STOCKHOLM UNIVERSITY
Department of Statistics
Spring 2022

Ulf Högnäs
Edgar Bueno (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2022-02-11

Time:	08.00 - 13.00
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.

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

GOOD LUCK!

Problem 1

The table below shows the total number of eligible voters (in thousands) and the percentage of them who voted in the election for Swedish parliament 2014, by age category and sex.

	Men		Women	
Agre group	Eligible voters 1000s	Voted, percent %	Eligible voters 1000s	Voted, percent %
18-24	420	79,3	393	83,3
25-29	290	78,9	286	84,0
30-34	269	82,5	265	85,4
35-39	276	85,3	267	86,8
40-44	307	87,0	302	89,0
45-49	321	85,5	320	88,2
50-54	300	86,7	292	88,8
55-59	281	86,9	279	91,0
60-64	277	88,2	277	90,8
65-69	272	91,6	281	92,2
70-74	250	91,3	259	90,5
75-79	152	87,4	178	86,7
80+	199	80,8	316	69,5
total	**3614**	**85,2**	**3715**	**86,4**

a. **What percentage of eligible female voters older than 69 voted in the election?** (5p) Choose the alternative closest to your own answer.

(A) 20.3%
(B) 21.9%
(C) 80.8%
(D) 82.2%
(E) 83.9%

B **Find the interval that contains the first quartile for age among men who voted.** (5p) Tip: First calculate the number of voters in each category.

(A) 18-24
(B) 25-29
(C) 30-34
(D) 35-39
(E) 40-44

Problem 2

The following table describes a random variable *X*. **Find the variance of *X*.** (5p)

x	-1	0	1
$P(X = x)$	0.4	0.2	0.4

a. **Find the variance of *X*.** (5p) Choose the alternative closest to your answer.

A) 0
B) 0.25
C) 0.40
D) 0.50
E) 0.80

A studio software company has analyzed its customer database and the sales figures for a particular software. The software is available in two versions: the "Lite" version and the more expensive "Pro" version. Based on sales records, 50% of the customers were in the age group 16-24, 30% were in the age group 25-34 and the remaining 20% were 35 and older. The relative frequencies of the two software versions for each of the customer categories is as follows:

	16-24	25-34	35+
Lite	80%	60%	25%
Pro	20%	40%	75%

b. **What is the probability that a randomly chosen customer purchased the Pro version?** (5p) Choose the alternative closest to your answer.

A) 0.22
B) 0.27
C) 0.37
D) 0.45
E) 0.48

A test for celiac disease (an autoimmune disorder that causes gluten intolerance) is 93% accurate when the person does have the disease, and 97% accurate when the person does not have the disease. Suppose that 1% of the population of some country has this disease.

c. **We test a randomly selected person from the population - and the test is positive (so the test indicates celiac disease). Find the probability that the test shows the correct result.** (5p) Choose the alternative closest to your answer.

A) 0.12
B) 0.25
C) 0.53
D) 0.77
E) 0.93

(Note: In this problem, a patient is tested at random. In practice, patients are tested because the show symptoms that are typical of the disease. When this is the case, the test will be much more accurate. You should not administer this test at random!)

Problem 3

A pizzeria sells pizzas for €10 each. Suppose that the number of pizzas sold during a particular week is approximately normally distributed with mean 1500 and standard deviation 180.

a. **Find the probability that the revenue (total sales) from pizza is greater than €18000 that week.** (5p) Choose the alternative closest to your answer.

A) 0%
B) 5%
C) 50%
D) 95%
E) 100%

The pizzeria sells only two variations of pizza: Margarita and Marinara. Past sales data shows that 25% of customers prefer Marinara, while 75% of customers prefer Margarita.

b. **Suppose that we randomly select 16 customers. What is the probability that at least 4 out of these 16 customers prefer the Marinara pizza?** (5p) Choose the alternative closest to your answer.

(A) 37%
(B) 40%
(C) 60%
(D) 63%
(E) 66%

c. Suppose that we randomly select 100 customers. **Find the approximate probability that at least 30 of these 100 customers prefer the Marinara pizza. Use approximation method taught in the course.** (5p) Choose the alternative closest to your answer.

A) 0%
B) 15%
C) 41%
D) 85%
E) 100%

Problem 4

A statistics student wants to estimate the mean age of first-year 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	Sample Standard Deviation	n
Stockholm	21.93	4.95	30
Uppsala	22.52	4.56	40

a. **Find a 90% confidence interval for the difference in mean age between the two populations of students (Stockholm minus Uppsala).** (5p) Choose the alternative closest to your answer.

(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)

As part of a scientific study, a random sample of 10 overweight volunteers are given a course in nutrition by a nutritionist. The body weight of each participant is measured at the beginning of the study, and then again six months later. Assume that the body weights are normally distributed. You can find the weights in kilograms below:

Volunteer	1	2	3	4	5	6	7	8	9	10
Weight before	80	95	78	105	90	90	91	79	115	85
Weight after	80	92	77	100	91	89	93	77	112	85

b. **Find a 95% confidence interval for the average weight change, *after* minus *before*, of a (future) participant in this type of study.** (5p) Choose the alternative closest to your answer.

A) (-2.24, -0.15)
B) (-2.37, -0.03)
C) (-2.50, -0.10)
D) (-2.70, 0.30)
E) (-11.9, 9.5)

Problem 5

The scientists in problem **4b** also selected a random sample of 10 overweight volunteers to use as control group. They suspected that just being part of a study at all might have an effect on weight loss. The body weights of the participants in the control group were also measured in the beginning of the study and then again after six months. Assume that the body weights are normally distributed in both groups and that the variances are equal. The weight changes (where a negative means a weight loss) and sample standard deviations can be found in the table below:

	Mean weight change	Standard deviation	n
Treatment Group	-1.2	2.01	10
Control Group	-0.5	2.00	10

Test whether the treatment group loses more weight (more negative weight change, use *treatment* minus *control*) than the control group. Use 5% level of significance.

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

A) Reject H_0 if $t_{obs} < -1.73$
B) Reject H_0 if $t_{obs} < -1.64$
C) Reject H_0 if $|t_{obs}| > 1.64$
D) Reject H_0 if $|t_{obs}| > 1.73$
E) Reject H_0 if $|t_{obs}| > 1.96$

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

A) -0.55
B) -0.78
C) -1.56
D) -2.14
E) -2.28

Problem 6

A candy manufacturer produces multi-colored button-shaped chocolates. There are five different colors: red, green, blue, yellow, and brown. The manufacturer claims that each color is equally frequent in production, but that the distribution of colors in each individual bag is random. A student collects a random sample of 300 candies and counts the colors. You can find the counts in the table below:

Red	Green	Blue	Yellow	Brown
72	65	67	52	44

Test at the 5% level of significance whether the population of candies is equally distributed between the five colors.

a) **State your hypotheses, test statistic, critical value and decision rule.** (5p)

b) **Calculate the test variable.** (5p)

c) **State your conclusions and give a verbal interpretation.** (5p)

d) **Explain briefly what a type-I error is. Illustrate a situation in which this test would result in a type-I error.** (5p)

Problem 7

An American businessman owns a chain of airport liquor stores. He wants to study the relationship between price and sales for a new brand of bourbon (a type of distilled alcoholic drink). The price varies between his airport stores. In some airports, he has paid for advertising billboards, in other airports, he has no advertising. He uses a random sample of 12 airports to estimate three models:

Model 1: $sales = \beta_0 + \beta_1 \cdot price + \varepsilon$

Model 2: $sales = \beta_0 + \beta_1 \cdot price + \beta_2 \cdot ad + \varepsilon$

Model 3: $sales = \beta_0 + \beta_1 \cdot price + \beta_2 \cdot ad + \beta_3 \cdot (price \cdot ad) \cdot + \varepsilon$

$sales$	the number of bottles sold in one week
$price$	the price of one bottle, in dollars
ad	a dummy variable; 1 means that he is paying for ads in the airport; 0 means no ads
$(price \cdot ad)$	an interaction term between *price* and the dummy variable *ad*

Parts of the output for each model can be found on the next pages

a) **Use Model 2 to find a point estimate for the number of bottles sold, given that the price of one bottle is $20 and that the businessman is paying for advertising at that airport. Round to the nearest integer.** (5p)

b) **Formally test whether Model 3 is better than Model 2 (i.e. test whether or not the interaction term should be included, given that *price* and *ad* are included in the model)**. **Use 5% level of significance.** (5p)

c) **Calculate and interpret the coefficient of determination for Model 2.** (5p)

d) **Perform an F-test of Model 3. State the hypotheses and use the p-value to reach a conclusion. Use 5% level of significance.** (5p)

MODEL 1

Regression Statistics	
Multiple R	0.252
R Square	0.064
Adjusted R Square	-0.030
Standard Error	6.868
Observations	12

ANOVA

	df	*SS*
Regression	1	32.029
Residual	10	471.637
Total	11	503.667

	Coefficients	*Standard Error*
Intercept	44.873	11.943
Price	-0.971	1.178

MODEL 2

Regression Statistics	
Multiple R	
R Square	
Adjusted R Square	
Standard Error	5.382
Observations	12

ANOVA

	df	*SS*
Regression	2	242.980
Residual	9	260.686
Total	11	503.667

	Coefficients	*Standard Error*
Intercept	50.381	9.580
Price	-2.063	1.008
Ad	9.287	3.441

MODEL 3

Regression Statistics	
Multiple R	0.748
R Square	0.559
Adjusted R Square	0.394
Standard Error	5.267
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	281.702	93.901	3.384	0.075
Residual	8	221.965	27.746		
Total	11	503.667			

	Coefficients	*Standard Error*
Intercept	64.111	14.933
Price	-3.556	1.603
Ad	-14.063	20.050
Price*Ad	2.402	2.034

---END OF EXAM---