Stockholms universitet

Department of Statistics

Correction sheet

Date: 14/02/2019

Room: Värtasalen

Course: Basic statistics for economists (eng)

Exam: Statistics for economists (eng)

Anonymous code: 0050-0LC

☒ I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	X	X	X	X	X			4
Teacher's notes 10	10	15	10	10	20	20			

1k

Points	Grade	Teacher's sign.
95	A	NO

STOCKHOLM UNIVERSITY
Department of Statistics

Autumn 2018 C-D

ANSWER FORM Exam – Basic statistics for economists
2019-02-14

Room: Vårtasalen

Anonymous code: 0050 - OLC *(write clearly!)*

Mark your answers with a clear cross (X) in the corresponding boxes below.

NOTE! Only one cross per question. If more than one alternative has been marked, zero points will be awarded for that question.

NOTE! If, after checking your calculations properly, you are convinced that the correct answer is not included among the given alternatives, write your answer in the margin to the right and explain you reasoning on the back.

		A	B	C	D	E	
Problem 1	a)		X				R
	b)			X			R
Problem 2	a)				X		R
	b)	X					R
Problem 3	a)	X					R
	b)					X	R
	c)		X				R
Problem 4	a)				X		—
	b)					X	R
	c)			X			R
Problem 5	a)				X		R
	b)	X					R

55/60

SU, DEPARTMENT OF STATISTICS

Room: Värtasalen Anonymous code: 0050-0LC Sheet number: 1

6. a) assumptions:
- Central Limit Theorem -> "positive towards the brand" is normally distributed
- observations within the samples are identically and independently distributed
- samples are independent of each other
- samples have been obtained through random sampling
- P(being positive to the brand) is a constant

u

test statistic:

$$z = \frac{\hat{p}_x - \hat{p}_y - 0}{\sqrt{\hat{p}_0 \cdot (1-\hat{p}_0) \cdot (\frac{1}{n_x} + \frac{1}{n_y})}} \quad \text{where } \hat{p}_0 = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}$$

here; $n_1 = n_2$
$$z = \frac{\hat{p}_2 - \hat{p}_1 - 0}{\sqrt{\hat{p}_0 \cdot (1-\hat{p}_0) \cdot \frac{2}{n}}} \quad \text{where } \hat{p}_0 = \frac{n(\hat{p}_2 + \hat{p}_1)}{2n}$$

↳ z is normally distributed with $N(0,1)$

5

b) $H_0: \hat{p}_2 - \hat{p}_1 = 0$ $H_1: \hat{p}_2 - \hat{p}_1 > 0$

decision rule & critical value:

H_0 is rejected if $z_{obs} > z_\alpha = z_{0.01} = 2.3263 = z_{crit}$

5

c) $$\hat{p}_0 = \frac{n(\hat{p}_2 + \hat{p}_1)}{2n} = \frac{800 \cdot (0.35 + 0.41)}{2 \cdot 800} = \frac{800 \cdot 0.76}{1600}$$
$$= 0.38$$

$$z_{obs} = \frac{\hat{p}_2 - \hat{p}_1 - 0}{\sqrt{\hat{p}_0 \cdot (1-\hat{p}_0) \cdot \frac{2}{n}}} = \frac{0.41 - 0.35}{\sqrt{0.38 \cdot 0.62 \cdot \frac{1}{400}}} = \frac{0.06}{\sqrt{0.000589}}$$
$$= 2.47225693$$ R

$z_{obs} \approx 2.4723 > 2.3263 = z_{crit}$

↳ H_0 is rejected. R

At a 1% significance level, there is a statistically significant increase in the proportion of those positive to the brand. R

5

d) Using the p-value is a different method of doing a hypothesis test. The p-value corresponds to α, i.e. describes the degree of confidence in a value. If α is small, z_α is large. Therefore, the decision rules are different. Here, the decision rule was „H_0 is rejected if $z_{obs} > z_\alpha$". Thus, the decision rule using α is „H_0 is rejected if p-value < α". The p-value thus describes to what extent a value is statistically significant.

$z_{obs} \approx 2.4723 \xrightarrow{\text{TABLE 2}}$ p-value ≈ 0.008

use table 1 !

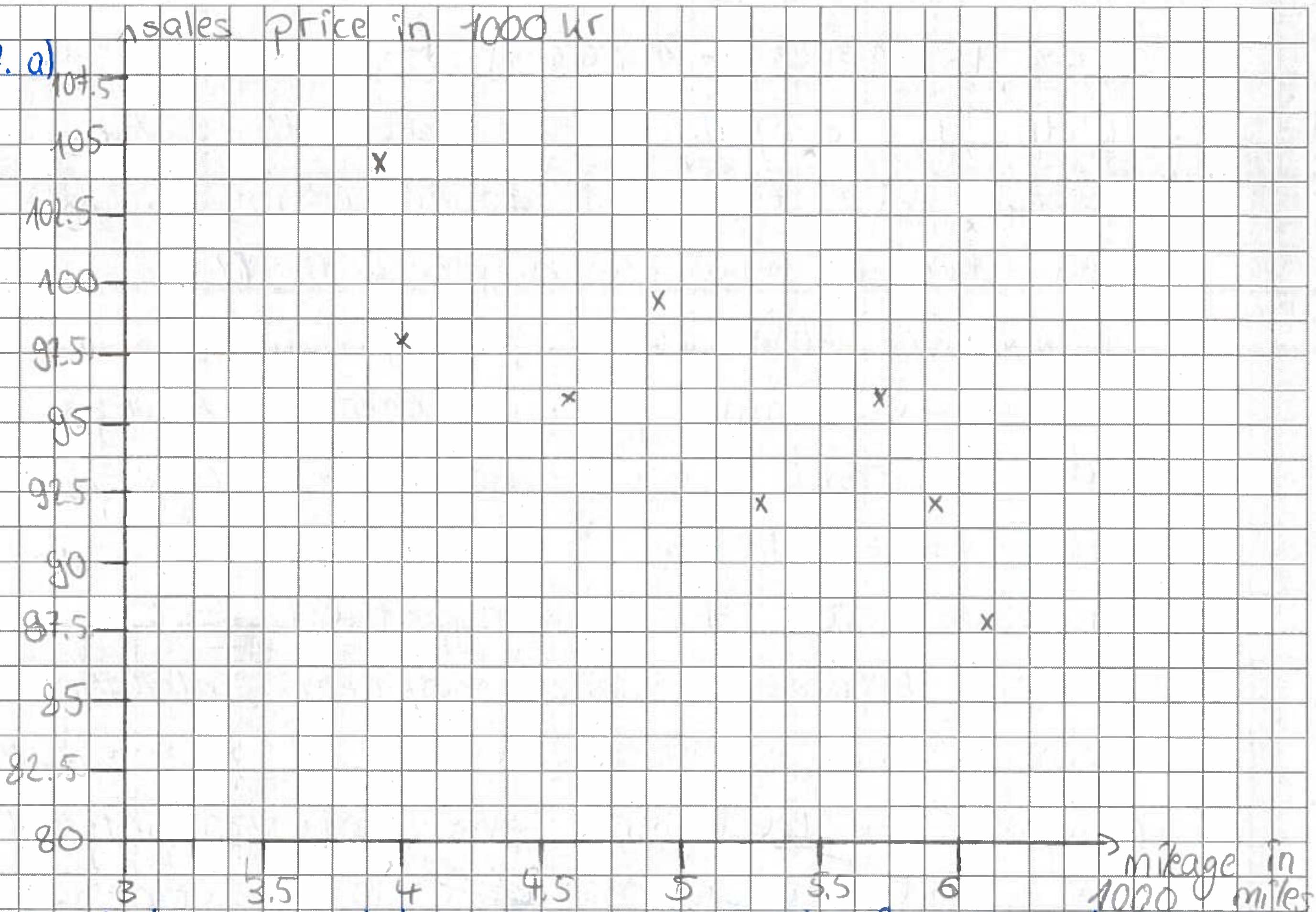
5

20

SU, DEPARTMENT OF STATISTICS

Room: Värtasalen **Anonymous code:** 0050-OLC **Sheet number:** 2

7. a)



A linear model seems appropriate for this data, since the sales price shows a clear downward movement as the mileage increases. However, just from looking at the scatter plot, it seems likely that SSE and the residual variance will be quite big (will see!) because the data points are far from forming a straight line. Thus, a second independent variable might be useful.

5

b) b_1 in model 1: $b_1 = -4.9496$

95% confidence interval for b_1:

$b_1 \pm t_{n-k-1;\, \alpha/2} \cdot s_{b_1}$

$b_1 \pm t_{6;\, 0.025} \cdot s_{b_1}$

$-4.9496 \pm 2.447 \cdot 1.300879$

-4.9496 ± 3.183250913

$\approx -4.9496 \pm 3.1833$

$\hookrightarrow [-8.1329; -1.7663]$ R

If we draw similar intervals (also with 95% confidence, i.e. $t_{6;0.025}$) around estimates for β_1 that are based on samples with $n = 8$ observations from the same population and drawn randomly, then 95% of these intervals will contain the true value of β_1. R

I would reject the null hypothesis that $\beta_1 = 0$ because a 95% confidence interval around b_1 is comparable to a test on b_1 with a 5% significance level. Since the confidence interval does not contain 0, there is no evidence of $\beta_1 = 0$, so the hypothesis is rejected. R /5

c) Model 2: R^2, R^2_{adj}:

$$R^2 = \frac{SSR}{SSE} = \frac{165.8457}{171.8750} = 0.9649204364 \approx 0.9649$$ R

$$R^2_{adj} = 1 - \frac{SSE : (n-k-1)}{SST : (n-1)} =$$

$$= 1 - \frac{MSE}{SST : 7} = 1 - \frac{1.205858}{24.55357143}$$

$$= 1 - 0.0491113076 3$$

$$= 0.9508886924$$

$$\approx 0.9509$$ R

SU, DEPARTMENT OF STATISTICS

Room: Vårtasalen Anonymous code: 0050-OLC Sheet number: 3

7.c) continued:

Both R^2 and R^2_{adj} are (much) higher for Model 2 than for Model 1, which indicates that Model 2 is better suited to explaining the variation in used car sales prices. Also, the difference between R^2_{adj} and R^2 is smaller for Model 2 than for Model 1, which tells us that the added independent variable b_2 describing the fuel type's influence on sales prices really improves the explanation." Thus, I prefer Model 2. 1/5

d) $H_0: \beta_1 = \beta_2 = 0$ H_1: at least one $\beta_j \neq 0$

The p-value given in the ANOVA printout is equal to 0.00023. Thus, I reject the null hypothesis. In a hypothesis test using the p-value, H_0 is rejected as long as the p-value is smaller than the significance level α. Therefore, H_0 would only be accepted for this model if the test was done at an extremely low significance level. In other words, it is highly unlikely that this model based on a car's mileage and its fuel type does not explain the variation in the variation in a car's sales price, which is equal to saying it is highly unlikely that H_0 is true. 5/5

20