STOCKHOLM UNIVERSITY
Department of Statistics
Fall 2020

Ulf Högnäs (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2021-02-17

Time:	13.00-19.00 (including one extra hour to arrange the electronic submission).
Approved aid:	Any books, notes, or digital resources. You are not allowed to communicate with anyone during the exam. This includes chats, messages, and internet forums.

- **Problems 1 – 5 MULTIPLE CHOICE QUESTIONS – max 50 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 or on one page. If you prefer, you can make a handwritten version, but please make it clear.
  - Mark exactly one answer and do not provide written solution.

- **Problems 6 – 7**: **COMPLETE WRITTEN SOLUTIONS – max 50 points**
  - 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 50 + 50 = 100. At least 50 points is required to pass (grades A-E).

  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

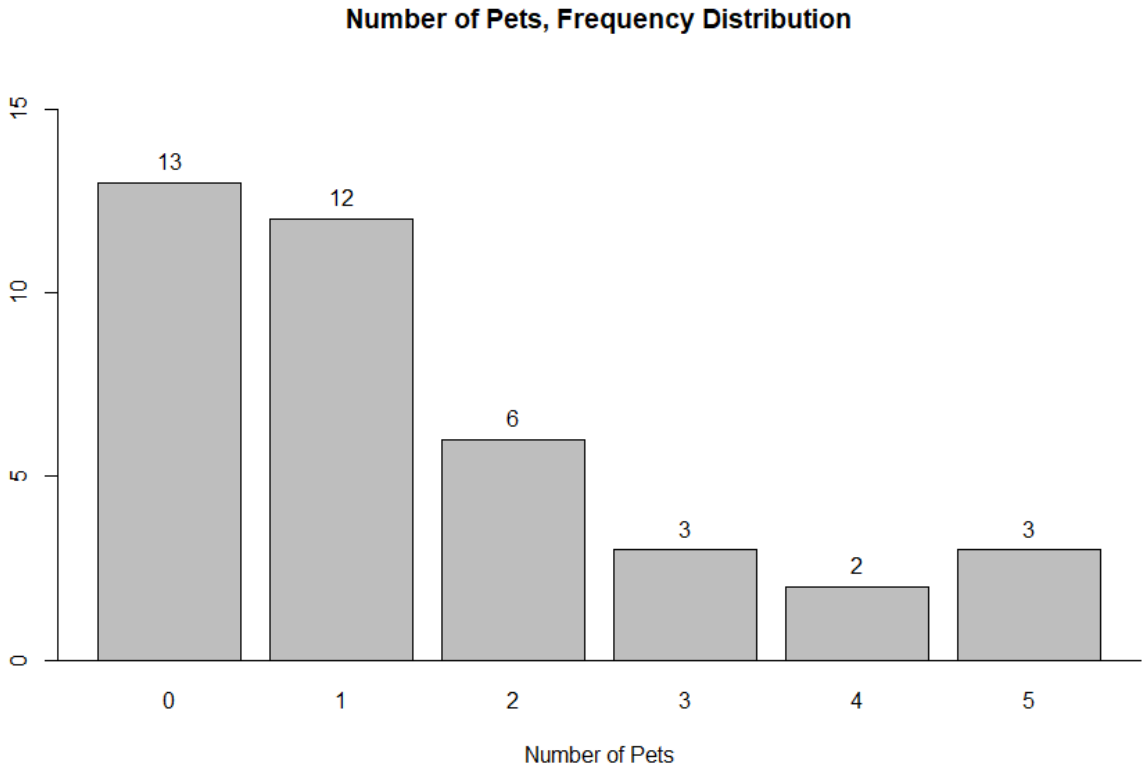
  NOTE! Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement for a higher grade.

  GOOD LUCK!

Answer form for multiple choice. You can make your own form, put please be clear and answer on one page. Do <u>not</u> submit solutions to the multiple-choice problems.

Number	Part	A	B	C	D	E
1	**a.**					
1	**b.**					
2	**a.**					
2	**b.**					
3	**a.**					
3	**b.**					
4	**a.**					
4	**b.**					
5	**a.**					
5	**b.**					

PROBLEM 1.



The diagram above shows the frequency distribution of the number of pets, for 39 households. For example, 13 households have zero pets. To make the diagram easier to read, the number above each bar shows the frequency.

a. **Find the inter quartile range (IQR) of the number of pets.** (5p)
   (A) 1
   (B) 2
   (C) 2.5
   (D) 3
   (E) 7.5

The table below shows the number of pets and the number of cars, for a sample of 5 of households:

Household	**1**	**2**	**3**	**4**	**5**
Cars	2	0	1	0	1
Pets	4	2	1	0	2

b. **Find the sample correlation between the number cars and the number of pets.** Choose the alternative closest to your answer. (5p)

(A) -0.2
(B) 0
(C) 0.54
(D) 0.77
(E) 0.95

PROBLEM 2

A blood test for the disease *Lycanthropy* is not 100% reliable. A positive test is supposed to indicate that a patient is infected, while a negative test is supposed to indicate that the patient is not infected. The probability that the test is positive given that the patient is infected is 90%. The probability that the test is negative given that the patient is not infected is also 90%. Suppose that 5% of the population is infected.

a. **Find the probability that a randomly selected patient member of the population tests negative (correctly or incorrectly).** Choose the alternative closest to your answer. (5p)
   (A) 0.78
   (B) 0.80
   (C) 0.82
   (D) 0.84
   (E) 0.86

A randomly selected patient is tested and the test is negative.

b. **Find the probability that the patient is infected by the disease, despite the negative test result.** Choose the alternative closest to your answer. (5p)
   (A) 0.006
   (B) 0.01
   (C) 0.05
   (D) 0.09
   (E) 0.11

(Note: If you find part b of this problem difficult, you might want to save it until last. Also, Lycanthropy is a fictional disease.)

PROBLEM 3

A biologist estimates that the weight of a randomly chosen male red fox is normally distributed with mean 8 kg and standard deviation 2 kg**.**

a. **Find the probability that a randomly chosen male fox weighs more than 11kg, according to the biologist's model.** Choose the alternative closest to your answer. (5p)
   (A) 0.012
   (B) 0.017
   (C) 0.025
   (D) 0.038
   (E) 0.067

John and Jane invite 11 other couples to their New Year's Eve party. Assume that each of the 11 couples will answer "yes, we are coming to the party" with 80% probability and that each couple will answer independently of the other couples. If more than 8 of the couples answer "yes," there will not be enough space at the dinner table.

b. **Find the probability that at most 8 of the 11 couples answer "yes."** Choose the alternative closest to your answer. (5p)
   (A) 17%
   (B) 38%
   (C) 62%
   (D) 84%
   (E) 99%

PROBLEM 4.

A statistics student wants to estimate the mean age of business students. She collects a simple random sample of 30 business students from Stockholm University and a sample of 40 business students from Uppsala University.

	Sample mean	Standard deviation	n
Stockholm	22.93	4.95	30
Uppsala	23.52	4.56	40

a. **Find a 90% confidence interval for the difference in mean age between the two populations of students (Stockholm minus Uppsala).** Choose the alternative closest to your answer. (5p)
   (A) (-1.46, 0.28)
   (B) (-2.29, 1.11)
   (C) (-2.49, 1.31)
   (D) (-2.86, 1.68)
   (E) (-3.26, 2.08)

An analyst at a marketing firm wants to survey Swedes about their opinions of the company *Legendary Nuts*, a big client of the marketing firm. The survey will include the question "do you have a favorable view of *Legendary Nuts?*" The analyst wants to find a 95% confidence interval for the proportion of Swedes who would answer "yes."

b. **What sample size is needed to guarantee a margin of error of at most 3% (3 percentage points)?** Assume 100% response rate. (5p)
   (A) 16
   (B) 267
   (C) 752
   (D) 1068
   (E) 2042

Hint: what proportion of "yes" answers should we assume when solving this problem?

PROBLEM 5

A shop owner sells a t-shirt with the logo of her store printed on it. She has three sizes: Small, Medium and Large. Previous years, she has sold the following distribution of sizes:

Type	Small	Medium	Large
Sales (%)	20%	50%	30%

The sales figures for last year are in. She sold a total of 200 t-shirts and you can find the sales for each size below:

Type	Small	Medium	Large
Sales, count	39	108	53

Treat these sales figures as an independent identically distributed sample from the local population. Test at the 5% level of significance whether the distribution of sizes have changed compared to previous years.

a. **Find the critical value.** (5p)
   (A) 1.6449
   (B) 1.653
   (C) 1.96
   (D) 5.991
   (E) 7.815

b. **Find the value of the test variable.** Choose the alternative closest to your answer (5p)
   (A) 1.48
   (B) 2.39
   (C) 5.27
   (D) 12.03
   (E) 16.31

PROBLEM 6

A company offers a preparation course for the Swedish scholastic aptitude test *Högskoleprovet*. As part of an assessment of the quality of their own course, they offered free courses to a random sample of students six students who have taken Högskoleprovet before. All the students completed the course and then attempted the exam again. The test consists of 160 questions (the number of correct answers is later converted to a "normalized score" from 0 to 2.0, but this is not relevant for the problem). The table shows the number of correct answers (score) of the students, before and after completing the course:

Student	**1**	**2**	**3**	**4**	**5**	**6**
Score before	60	78	75	63	95	77
Score after	79	105	87	77	117	93

Assume that the score of each individual student is (approximately) normally distributed, both before and after taking the course**. Test at 5% level of significance whether taking the course is associated with an increase (score after minus score before) in average score of at least 10.**

a. **State the hypotheses, test variable, critical value and decision rule.** (5p)
b. **Calculate the test variable and interpret the outcome of the test.** (5p)
c. **Use the formula sheet to find an approximate p-value of the test. Briefly explain why we can use the p-value to determine to outcome of the test.** (5p)

For part d and part e, assume that the true average (population) increase is 10 and that the population variance is 100. The company decides to conduct another study with six new students; otherwise, the study is identical to the first.

d. **Find the probability that a randomly chosen student does not improve their score after taking part in the study.** (5p)
e. **Find the probability that none of the six students improve their score after taking part in the study.** (5p)

PROBLEM 7

An employee at a diamond wholesaler collects a random sample of 100 round cut diamonds from the vault. He notes the carat, color, clarity, and quality of cut. Last, he notes the estimated retail price, determined by the experts at the firm.

PRICE: Estimated price in USD

CARAT: weight in carats (1 carat = 200 milligram)

COLOR: Quality of the color, dummy variable where 1= "best or second-best color"

CLARITY: dummy variable where 1= "best or second-best clarity" (VVS1 or VVS2)

CUT: dummy variable where 1= "Premium or Ideal cut" (and 0="lesser cut")

He estimates two linear regression models:

MODEL 1: $PRICE = \beta_0 + \beta_1 CARAT + \beta_2 CLARITY + \beta_3 CUT + \beta_4 COLOR + \varepsilon$

MODEL 2: $PRICE = \beta_0 + \beta_1 CARAT + \beta_2 CLARITY + \beta_3 CUT + \varepsilon$

The output of each model can be found below.

a. **Find the estimated price of a diamond according to model 2, given that the diamond weighs 1 carat, has a clarity rating that is <u>less good</u> than second best, and a "premium" cut.** (5p)

For part b and part c, you are asked to test at the <u>1% level of significance</u> whether $\beta_2 > 500$, given that CARAT and CUT is included in the model.

b. **Clearly state hypotheses, test variable, critical value and decision rule.** (5p)
c. **Use the output from MODEL 2 to calculate the test variable. Clearly state the conclusion and interpretation of the test result.** (5p)
d. **Find the adjusted coefficient of termination for both models. Interpret the results and the difference in results.** (5p)
e. **Note that the coefficient for COLOR is negative, even though a "1" means "best or second-best color." Use your statistics knowledge to explain how this is possible.** (5p)

MODEL 1

Regression Statistics	
Multiple R	[redacted]
R Square	
Adjusted R Square	
Standard Error	1085.969629
Observations	100

ANOVA

	df	*SS*	*MS*	*F*
Regression	4	966730149.7	2.42E+08	204.9321
Residual	95	112036353.3	1179330	
Total	99	1078766503		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-2576.468	324.904	[redacted]	
CARAT	7247.508	262.216		
CLARITY	1080.167	278.370		
CUT	483.917	227.033		
COLOR	-28.244	241.200		

MODEL 2

Regression Statistics	
Multiple R	0.946640919
R Square	[redacted]
Adjusted R Square	
Standard Error	1080.376688
Observations	100

ANOVA

	df	*SS*	*MS*	*F*
Regression	3	966713979.4	3.22E+08	276.0745
Residual	96	112052523.5	1167214	
Total	99	1078766503		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-2592.874	291.632	[redacted]	
CARAT	7253.921	255.112		
CLARITY	1086.982	270.817		
CUT	485.726	225.340		

END OF EXAM