953$$

$$\bar{x} = \frac{\sum x}{n} = \frac{11.5}{4} = 2.875 \qquad \bar{y} = \frac{\sum y}{n} = \frac{330}{4} = 82.5$$

Slope : $b = \frac{SS_{xy}}{SS_x} = \frac{52.35}{4.4475} = 11.77$

y-intercept : $a = \bar{y} - b\bar{x} = 82.5 - (11.77)(2.875) = 48.66$

Thus the least-squares line is given by $y = 48.66 + 11.77x$.

b) Find a 95% confidence interval for the y-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = n-2 = 4-2 = 2$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $a = 48.66$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(2,0.025)} = 4.303$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{\left(SS_{xy}\right)^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} s_e \sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{SS_x}} = 4.303 \cdot 12.977 \sqrt{\dfrac{1}{4} + \dfrac{2.875^2}{4.4475}} = 81.08$

d) $a - E < \alpha < a + E$

$48.66 - 81.08 < \alpha < 48.66 - 81.08$

$-32.42 < \alpha < 129.74$

Step 5 The 95% confidence interval for the regression coefficient α is -32.42 to 129.74.

c) Find a 95% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = n-2 = 4-2 = 2$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $b = 11.77$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(2,0.025)} = 4.303$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{\left(SS_{xy}\right)^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 4.303 \dfrac{12.977}{\sqrt{4.4475}} = 26.48$

d) $b - E < \beta < b + E$

$11.77 - 26.48 < \beta < 11.77 + 26.48$

$-14.71 < \beta < 38.25$

Step 5 The 95% confidence interval for the regression coefficient β is -14.71 to 38.25.

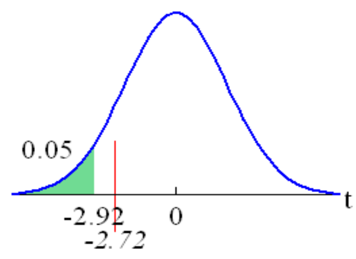
d) Determine if the y-intercept α of the regression line is less than 100 at the 5% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o : \alpha = 100$

$H_a : \alpha < 100$



Step 3 a) Test statistic: t with $df = 2$

b) Left-tailed test with α = 0.05

c) $t_{(df,1-\alpha)} = t_{(2,0.95)} = -2.920$

Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\overline{x}^2}} = \dfrac{48.66-100}{12.977}\sqrt{\dfrac{4 \cdot 4.4475}{4.4475 + 4(2.875)^2}} = -2.72$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the y-intercept α of the regression line is less than 100.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o : \alpha = 100$

$H_a : \alpha < 100$

Step 3 a) Test statistic: t with $df = 2$

b) Left-tailed test with α = 0.05

Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\overline{x}^2}} = \dfrac{48.66-100}{12.977}\sqrt{\dfrac{4 \cdot 4.4475}{4.4475 + 4(2.875)^2}} = -2.72$

c) $0.054 < p\text{-value} < 0.057$

Step 5 a) p-value > 0.054 > α = 0.05

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the y-intercept α of the regression line is less than 100.

e) Determine if the slope β of the regression line is positive at 5% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

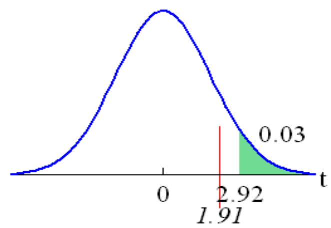
Step 2 $H_o: \beta = 0$

$H_a: \beta > 0$

Step 3 a) Test statistic: t with $df = 2$

b) Right-tailed test with α = 0.05

c) $t_{(df,\alpha)} = t_{(2,0.05)} = 2.920$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{11.77-0}{12.977}\sqrt{4.4475} = 1.91$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is positive.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 0$

$H_a: \beta > 0$

Step 3 a) Test statistic: t with $df = 2$

b) Right-tailed test with α = 0.05

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{11.77-0}{12.977}\sqrt{4.4475} = 1.91$

c) $0.092 < p\text{-value} < 0.099$

Step 5 a) p-value > 0.092 > α = 0.05

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is positive.

f) If a book has a thickness of 2.2 cm, find a 95% confidence interval for the predicted cost.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 2$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\hat{y} = 48.66 + 11.77 \cdot 2.2 = 74.55$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(2, 0.025)} = 4.303$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{\left(SS_{xy}\right)^2}{SS_x}}{n-2}} = \sqrt{\dfrac{953 - \frac{52.35^2}{4.4475}}{2}} = 12.977$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{1 + \dfrac{1}{n} + \dfrac{\left(x - \overline{x}\right)^2}{SS_x}}$

$= 4.303 \cdot 12.977 \sqrt{1 + \dfrac{1}{4} + \dfrac{\left(2.2 - 2.875\right)^2}{4.4475}} = 64.93$

c) $\hat{y} - E < y < \hat{y} + E$

$74.55 - 64.93 < y < 74.55 + 64.93$

$9.62 < y < 139.48$

Step 5 The 95% confidence interval for predicted cost of a book with a thickness of 2.2 cm is \$9.62 to \$139.48

2. Do reading and TV viewing compete for leisure time? To find out, a communication specialist interviewed a sample of children regarding the number of books they had read during the last year and the number of hours they had spent watching TV on a daily basis.

Daily Hours of TV Viewing	**Yearly Number of Books Read**
3	0
1	7
2	2
2	1
0	5
1	4
3	3
2	3
7	0
4	1

a) Find the equation of the least-squares line.

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 97 - \frac{25^2}{10} = 34.5$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 36 - \frac{25 \cdot 26}{10} = -29$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 114 - \frac{26^2}{10} = 46.4$$

$$\bar{x} = \frac{\sum x}{n} = \frac{25}{10} = 2.5 \qquad \bar{y} = \frac{\sum y}{n} = \frac{26}{10} = 2.6$$

$$\text{Slope}: \ b = \frac{SS_{xy}}{SS_x} = \frac{-29}{34.5} = -0.841$$

$$y\text{-intercept}: \ a = \bar{y} - b\bar{x} = 2.6 - (-0.841)(2.5) = 4.701$$

Thus the least-squares line is given by $y = 4.701 - 0.841x$.

b) Find a 99% confidence interval for the y-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 8$

b) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate: $a = 4.701$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(8,0.005)} = 3.355$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} s_e \sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{SS_x}} = 3.355 \cdot 1.659 \sqrt{\dfrac{1}{10} + \dfrac{2.5^2}{34.5}} = 2.952$

d) $a - E < \alpha < a + E$

$4.701 - 2.952 < \alpha < 4.701 + 2.952$

$1.749 < \alpha < 7.653$

Step 5 The 95% confidence interval for the regression coefficient α is 1.749 to 7.653.

c) Find a 99% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 8$

b) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate: b = -0.841

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(8,0.005)} = 3.355$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 3.355 \dfrac{1.659}{\sqrt{34.5}} = 0.948$

d) $b - E < \beta < b + E$

$-0.841 - 0.948 < \beta < -0.841 + 0.948$

$-1.788 < \beta < 0.107$

Step 5 The 95% confidence interval for the regression coefficient β is -1.788 to 0.107.

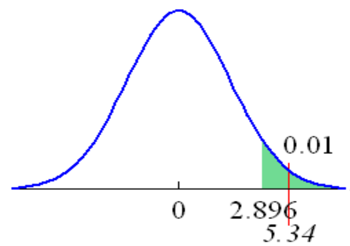
d) Determine if the y-intercept α of the regression line is positive at the 1% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \alpha = 0$

$H_a: \alpha > 0$



Step 3 a) Test statistic: t with $df = 8$

b) Right-tailed test with $\alpha = 0.01$

c) $t_{(df,\alpha)} = t_{(8,0.01)} = 2.896$

Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\overline{x}^2}} = \dfrac{4.701-0}{1.659}\sqrt{\dfrac{10\cdot 34.5}{34.5+10(2.5)^2}} = 5.34$

Step 5 a) t is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that the y-intercept α of the regression line is positive.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \alpha = 0$

$H_a: \alpha > 0$

Step 3 a) Test statistic: t with $df = 8$

b) Right-tailed test with $\alpha = 0.01$

Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\overline{x}^2}} = \dfrac{4.701-0}{1.659}\sqrt{\dfrac{10\cdot 34.5}{34.5+10(2.5)^2}} = 5.34$

c) $p\text{-value} < 0.002$

Step 5 a) $p\text{-value} < 0.002 < \alpha = 0.01$

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that the y-intercept α of the regression line is positive.

e) Determine if the slope β of the regression line is negative at 1% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

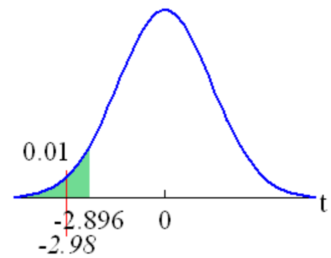
Step 2 $H_o: \beta = 0$

$H_a: \beta < 0$

Step 3 a) Test statistic: *t* with *df* = 8

b) Left-tailed test with α = 0.01

c) $t_{(df,1-\alpha)} = t_{(8,0.99)} = -2.896$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{-0.841-0}{1.659}\sqrt{34.5} = -2.98$

Step 5 a) *t* is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is negative.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 0$

$H_a: \beta < 0$

Step 3 a) Test statistic: *t* with *df* = 8

b) Left-tailed test with α = 0.01

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{-0.841-0}{1.659}\sqrt{34.5} = -2.98$

c) $0.009 < p\text{-value} < 0.010$

Step 5 a) p-value < α = 0.01

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is negative.

f) If a child watches 2 hours of television per day, a 99% confidence interval for the predicted yearly number of books read.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 8$

b) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate: $\hat{y} = 4.701 - 0.841 \cdot 2 = 3.020$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(8, 0.005)} = 3.355$

b) $$S_e = \sqrt{\frac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\frac{46.4 - \frac{(-29)^2}{34.5}}{8}} = 1.659$$

c) $$E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}}$$

$$= 3.355 \cdot 1.659 \sqrt{1 + \frac{1}{10} + \frac{(2-2.5)^2}{34.5}} = 5.858$$

d) $$\hat{y} - E < y < \hat{y} + E$$

$$3.020 - 5.858 < y < 3.020 + 5.858$$

$$-2.838 < y < 8.878$$

Step 5 The 99% confidence interval for predicted yearly number of books read is 0 to 8.878 books.

3. Eight people applying for a job as a graphic designer were given two tests, one measuring the applicant's logical reasoning ability (on a scale of 1 to 20), the other measuring the applicant's creativity (on a scale of 1 to 30). Here are the results

Reason (scale 1 to 20)	**Creativity (scale 1 to 30)**
13	18
13	20
18	31
14	25
9	23
9	21
5	5
10	21

a) Find the equation of the least-squares line.

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 1145 - \frac{91^2}{8} = 109.875$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 2033 - \frac{91 \cdot 164}{8} = 167.5$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 3746 - \frac{164^2}{8} = 384$$

$$\overline{x} = \frac{\sum x}{n} = \frac{91}{8} = 11.375 \qquad \overline{y} = \frac{\sum y}{n} = \frac{164}{8} = 20.5$$

Slope : $b = \frac{SS_{xy}}{SS_x} = \frac{167.5}{109.875} = 1.524$

y-intercept : $a = \overline{y} - b\overline{x} = 20.5 - (1.524)(11.375) = 3.159$

Thus the least-squares line is given by $y = 3.159 + 1.524x$.

b) Find a 90% confidence interval for the y-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 6$

b) Level of confidence: $1-\alpha = 0.90$ or $\alpha = 0.10$

Step 3 Point estimate: $a = 3.159$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(6,0.05)} = 1.943$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{SS_x}} = 1.943 \cdot 4.631 \sqrt{\dfrac{1}{8} + \dfrac{11.375^2}{109.875}} = 10.270$

d) $a - E < \alpha < a + E$

$3.159 - 10.270 < \alpha < 3.159 + 10.270$

$-7.111 < \alpha < 13.429$

Step 5 The 90% confidence interval for the regression coefficient α is -7.111 to 13.429.

c) Find a 90% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 6$

b) Level of confidence: $1-\alpha = 0.90$ or $\alpha = 0.10$

Step 3 Point estimate: $b = 1.524$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(6,0.05)} = 1.943$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 1.943 \dfrac{4.631}{\sqrt{109.875}} = 0.858$

d) $b - E < \beta < b + E$

$1.524 - 0.858 < \beta < 1.524 + 0.858$

$0.666 < \beta < 2.383$

Step 5 The 90% confidence interval for the regression coefficient β is 0.666 to 2.383.

d) Determine if the y-intercept α of the regression line is different than 5 at the 10% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

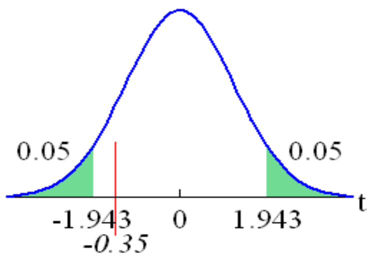
Step 2 $H_o : \alpha = 5$

$H_a : \alpha \neq 5$

Step 3 a) Test statistic: t with $df = 6$

b) Two-tailed test with $\alpha = 0.10$

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(6, 0.05)} = 1.943$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

b) $t = \dfrac{a - \alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{3.159 - 5}{4.631}\sqrt{\dfrac{8 \cdot 109.875}{109.875 + 8(11.375)^2}} = -0.35$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is insufficient evidence at the 10% level of significance to conclude that the y-intercept α of the regression line is different than 5.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o : \alpha = 5$

$H_a : \alpha \neq 5$

Step 3 a) Test statistic: t with $df = 6$

b) Two-tailed test with $\alpha = 0.10$

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

b) $t = \dfrac{a - \alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{3.159 - 5}{4.631}\sqrt{\dfrac{8 \cdot 109.875}{109.875 + 8(11.375)^2}} = -0.35$

c) $2 \cdot 0.352 < p\text{-value} < 2 \cdot 0.387$

$0.704 < p\text{-value} < 0.774$

Step 5 a) p-value $> 0.704 > \alpha = 0.10$

b) Fail to reject H_o.

∴ There is insufficient evidence at the 10% level of significance to conclude that the y-intercept α of the regression line is different than 5.

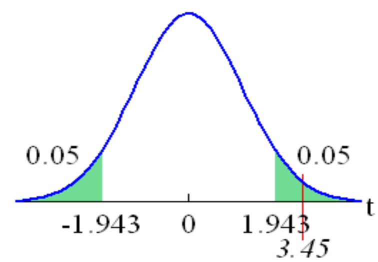
e) Determine if the slope β of the regression line is different than zero at the 10% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 0$

$H_a: \beta \neq 0$



Step 3 a) Test statistic: t with $df = 6$

b) Two-tailed test with $\alpha = 0.10$

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(6,0.05)} = 1.943$

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{1.524-0}{4.631}\sqrt{109.875} = 3.45$

Step 5 a) t is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 10% level of significance to conclude that the slope β of the regression line is different than zero.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 0$

$H_a: \beta \neq 0$

Step 3 a) Test statistic: t with $df = 6$

b) Two-tailed test with $\alpha = 0.10$

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{1.524-0}{4.631}\sqrt{109.875} = 3.45$

c) $2 \cdot 0.006 < p\text{-value} < 2 \cdot 0.007$

$0.012 < p\text{-value} < 0.014$

Step 5 a) $p\text{-value} < 0.014 < \alpha = 0.01$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is different than zero.

f) If a person scores 12 on reason, find a 90% confidence interval for the predicted score in creativity.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 6$

b) Level of confidence: $1-\alpha = 0.90$ or $\alpha = 0.10$

Step 3 Point estimate: $\hat{y} = 3.159 + 1.524 \cdot 12 = 21.453$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(6, 0.05)} = 1.943$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{384 - \frac{167.5^2}{109.875}}{6}} = 4.631$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{SS_x}}$

$= 1.943 \cdot 4.631 \sqrt{1 + \dfrac{1}{8} + \dfrac{(12 - 11.375)^2}{109.875}} = 9.559$

d) $\hat{y} - E < y < \hat{y} + E$

$21.453 - 9.559 < y < 21.453 + 9.559$

$11.894 < y < 31.012$

Step 5 The 90% confidence interval for predicted for the predicted score on creativity for a score of 12 on reason is 11.89 to 30.

4. A company wants to explore the relationship between its annual advertising spending x (in \$) and its annual sales y (in 1000\$). A random sample of 28 weeks was taken, and the following results were calculated.

$SS_x = 3.819 \quad SS_y = 6434 \quad SS_{xy} = 156.5 \quad \bar{x} = 1.440 \quad \bar{y} = 57.69$

a) Find the equation of the least-squares line.

Slope : $b = \dfrac{SS_{xy}}{SS_x} = \dfrac{156.5}{3.819} = 40.98$

y-intercept : $a = \bar{y} - b\bar{x} = 57.69 - (40.98)(1.440) = -1.321$

Thus the least-squares line is given by $y = -1.321 + 40.98x$.

b) Find a 95% confidence interval for the *y*-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: *t* with $df = 23$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $a = -1.321$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(23, 0.025)} = 2.069$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{SS_x}} = 2.069 \cdot 0.950 \sqrt{\dfrac{1}{25} + \dfrac{1.440^2}{3.819}} = 1.500$

d) $a - E < \alpha < a + E$

$-1.321 - 1.500 < \alpha < -1.321 + 1.500$

$-2.821 < \alpha < 0.179$

Step 5 The 95% confidence interval for the regression coefficient α is -2.281 to 0.179.

c) Find a 95% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: *t* with $df = 23$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $b = 40.98$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(23, 0.025)} = 2.069$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 2.069 \dfrac{0.950}{\sqrt{3.819}} = 1.005$

d) $b - E < \beta < b + E$

$40.98 - 1.01 < \beta < 40.98 + 1.01$

$39.97 < \beta < 41.99$

Step 5 The 95% confidence interval for the regression coefficient β is 39.97 to 41.99.

d) Determine if the y-intercept α of the regression line is different than 0 at the 5% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

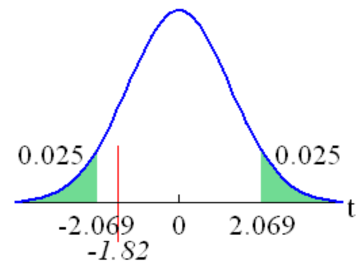
Step 2 $H_o: \alpha = 0$

$H_a: \alpha \neq 0$

Step 3 a) Test statistic: t with $df = 23$

b) Two-tailed test with α = 0.05

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(23, 0.025)} = 2.069$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{-1.321}{0.950}\sqrt{\dfrac{25 \cdot 3.819}{3.819 + 25(1.440)^2}} = -1.82$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is insufficient evidence at the 5% level of significance to conclude that the y-intercept α of the regression line is different than 0.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \alpha = 0$

$H_a: \alpha \neq 0$

Step 3 a) Test statistic: t with $df = 23$

b) Two-tailed test with α = 0.05

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{-1.321}{0.950}\sqrt{\dfrac{25 \cdot 3.819}{3.819 + 25(1.440)^2}} = -1.82$

c) $2 \cdot 0.035 < p\text{-value} < 2 \cdot 0.043$

$0.070 < p\text{-value} < 0.086$

Step 5 a) p-value > 0.070 > α = 0.05

b) Fail to reject H_o.

∴ There is insufficient evidence at the 5% level of significance to conclude that the y-intercept α of the regression line is different than 0.

e) Determine if the slope β of the regression line is greater than 40 at the 5% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

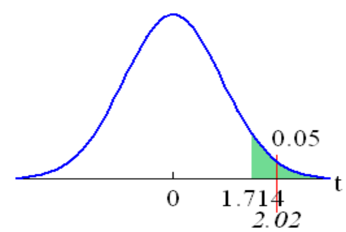
Step 2 $H_o: \beta = 40$

$H_a: \beta > 40$

Step 3 a) Test statistic: *t* with *df* = 23

b) Right-tailed test with α = 0.10

c) $t_{(df,\alpha)} = t_{(23,0.05)} = 1.714$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{40.98-40}{0.950}\sqrt{3.819} = 2.02$

Step 5 a) *t* is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is greater than 40.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 40$

$H_a: \beta > 40$

Step 3 a) Test statistic: *t* with *df* = 23

b) Right-tailed test with α = 0.10

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{40.98-40}{0.950}\sqrt{3.819} = 2.02$

c) $0.023 < p\text{-value} < 0.029$

Step 5 a) p-value $< 0.029 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is greater than 40.

f) If the amount spent on advertising is \$2000, find a 95% confidence interval for the predicted sales.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 23$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\hat{y} = -1.321 + 40.98 \cdot 2 = 80.639$

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(23, 0.025)} = 2.069$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{\left(SS_{xy}\right)^2}{SS_x}}{n-2}} = \sqrt{\dfrac{6434 - \frac{156.5^2}{3.819}}{23}} = 0.950$

c) $E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{1 + \dfrac{1}{n} + \dfrac{\left(x - \overline{x}\right)^2}{SS_x}}$

$= 2.069 \cdot 0.950 \sqrt{1 + \dfrac{1}{25} + \dfrac{\left(2 - 1.44\right)^2}{3.819}} = 2.081$

d) $\hat{y} - E < y < \hat{y} + E$

$80.639 - 2.081 < y < 80.639 + 2.081$

$78.558 < y < 82.720$

Step 5 The 95% confidence interval for predicted for the predicted sales if \$2000 is spent on advertising is \$78 558 to \$82 720.