Stockholm University

STOCKHOLM UNIVERSITY

Department of Statistics

Autumn 2019, period C-D

Ulf Högnäs (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2020-02-13

Time:	16.00 - 21.00 (4PM – 9PM)
Approved aid:	Hand-held calculator with no stored text, data or formulas
Provided aid:	*Formula Sheet and Probability Distribution Tables*, returned after the exam, English-Swedish dictionaries available on-site

- **Problems 1 – 5: MULTIPLE CHOICE QUESTIONS – max 60 points**
  - A total of 12 multiple choice questions with five alternative answers per question one of which is the correct answer. Mark your answers on the attached **answer form**.
  - Marking more than one alternative will result in zero points for that question.
  - Written solutions should not submitted; only your answers on the answer form will be considered in the assessment and final grading.

- **Problems 6 – 7: COMPLETE WRITTEN SOLUTIONS – max 40 points**
  - Use only the provided **answer sheets** when submitting your solutions and answers.
  - For full marks, clear, comprehensive and well-motivated solutions are required. Unclear and unexplained solutions may result in point deductions even if the final answer is correct.
  - Check your calculations and solutions before submitting. Careless mistakes may result in unnecessary point deductions.

- The maximum number of points is stated for each question. The maximum total number of points is 60 + 40 = 100. At least 50 points is required to pass (grades A-E). The grading scale is as follows:

  A: 90 – 100 points
  B: 80 – 89 points
  C: 70 – 79 points
  D: 60 – 69 points
  E: 50 – 59 points
  Fx: 40 – 49 points
  F: 0 – 40 points

  NOTE! Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement for a higher grade.

- **GOOD LUCK!**

Problem 1

a) The table below shows ten five-year-olds' savings, in Swedish crowns. **Find the first and third quartiles, according to the method taught in this class.** (5p)

44	61	34	3	17	26	19	87	23	86

(A) $Q_1 = 16.75\,;\ Q_3 = 67.25$

(B) $Q_1 = 16.75\,;\ Q_3 = 67.75$

(C) $Q_1 = 18.25\,;\ Q_3 = 67.25$

(D) $Q_1 = 18.50\,;\ Q_3 = 67.25$

(E) $Q_1 = 18.50\,;\ Q_3 = 67.75$

b) The table below shows the joint frequency distribution of two variables x and y, from a sample of size $n = 10$:

Frequency	$y = 0$	$y = 1$
$x = 0$	4	1
$x = 1$	4	1

Find the covariance between x and y. (5p)

(A) -0.222

(B) -0.200

(C) 0

(D) 0.200

(E) 0.222

Problem 2

a) The European Union commissioned a study on how often people, aged 15 and older, engage in physical activity (B_1 = *Regularly;* B_2 = *Sometimes;* B_3 = *Seldom or Never*). The data was also been divided into two age groups (A = *15-24 years old,* $\bar{A}$ = *25 and older*). A person is selected at random and the joint probability for each of the six possible combinations are presented in the table below. Note that B_1, B_2, and B_3 are mutually exclusive and that $\bar{A}$ is the complement of A.

	B_1	B_2	B_3
A	0.03	0.05	0.08
$\bar{A}$	0.12	0.21	0.51

Which of the following statements is false? (5p)

(A) $P(\bar{A}) = 0.84$

(B) $P(B_1 \mid A) = 0.20$

(C) $P(B_2 \mid \bar{A}) = 0.25$

(D) $P(B_2 \cup \bar{A}) = 0.89$

(E) $P(A \cap \overline{B_2}) = 0.11$

b) The table below describes a random variable X. **Find the expected value μ and standard deviation σ of X.** (5p)

x	-2	-1	0	1	2
$P(X = x)$	0.2	0.3	0.2	0.1	0.2

(A) $\mu = -0.2\,;\, \sigma = 1.2$

(B) $\mu = -0.2\,;\, \sigma = 1.4$

(C) $\mu = -0.4;\, \sigma = 1.4$

(D) $\mu = -0.2\,;\, \sigma = 1.96$

(E) $\mu = -0.4\,;\, \sigma = 1.96$

Problem 3

a) As part of a survey, people in an i.i.d. random sample of 20 Swedes are asked "Did you watch the television program *Eurovision Song Contest* last year?" Suppose that the true proportion of Swedes who did watch the contest last year is 25% and that each person answers truthfully. **What is the probability that 8 or more people (out of the 20 in the sample) answer "yes"?** Choose the alternative closest to your answer. (5p)

(A) 1%

(B) 3%

(C) 4%

(D) 10%

(E) 96%

b) Two students, Anna and Bea, take an exam. Suppose that their respective scores could be modelled with two normally distributed random variables, X and Y:

Anna: $X \sim N(70, 6^2)$
Bea: $Y \sim N(60, 8^2)$

Assume that X and Y are independent. **Find the probability that Anna scores higher than Bea, according to this model.** Choose the alternative closest to your answer. (5p)

(A) 16%

(B) 24%

(C) 76%

(D) 79%

(E) 84%

c) During one month, the E-commerce website *Pebbles 'R 'Us* gets 600 unique visits. Each unique visit has a 2% probability of resulting in a sale. Assume that the visits are statistically independent and that each visit results in at most one sale, that month. **Find the approximate probability that these 600 visits result in 20 or more sales.** Choose the alternative closest to your answer. Hint: a sale could be viewed as a "success." (5p)

(A) 0%

(B) 1%

(C) 5%

(D) 10%

(E) 25%

Problem 4

a) In an i.i.d. sample of 900 German school children, 90 children are found to be left-handed. **Based on this sample, which of the following is a 90% confidence interval for the proportion of left-handed schoolchildren in Germany?** Choose the alternative closest to your answer. (5p)

(A) (0.0804, 0.1196)

(B) (0.0836, 0.1164)

(C) (0.0941, 0.1059)

(D) (0.0951, 0.1049)

(E) (0.0998, 0.1002)

b) A nutritional scientist compares the kilo calorie (kcal) content of two popular vegan energy bars, *Raw* and *Omega* (she suspect that the values given on the packaging are false). She collects a random sample of 10 energy bars each of the two relevant brands and measures the kilo calorie contents. Assume that the true kcal population variances for the two brands are the same and that the kcal contents are normally distributed. **Based on the scientist's table below, create a 95% confidence interval for the difference in mean kcal,** $\mu_{Raw} - \mu_{Omega}$. Choose the alternative closest to your answer. (5p)

Brand	n	$\bar{x}$	s^2
Raw	10	200	62.0
Omega	10	190	66.0

(A) (1.907, 18.093)

(B) (2.483, 17.517)

(C) (2.988, 17.012)

(D) (4.397, 15.603)

(E) (4.685, 15.315)

c) **Which of the following statements is correct?** (5p)

(A) A statistically significant result is synonymous with a p-value larger than α.

(B) The width of a confidence interval is also called the *margin of error*.

(C) In order to reduce the standard error of an estimator by 50%, a sample size twice as large is typically needed.

(D) If the sample size is sufficiently large and the population variance is finite, we can use the central limit theorem to say that the sample mean will be approximately normally distributed.

(E) The expected value of a *biased* estimator is equal to the parameter it is estimating.

Problem 5

A candy company produces chocolate dipped peanuts. Each peanut has a colorful outer shell and there are $K = 4$ possible colors. The company claims that each color has the same probability of occurring, e.g. if a bag contains certain amount of peanuts, the expected frequency of each color is the same. You randomly select a bag containing $n = 200$ peanuts and count the colors: 45 are red, 40 are green, 60 are yellow, and 55 are brown. You decide to use a χ^2-test to determine whether each color is equally likely, or if the probabilities of the colors differ at a 5% significance level.

a) **What is the decision rule and what is the null hypothesis?** (5p)

(A) Reject H_0 if $\chi^2_{obs} > 3.841$; H_0: each color has probability 25%

(B) Reject H_0 if $\chi^2_{obs} > 5.991$; H_0: each color has probability 25%

(C) Reject H_0 if $\chi^2_{obs} > 5.991$; H_0: the probabilities are <u>not</u> all 25%

(D) Reject H_0 if $\chi^2_{obs} > 7.815$; H_0: each color has probability 25%

(E) Reject H_0 if $\chi^2_{obs} > 7.815$; H_0: the probabilities are <u>not</u> all 25%

b) **What is the value of the test variable?** Choose the value closest to your answer. (5p)

(A) 0.0115

(B) 3.31

(C) 5.00

(D) 7.93

(E) 8.44

Problem 6

In a sleep study a group of scientists recorded the number hours slept by eight randomly selected (and consenting) patients during one night. A week later, the same eight patients were given a sleep-inducing drug before their sleep was recorded once more. The table shows the hours slept by each patient, with and without the drug. Test at the 5% level whether patients sleep longer under the influence of the sleep-inducing drug.

Patient	1	2	3	4	6	7	8
No Drug	8.6	6.9	7.8	7.2	7.6	7.6	7.7
With Drug	9.5	7.6	7.7	8.6	9.2	9.5	8.5

a) State your assumptions, hypotheses, test statistic, critical value and decision rule. (8p)
b) Finish your calculations, state your conclusions and give a verbal interpretation. (6p)
c) Briefly explain the difference between a Type I error and a Type II error in statistics. Give a brief example of a Type I error. You may use the study in this problem to construct your example. (6p)

Problem 7

A book publisher has recently published a new biography. The table below contains data from a random sample of eight bookstores, where y is books sold and x is price in euro. To make your calculations easier, we have provided some additional figures.

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$\hat{y}_i$
1	15	29	225	841	435	38,38596
2	20	38	400	1444	760	29
3	19	34	361	1156	646	30,87719
4	21	39	441	1521	819	27,12281
5	15	43	225	1849	645	38,38596
6	26	10	676	100	260	17,73684
7	19	19	361	361	361	30,87719
8	25	20	625	400	500	19,61404
Summa	160	232	3314	7672	4426	232

a) Write down the regression model with price (X) as independent variable and number of books sold (Y) as dependent variable. Estimate the model parameters with the least square method. (6p)

b) Calculate the residual variance and the coefficient of determination. Interpret the coefficient of determination. (6p)

c) Test formally if the slope coefficient is significantly different from zero at 5% level of significance. State your hypotheses, test variable, decision rule, calculations and conclusion. (8p)

Department of Statistics

Stockholms universitet

Correction sheet

Date: 13/02/2020

Room: Värtasalen

Exam: Statistics for Economists (eng)

Course: Basic Statistics for Economists

Anonymous code: 0112-GDO

☒ 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			3
Teacher's notes									

Points	Grade	Teacher's sign.
81	B	

STOCKHOLM UNIVERSITY
Department of Statistics

Autumn 2019 C-D

ANSWER FORM Exam – Basic Statistics for Economists
2020-02-13

Room: VÄ

Anonymous code: 0112-GDO *(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
1a	☐	☐	☐	☒	☐
1b	☐	☐	☒	☐	☐
2a	☐	☒	☐	☐	☐
2b	☐	☐	☐	☒	☐
3a	☐	☐	☐	☒	☐
3b	☐	☐	☐	☐	☒
3c	☐	☒	☐	☐	☐
4a	☐	☒	☐	☐	☐
4b	☐	☒	☐	☐	☐
4c	☐	☐	☒	☐	☐
5a	☐	☐	☐	☒	☐
5b	☐	☐	☒	☐	☐

50/60

SU, DEPARTMENT OF STATISTICS

Room: VA Anonymous code: 0112-GDO Sheet number: 1

N 6

a) Assumptions: iid; X and Y are dependent because the same people are being tested, normal distribution.

Hypotheses: $H_0: \mu_D = \mu_X - \mu_Y = 0$

$H_1: \mu_D = \mu_X - \mu_Y > 0$

Test Statistic:

$$t_{n-1} = \frac{\bar{d} - \mu_0}{S_d / \sqrt{n}}$$

Critical value: $t_{n-1;\alpha} = t_{7-1;0,05} = 1,943$

Decision rule: reject H_0 if $t_{obs} > 1,943$ /8

b) $\bar{d} = \frac{1}{n} \cdot \sum_{i=1}^{7} (x_i - y_i) = \frac{(9,5-8,6) + (7,6-6,9) + \dots + (8,5-7,7)}{7-1}$ no

$= 1,2$ ~~1,02~~

$$S_d^2 = \frac{\sum_{i=1}^{7} ((x_i - y_i) - \bar{d})^2}{n-1} = \frac{(0,9-1,2)^2 + (0,7-1,2)^2 + \dots + (0,8-1,2)^2}{7-1}$$

$= 0,48$

$d \sim N(1,2 ; 0,48)$

$$t_{obs} = \frac{1,2 - 0}{\sqrt{\frac{0,48}{7}}} = 4,5356$$

$t_{obs} = 4,5356 > t_{crit} = 1,943 \Rightarrow$ we reject H_0, the drug does increas the sleeping time on average

/4

null

Hypotheses

c)

Decision	Right	Wrong
Accept	correct	Type I α
Reject	Type II β	$1-\beta$

Type I error occurs when we fail to reject a wrong null hypotheses, the probability of such an error occuring is α.
For example, if the true mean of some study is, for example 10, but we accept a hypotheses that the mean is equal to, for example 7, we commit a Type I error.
A type II error occurs when we wrongfully reject a true hypotheses, the probability of such an error occuring is β, while $(1-\beta)$ is called the strength of the test which denotes what is the probability of us rejecting a wrong hypotheses.
Example: going back to the previous example, if we reject a true hypotheses which does state that the mean is 10, we will make a type II error, by denying a true mean.

4

SU, DEPARTMENT OF STATISTICS

Room: VÄ Anonymous code: 0112-GDO Sheet number: 2

N 7

a) Model: $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_1$

$$\beta_1 = \frac{Cov(x,y)}{S_x^2}$$

$$\bar{x} = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{15+20+19+\dots+25}{8} = 20$$

$$\bar{y} = \frac{\sum_{i=1}^{8} y_i}{n} = \frac{29+38+34+\dots+20}{8} = 29$$

$$S_x^2 = \frac{\sum_{i=1}^{8} x_i - n\bar{x}^2}{n-1} = \frac{15^2+20^2+19^2+\dots+25^2-8\cdot 20^2}{8\cdot 1} = 16.2857$$

$$Cov(x,y) = \frac{\sum_{i=1}^{8} x_i \cdot y_i - n\cdot\bar{x}\cdot\bar{y}}{n-1} = \frac{15\cdot 29+20\cdot 38+\dots+25\cdot 20-8\cdot 20\cdot 29}{8-1}$$

$$= -30,5714$$

$$\beta_1 = \frac{-30,5714}{16.2857} = -1,8772$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x} = 29 - (-1,8772 \cdot 20) = 66.544.$$

b) $e_i = y_i - \hat{y}_i$

i	1	2	3	4	5	6	7	8
e_i	-9,3859	9	3,12281	11,87719	4,61404	-7,7368	-11,87719	0,38596

$$\bar{e} = \frac{\sum e_i}{n-1} = 1,42 \cdot 10^{-6} \approx 0$$

6

$$S_e^2 = \frac{\sum e_i^2 - n \cdot \bar{e}^2}{n-1} = \frac{542,2804 - 0}{8-1} = 77.545$$

Note: I assume that $n \cdot \bar{e}^2 = 0$, because the actual number is very small

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$SST = \sum_{i=1}^{8} (y_i - \bar{y})^2 = (29-29)^2 + (38-29)^2 + \dots + (20-29)^2 = 944$$

$$SSE = \sum_{i=1}^{8} e_i^2 = (-9.3859)^2 + (9)^2 + \dots + (0,385936)^2 = 542.28$$

$$R^2 = 1 - \frac{542.28}{944} = 0,42555$$

Interpretation: $R^2 = 0,42555$ means that the model explains $\approx 42.555\%$ of why y_i takes on a certain value. Thus, the book price only explains $\approx 42.555\%$, while the rest is influenced by other independent variables outside of this model.

/6

c) Hypotheses: $H_0: \beta_0 = 0$

$H_1: \beta_0 > 0$

Test variable: $t_{n-k-1} = \frac{b_j - \beta_j}{S_{b_j}}$

Decision rule: reject H_0 if $t_{obs} > t_{crit} = 1,943$

$t_{crit} = t_{8-1-1;\,0,05} = 1,943$

$$S_{b_0}^2 = S_e^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1) S_x^2} \right) = 77.545 \cdot \left(\frac{1}{8} + \frac{20^2}{7 \cdot 16.2857} \right)$$

$= 281.781 \Rightarrow 281.781 = t_{obs} > 1,943 \Rightarrow$ we reject H_0, β_0 is not significant from 0

/3