

**EXAM – BASIC STATISTICS FOR ECONOMISTS**  
2022-01-14

---

<b>Time:</b>	08.00 - 13.00
<b>Approved aid:</b>	Hand-held calculator with no stored text, data or formulas
<b>Provided aid:</b>	<i>Formula Sheet and Probability Distribution Tables</i> , 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 be 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.

- Solutions will be posted on Athena shortly after the exam.

**GOOD LUCK!**

### Problem 1

Twelve students from the same school took the *Swedish Scholastic Aptitude Test (Högskoleprovet)*. The results on this test are normalized to a scale ranging from 0.0 to 2.0 Their scores (ordered lowest to highest) can be found in the table below:

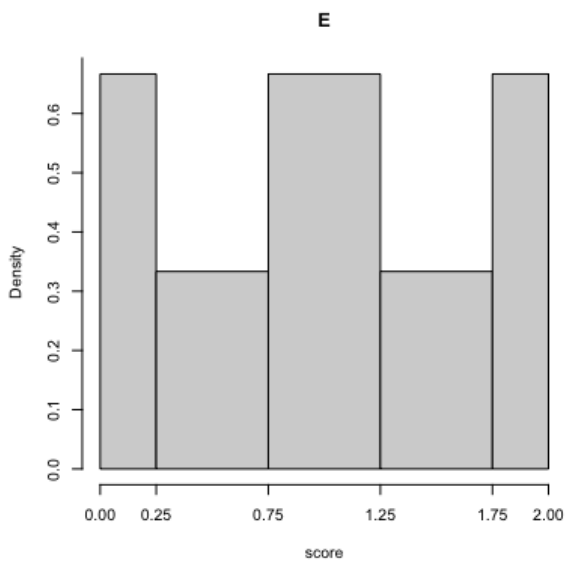
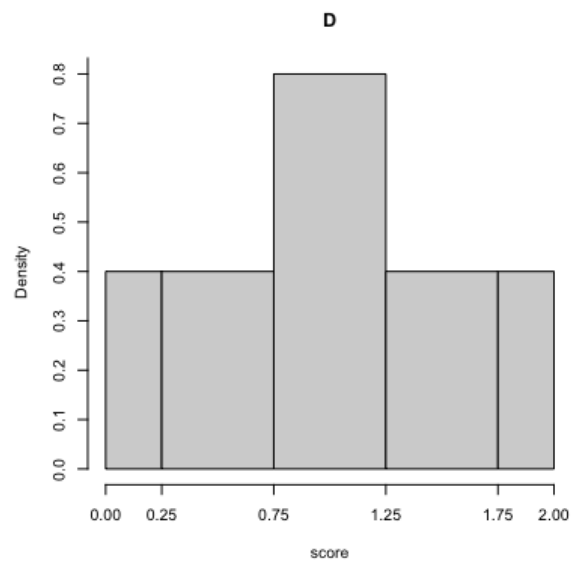
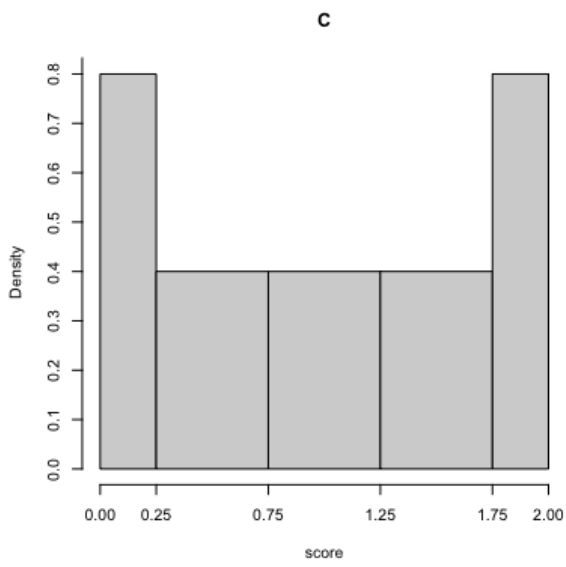
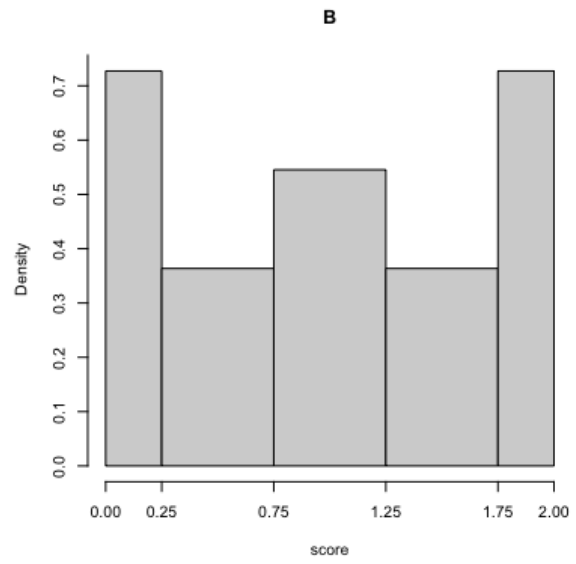
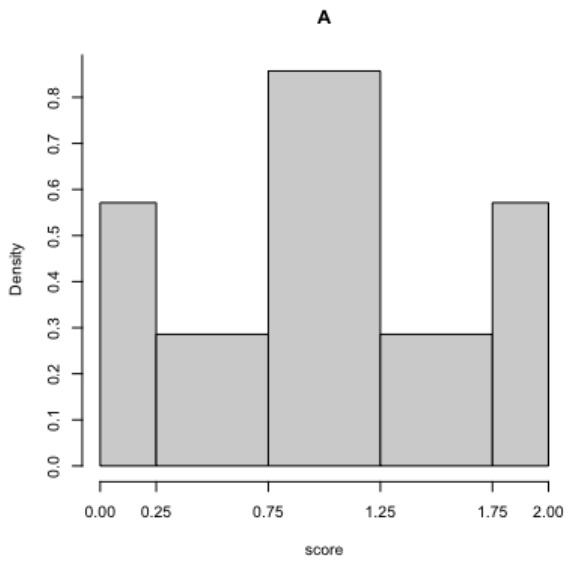
<b>Student</b>	1	2	3	4	5	6	7	8	9	10	11	12
<b>Score</b>	0.1	0.1	0.3	0.7	0.8	0.9	0.9	1.1	1.5	1.5	1.8	2.0

a. **Find the Inter Quartile Range of the twelve scores.** Choose the alternative closest to your answer. (5p)

- A) 0.8
- B) 0.9
- C) 1.1
- D) 1.2
- E) 1.3

b. On the next page, you can find five histograms marked A-E. **Find the histogram that correctly represents the twelve scores.** (5p)

- A) A
- B) B
- C) C
- D) D
- E) E



## Problem 2

a. A class of fifth-graders have 26 students; 13 are girls and 13 are boys. Three students from the class are chosen at random to help out in the lunch room. **Find the probability that the group of three contains both boys and girls.** (5p) Choose the alternative closest to your answer.

- A) 0.72
- B) 0.75
- C) 0.78
- D) 0.81
- E) 0.84

b. According to experts, the probability that a certain ice-hockey team wins their next match is 50%. The probability that the team wins the match after that is also 50%, and the probability that the team loses both matches is 30%. There will be overtime and penalties, if necessary, to ensure that there are no ties. **What is the probability that the team wins exactly one match?** (5p) Choose the alternative closest to your answer.

- A) 20%
- B) 25%
- C) 30%
- D) 40%
- E) 50%

### Problem 3

a. A factory produces electrical components. It is known that 5% of all components produced have some defect. A manager draws an i.i.d. sample of 10 components. **What is the probability that at least two have some defect?** (5p) Choose the alternative closest to your answer.

- A) 0.07
- B) 0.09
- C) 0.11
- D) 0.13
- E) 0.15

b. A financial analyst uses the following model for the yearly return (change) of two stocks, *Magnavox* and *Pong Inc.* If  $X$  is the percentage return of *Magnavox* and  $Y$  is the percentage return of *Pong Inc.*, then

$$X \sim N(15, 25^2)$$

$$Y \sim N(10, 20^2)$$

$$\rho_{X,Y} = 0.7$$

**Find the probability that  $X$  yields higher return than  $Y$  (over one year).** (5p) Choose the alternative closest to your answer. *Hint: Can you find the covariance between  $X$  and  $Y$ ?*

- A) 0.44
- B) 0.55
- C) 0.56
- D) 0.61
- E) 0.65

c.

Tom and Kelly are both fighter pilots in the Swedish air force. They would like their baby daughter to be a fighter pilot too. They know the air force has a height requirement: to be a pilot, you have to be between 160 cm and 190 cm tall. Taking into account their own heights, they consider that the height of her daughter as an adult can be modeled as the outcome of a normal distribution with mean 170 cm and standard deviation 7 cm. **Find the probability that their daughter will be between 160 cm and 190 cm tall, according to the parents' model.** (5p) Choose the alternative closest to your answer.

- A) 90%
- B) 92%
- C) 94%
- D) 96%
- E) 98%

#### Problem 4

a. A biologist collects a i.i.d. random sample of 10 apples from an apple orchard. She measures and weighs each apple. The mean weight of the 10 apples is 185 grams and the sample standard deviation is 35 grams. Assume that the apples' weights follow a normal distribution. **Find a 90% confidence interval for the mean weight of an apple from the orchard.** (5p) Choose the alternative closest to your answer.

- A) (160, 210)
- B) (164, 205)
- C) (167, 203)
- D) (172, 198)
- E) (179, 191)

b. In a medical study, scientists compared the efficiency of two nicotine patches of two different strengths, 22 mg and 44 mg. Nicotine patches are used to help tobacco smokers to not smoke. A sample of 100 volunteer subjects were divided into two groups: 50 were given the 22 mg patch and 50 were given the 44 mg patch. After six months, the subjects were asked if they had smoked since the study started or if they had stopped completely. The result can be found in the table below:

Group	44 mg	22 mg
n	50	50
Stopped smoking	16	11

**Find a 95% confidence interval for the difference in proportion between the two groups (smokers using the 44 mg patch and smokers using the 22 mg patch).** (5p) Choose the alternative closest to your answer.

- A) (-0.07, 0.27)
- B) (-0.04, 0.24)
- C) (-0.02, 0.22)
- D) (0.01, 0.19)
- E) (0.02, 0.18)

c) Assume that the natural lifespan of a lab mouse kept in a cage has a known population standard deviation of 80 days and that the lifespan is normally distributed. A researcher wants to estimate the population lifespan of such a mouse, with a 95% confidence interval.

**Find the minimum sample size that the researcher needs to use to guarantee that the margin of error is at most 5 days.** (5p) Choose the alternative closest to your answer.

- A) 40
- B) 246
- C) 385
- D) 984
- E) 1537

### Problem 5

The marketing department of an online gaming site creates two alternative two-minute videos to market their latest game. They call the videos *Version A* and *Version B*. They design the website so that visitors to the site are randomly shown either *A* or *B*. They then record the number of visitors and the number of downloads, for each version.

Version	Downloads	Total Visitors per Version
A	120	995
B	105	1005

Test at the 5% level whether the proportion of downloads per visitor is greater for *Version A* than for *Version B*. Call the test variable  $z_{obs}$ .

a. Find the decision rule of the test. (5p)

- A) Reject  $H_0$  if  $z_{obs} < 1.6449$
- B) Reject  $H_0$  if  $|z_{obs}| < 1.96$
- C) Reject  $H_0$  if  $z_{obs} > 1.6449$
- D) Reject  $H_0$  if  $|z_{obs}| > 1.96$
- E) Reject  $H_0$  if  $|z_{obs}| > 1.6449$

b. Find the value of the test variable. (5p) Chose the alternative closest to your answer.

- A) 0.83
- B) 0.98
- C) 1.07
- D) 1.14
- E) 1.28

## Problem 6

A Swedish sociologist studying discrimination conducts an experiment. She sends out 200 job applications to 200 different advertised entry-level jobs in Sweden. In half of the 200 job applications, she uses a fictive Swedish name (like Karl Andersson) and in the other half, she uses a fictive foreign sounding first name (like Shady Gamhour). Otherwise, she takes great care to make all applications as similar as possible to each other (same CV, same personal letter, same time sent relative to the deadline, and so on). The results can be found in the table below:

	<b>Interview</b>	<b>No interview</b>
<b>Swedish name</b>	36	64
<b>Foreign name</b>	25	75

**Test at the 1% level whether interview/no interview is independent of type of name used.**

- a. State hypotheses and the test variable. (5p)**
- b. State the critical value and decision rule. (5p)**
- c. Calculate the test statistic and draw conclusion. (5p)**
- d. Briefly explain what a p-value is for this test. You do not have to estimate the p-value. (5p)**  
(Tip: you can draw a picture)

*Note: this problem is inspired by scientific studies of discrimination by the late sociologist Devah Pager, but the context and numbers here are made up. Similar studies have been conducted in Sweden. See for example*

Bursell, M., Bygren, M., & Gähler, M. (2021). Does employer discrimination contribute to the subordinate labor market inclusion of individuals of a foreign background? *Social Science Research*, 98, 102582. <https://doi.org/10.1016/j.ssresearch.2021.102582>



## Problem 7

A student wants to examine the relationship between median income and median rent, in U.S. states. To make calculations easier, we have limited the data to a sample of 11 states. You can find data, plus the products of the variable *rent* and *income* in the table below:

State	income	rent	population	income*rent
California	29	1360	39.8	39440
Connecticut	35	1120	3.6	39200
Vermont	29	940	0.6	27260
Wisconsin	30	810	5.8	24300
Virginia	33	1170	8.5	38610
Delaware	32	1080	1	34560
Texas	28	950	28.7	26600
Oregon	27	990	4.2	26730
Kansas	29	800	2.9	23200
Alaska	33	1200	0.7	39600
Alabama	24	750	4.9	18000
<b>Sum</b>	<b>329</b>	<b>11170</b>	<b>100.7</b>	<b>337500</b>

*income* the median income in the state, in thousands of dollars per year

*rent* the median rent in the state, in dollars per month

*population* the population of the state, in millions

For part a and b, you should consider Model 1:

$$rent = \beta_0 + \beta_1 \cdot income + \varepsilon$$

**a. Find the variance of *income* and the covariance between *income* and *rent*. (5p)**

**b. Estimate the coefficients of Model 1. Clearly state the estimated model. (5p)**

For part c and d, you should consider Model 2:

$$rent = \beta_0 + \beta_1 \cdot income + \beta_2 \cdot population + \varepsilon$$

Part of the output from model 2 can be found on the next page.

**c. Use a formal test to test whether Model 2 is better than Model 1 (i.e., test whether *population* should be included in the model). (5p)**

**d. Briefly explain what a residual is. Explain why it is a good idea to plot the residuals of your linear regression model. (5p)**

**Model 2**

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>
Regression	2	231725.89	115862.95
Residual	8	129746.84	16218.35
Total	10	361472.73	

	<i>Coefficients</i>	<i>Standard Error</i>
Intercept	-315.40	401.02
income	41.90	13.10
population	8.48	3.20

--- END OF EXAM ---