STOCKHOLM UNIVERSITY
Department of Statistics
Spring 2021

Ulf Högnäs (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2021-08-17

Time: 09.00-15.00 (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 10 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
  - Do not provide written solution for these problems.
- **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.

  Follow the instructions carefully when you submit your answers. The instructions were sent out in a separate email.

  NOTE! If the course coordinator needs to send out information to all students during the exam, this is done to your registered email address. Therefore, check your email during the exam.

  GOOD LUCK!

All students get the same seven problems, regardless of anonymity code.

Answer form for multiple choice. You can make your own form, put please be clear and answer on <u>one</u> 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.					

1. Twenty high school students were asked how many hours they worked on a particular homework assignment. The answers are summarized in the table below:

Hours Worked	0	1	2	3	4
Frequency	4	12	1	1	2

a) **Find the inter quartile range of the number of hours worked.** (5p)

(A) 0
(B) 1
(C) 1.5
(D) 2
(E) 2.5

b) **Find the sample standard deviation of the number hours studied.** Choose the alternative closest to your answer. (5p)

(A) 1.16
(B) 1.36
(C) 1.58
(D) 3.12
(E) 4.63

2. A factory produces radio-controlled toy cars. The probability that randomly selected car has a visible flaw is 0.02. The probability that a randomly selected car does not function properly is 0.01. The probability that the car has a visible flaw and does not function properly is 0.002. The shift manager selects a finished car at random.

a) **If the car has a visible flaw or does not function properly (or both), the car will be discarded. Find the probability that the car will be discarded.** Choose the alternative closest to your answer. (5p)

(A) 0.026
(B) 0.028
(C) 0.030
(D) 0.032
(E) 0.034

b) **Find the probability that a randomly chosen car does not function properly, given that it has a visible flaw.** Choose the alternative closest to your answer. (5p)

(A) 0.01
(B) 0.06
(C) 0.10
(D) 0.18
(E) 0.20

3. Max has a gambling problem. One day, he buys eleven scratch cards at one time (a scratch card is a type of lottery ticket). Each card is a winning ticket with probability 0.25. Since the total number of issued tickets is very large, you may assume that the probability of a win for each ticket is independent of the other tickets.

a) **Find the probability that at least four of the eleven scratch cards are winning tickets.** Choose the alternative closest to your answer. (5p)

(A) 0.034
(B) 0.050
(C) 0.11
(D) 0.17
(E) 0.29

Serhiy and William run a 200 meter race. Assume that Serhiy's race time is normally distributed with mean 25 seconds and standard deviation 1.5 seconds. Assume that William's race time is normally distributed with mean 26 seconds and standard deviation 2 seconds. During this race, the correlation between the two race times is 0.2.

b) **Find the probability that Serhiy wins the race against William.** Choose the alternative closest to your answer. (5p)

(A) 0.33
(B) 0.42
(C) 0.58
(D) 0.67
(E) 0.75

4. A random sample of 10 men in New Jersey and 10 men in Arizona revealed the following average heights and standard deviations for the two samples. Assume that height is normally distributed and that the population variances are equal.

	New Jersey	**Arizona**
Mean height (cm)	177.4	176.2
Standard deviation (cm)	5.1	5.1
n	10	10

a) **Find a 99% confidence interval for the difference in mean male height between the two states, $\mu_{NJ} - \mu_{AZ}$.** Choose the alternative closest to your answer. (5p)

(A) (-3.44, 5.84)
(B) (-3.91, 6.31)
(C) (-4.67, 7.07)
(D) (-5.36, 7.76)
(E) (-6.02, 8.42)

In a survey of 1000 randomly selected British adults, 445 answered that they have given money to charitable organizations in the past 12 months.

b) **Assuming that the sample is representative of the whole population, find the margin of error for the proportion of British adults who have given money to charitable organizations in the past 12 months. Use a 95% confidence interval.** Choose the alternative closest to your answer. (5p)

(A) 1.6%
(B) 2.1%
(C) 3.1%
(D) 5.3%
(E) 6.2%

5. When *Bank of Tuvalu* issues debit cards to their customers, they let the customer choose the color of the card, either green or blue. Out of the first 160 customers, 72 were female and 88 were male. Out of the 72 female customers, 48 chose blue and the rest chose green. Out of the 88 male customers 66 chose blue and the rest chose green.

Treat the 160 first customers as a simple random sample. Test at 1% level of significance whether gender (male/female) is independent of card color choice.

a) **Find the critical value of the test.** Choose the alternative closest to your answer. (5p)

(A) 3.841
(B) 5.991
(C) 6.635
(D) 9.210
(E) 13.277

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

(A) 1.12
(B) 1.34
(C) 1.46
(D) 1.58
(E) 1.70

6. A group of university students collect a random sample of 10 cans of *White Lightning Cola* from stores around Sweden. According to the packaging, each container is supposed to contain 330 ml of soda. The amount of soda, measured in milliliters, in each of the 10 cans can be found in the table below:

i	1	2	3	4	5	6	7	8	9	10
volume	331	329	328	328	328	329	331	329	328	329

a) **Find the sample mean and sample variance of the sample.** Remember to show your work! (5p)

Test at a 10% level of significance whether the mean content of a can of White Lightning Cola is less than 330 ml. Assume that the amount of soda in a randomly selected can is normally distributed.

b) **State necessary assumptions, hypotheses and the test variable.** (5p)

c) **State the critical value and decision rule.** (5p)

d) **Calculate the test statistic and draw conclusion.** (5p)

Most of the students in the group claim that the taste of White Lightning Cola is identical to the taste of Coca-Cola. Jane, one of the students, claims that she can tell the difference. The other students pour eight glasses of soda in unmarked glasses, four Coca-Cola and four White Lightning Cola, while Jane is not present. Jane is then told to guess which four glasses contain White Lightning Cola. She picks out all four correctly.

e) **Find the probability that someone guesses four out of four correctly, if they cannot tell the difference at all and they just guess.** (5p)

7. Chad, a real estate agent, collects data using a random sample of 100 sold houses in a neighborhood of Seattle, Washington.

PRICE	The final price of the sold house, in thousands of USD
SQFT	The living area of the house, measured in square foot
BEDROOMS	The number of bedrooms
LAT	The latitude, or north-south position, of the house. A higher number means that the house is further north.
LONG	The longitude, or east-west position, of the house. A higher number means that the house is further east.
WATERFRONT	A dummy variable. A one means that the house sits on the waterfront (shore). A zero means that the house does not sit on the waterfront.

Chad estimates three models:

MODEL 1: $PRICE = \beta_0 + \beta_1 * SQFT + \varepsilon$

MODEL 2: $PRICE = \beta_0 + \beta_1 * SQFT + \beta_2 * LONG + \varepsilon$

MODEL 3: $PRICE = \beta_0 + \beta_1 * SQFT + \beta_2 * BEDROOMS + \beta_3 * LAT + \beta_4 * LONG + \beta_5 * WATERFONT + \varepsilon$

You can find part of the output from Model 1 on the following pages. Use the output to solve the problems.

(a) **Find a 95% confidence interval for the variable BEDROOMS in model 3. Interpret the result.** (5p)

For parts (b) and (c), you are asked to test whether the variable LONG is significantly different from zero, given that SQFT is included in the model.

(b) **State the hypotheses, test variable, critical value, and decision rule. (5p)**

(c) **Calculate the value of the test statistic and state your conclusions. (5p)**

For part (d), you will need the following information about the variable SQFT:

SQFT	
mean	1650
variance	552284

d) **Use model 1 to find a 90% prediction interval for the price of a home given that the living area of the house is 1500 square feet.** Interpret the result. (5p)

e) The correlation between the variables LONG and SQFT is negative (-0.55). **Use this information to briefly explain why the coefficient for SQFT is smaller in model 2, compared to model 1.** (5p)

price	sqft_living	bedrooms	lat	long	waterfront
205	1210	3	47.4873	-122.345	0
225	1790	3	47.508	-122.337	0
825	3810	4	47.5018	-122.38	0
151.1	840	3	47.496	-122.349	0
335	1260	3	47.5114	-122.361	0
230	900	2	47.5083	-122.336	0
249.95	1290	4	47.513	-122.358	0
300	1420	3	47.5069	-122.367	0

Figure 1: An excerpt of the data. The whole data set contains 50 rows.

MODEL 1

Regression Statistics	
Multiple R	0.81609964
R Square	0.66601862
Adjusted R Square	0.65906068
Standard Error	144.228749
Observations	50

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	1991173.167	1991173.17	95.7205887
Residual	48	998492.731	20801.9319	
Total	49	2989665.898		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-83.26899	50.08658182	[redacted]	
sqft	0.27125339	0.02772506		

MODEL 2

Regression Statistics	
Multiple R	0.83645071
R Square	0.69964979
Adjusted R Square	0.68686892
Standard Error	138.221746
Observations	50

ANOVA

	df	*SS*	*MS*	*F*
Regression	2	2091719.103	1045859.55	54.7419949
Residual	47	897946.7952	19105.251	
Total	49	2989665.898		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-441400.4	192373.3519		
sqft	0.23114205	0.03180726		
long	-3607.4344	1572.506903		

MODEL 3

Regression Statistics	
Multiple R	0.90773214
R Square	0.823977637
Adjusted R Square	0.803975096
Standard Error	109.3626219
Observations	50

ANOVA

	df	*SS*	*MS*	*F*
Regression	5	2463417.84	492683.569	41.1936477
Residual	44	526248.055	11960.1831	
Total	49	2989665.9		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-411872.8884	153168.913		
sqft	0.18850848	0.02916602		
bedrooms	14.79040986	22.4611531		
lat	-4030.368552	2215.47294		
long	-4930.924229	1485.722		
waterfront	540.6605886	115.977041		

--- END OF EXAM ---