Mine Çağlar

January 8, 2006

Math 202: Statistics for Social Sciences

Fall 2006 FINAL EXAM

Calculator OK, 2 hours and 15 minutes.

Instructions: There are seven parts to this exam I-VII. Please inspect the exam and make sure you have all 6 pages of questions. Do all your work on these pages. If you use the back of a page, make sure to indicate that.

Remember: ***You must show your work to get proper credit.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME: SOLUTION KEY

Part I:	/15
Part II:	/15
Part III:	/15
Part IV:	/15
Part V:	/15
Part VI:	/15
Part VII:	/10
Total:	/100

<u>Part I.</u> (15 points)

1. (5 points) Fill in the blanks with the appropriate test (i. through viii.) for each situation (I through V). Some of the tests may be appropriate in several situations and some may not be used at all.

Tests:

i. one-sample z-test
ii. one-sample t-test
iii. two-sample z-test
iv. one-sample z-test for percentages (proportions)
v. two-sample z-test for percentages (proportions)
vi. χ^2-test with a null hypothesis that tells you the distribution in the population
vii. χ^2-test for independence
viii. F-test in ANOVA

Situtations:

I. You are drawing 80 times at random with replacement from a box. To test the hypothesis that the average of the box is 2.5, you would use one-sample z-test (i)

II. You are drawing 80 times at random with replacement from a box. To test the null hypothesis that the box is $\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}}$ you would use (vi) χ^2-test with a null hypothesis.

III. You conduct a survey where each subject (randomly sampled from a single population) is asked two questions for variables 1 and 2. To test the null hypothesis that the two variables are unrelated to each other, you would use (vii) χ^2-test. for independence

IV. To test the null hypothesis that the population means of four different populations are equal, you would use (viii) F-test. in ANOVA.

V. You are drawing 80 times at random from a box which contains only 0's and 1's (many 0's and many 1's). To test the null hypothesis that the percentage of 1's in the box is 40%, you would use (iv) one-sample z-test for percentages.

2. a) (7 points) Perform the test in situation V and state your conclusion assuming that 27 of the 80 draws are observed to be 1's. Set up the appropriate null and alternative hypotheses first.

$H_0: p = 0,4$

$H_1: p < 0,4$

$SE = \sqrt{\dfrac{(0,4)\cdot(0,6)}{80}} = 0,054 = 5,4\ \%$

$\hat{p} = \dfrac{27}{80} = 0,3375 = 33,8\ \%$

$z = \dfrac{33,8 - 40}{5,4} = -1,15$

p-value $= \dfrac{100-75}{2} = 12,5 > 5$. we do not reject H_0.

There is no significant evidence to say that the percentage of 1's in the box is different than 40

b) (3 points) Fill in the blanks:

"In random sampling (of many random samples), the distribution of sample percentages of 1's is approximately normally distributed"

"An example of a variable which may be assigned the values 0 and 1 could be the condition of smoking or not. "

Part II. (15 points) The following table gives the Census results for the distribution of occupied housing units by number of rooms.

Number of rooms in unit	*Owner-occupied (percent)*	*Renter-occupied (percent)*
1	2.0	9.2
2	3.8	12.9
3	11.9	32.5
4	14.5	26.5
5	16.7	12.5
6	22.3	4.8
7	11.7	1.0
8	6.5	0.3
9	10.5	0.4
Total	99.9	100.1

1. (4 points) What is the 10^{th} percentile of the distribution for owner-occupied units?

3 (number of rooms)

2. (5 points) What is the average number of rooms for renter-occupied units?

$$\bar{X} = \frac{(9,2).1 + (12,9).2 + (32,5).3 + (26,5).4 + (12,5).5 + (4,8).6}{100} + \frac{1.7 + (0,3).8 + (0,4).9}{100} = \frac{342,8}{100} = 3,43$$

3. (3 points) What is the median number of rooms for renter-occupied units?

$\underset{1 \text{ room}}{9,2} + \underset{2 \text{ rooms}}{12,9} + \underset{3 \text{ rooms}}{32,5} = 54,6$ Median is 3.

4. (3 points) How is the distribution of renter-occupied units likely to be: skewed or symmetric? If skewed, to the right or left?

Skewed to the right since while 1,2,3,4,5 rooms have large percentages and 6,7,8,9 have small percentages forming a right tail.

OR

The mean (average) is greater than the median. It's likely to be skewed to the right.

Part III. (15 points) The following are ratings of a television program by several viewers:

7.8, 7.0, 10.0, 9.1, 6.3, 7.2, 5.3, 8.0, 9.5, 9.7, 7.8, 10.0, 6.3, 5.9, 5.0, 6.1

Mean = 7.6 and **Standard Deviation** = 1.70

1. (3 points) What percent of the ratings is between 7 and 10?

$\frac{10}{16} \cdot 100 = 62{,}5\ \%$ or $\frac{7}{6} \cdot 100 = 43{,}75\ \%$

2. (3 points) What is the 85th percentile of the ratings?

$$\begin{array}{cc} 100 & 85 \\ 16 & x \\ \hline \end{array}$$

$x = 13{,}6$

So, if we order these rating from lowest to highest, the 13 or 14 th of these ratings will be the 85th perce since below it there would be 13 ratings. So it's 9.5 9.7

3. (7 points) Is the average rating significantly less than 9, or is the difference just due to a chance variation? Answer by performing a test of significance. Show all steps.

$H_0: \mu = 9$

$H_1: \mu < 9$

$SE = \frac{1.7}{4} \cong 0.425$ $t = \frac{7.6 - 9}{0.425} \cong -3{,}29$

From t-table with 15 degree of freedom $2{,}95 < 3{,}29$,

so $p < 0.05$.

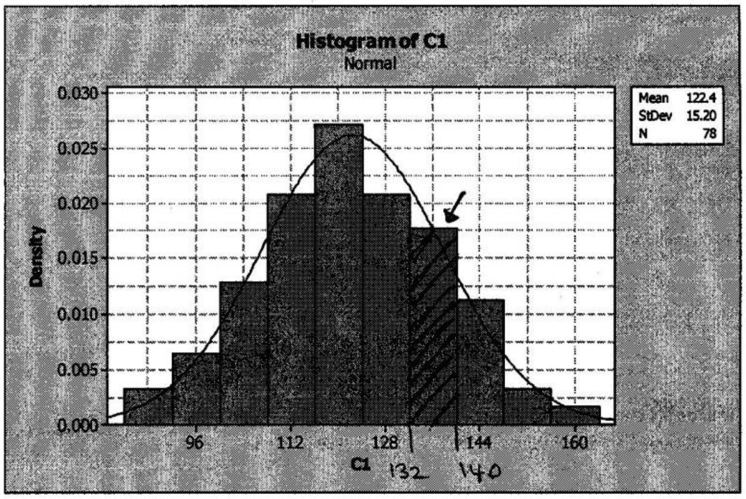
We reject H_0.

Average rating is significantly less than 9.

4. (2 points) What is your assumption about the population distribution for performing the specific test in question 3?

It is normally distributed

Part IV. **(15 points)** Consider the following histogram, and summary statistics for the blood pressures of subjects in a certain study.



Descriptive Statistics: C1

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
C1	78	122.36	1.72	15.20	89.67	113.16	122.14	134.04	159.90

Q1 stands for 25th percentile, Q3 stands for 75th percentile.

1. (3 points) Assuming that the blood pressure follows a normal curve, estimate the percentage of the people in the population with blood pressure in the interval [122.36-15.20, 122.36+15.20], that is [107.16, 137.56].

   1 SD from mean, so about 68 % is in this interval.

2. (3 points) Exactly what percent of the sample is in the interval [132,140]? Hint: Use the histogram.

   Area of the shown rectangle: $(140-132)\times(0{,}0175) = 0{,}14$

   So, 14 % of the sample is in [132,140].

3. (3 points) Exactly what percent of the sample is between 113.6 and 122.14? Hint: Use the Minitab output.

   113.16 : 25th percentile (25 % is below 113.16)

   122,14 : median (50 % is below 122.14)

   so exactly 25 % is between 113,16 and 122,14

4. (3 points) True or false: "An approximate 95% CI for the average blood pressure in the population runs between 118.92 and 125.8."

   $z_1 = 122{,}36 - 2\times 1{,}72 =$ 118.92 $z_2 = 122{,}36 + 2\times 1{,}72 =$ 125.8 .

If false, construct a 95% CI for the average blood pressure in the population.

95% CI = [118.92, 125.8]. The statement is true.

5. (3 points) Is the average blood pressure in the population contained in the confidence interval constructed in question 4? Why or why not?

   we do not know the average blood pressure in the population. It is most probably contained in the CI constructed in question 4. (with 95 % probability)

Part V. (15 points) A university administrator wishes to determine whether the instructor's (education) degree is related to the students' opinion of the quality of instruction received. A sample of students' evaluations of various instructors is given in the following data. At 1% level of significance, can the administrator conclude that the degree of the instructor is related to students' opinion about that instructor's effectiveness in the classroom? $d.f = 2\times 2 = 4$

observed:

Rating	Bachelor's (=university)	Master's	Doctorate	
Excellent	BE_o 14	ME_o 9	DE_o 4	sum: 27
Average	BA_o 16	MA_o 5	DA_o 7	sum: 28
Poor	BP_o 3	MP_o 12	DP_o 16	sum: 31
	sum: 33	sum: 26	sum: 27	total: 86

expected values: $BE_e = \frac{33.27}{86} = 10.4$ $ME_e = \frac{26.27}{86} = 8,2$ $DE_e = \frac{27.27}{86} = 8.5$

$BA_e = \frac{33.28}{86} = 10,7$ $MA_e = \frac{26.28}{86} = 8,5$ $DA_e = \frac{27.28}{86} = 8,8$ $BP_e = 11,9$ $MP_e = 9,4$ $DP_e = 9,7$

$\chi^2 = \frac{(14-10.4)^2}{10,4} + \frac{(9-8,2)^2}{8,2} + \frac{(4-8,5)^2}{8,5} + \frac{(16-10,7)^2}{10,7} + \frac{(5-8,5)^2}{8,5} + \frac{(7-8,8)^2}{8,8} + \frac{(3-11,9)^2}{11,9} + \frac{(12-9.4)^2}{9.4} + \frac{(16}{9}$

$\chi^2 = 19,6$. critical value $= 13,28$ $19,6 > 13,28$ So, there is significant evidence to conclude the degree of the instructor is related to students' opinion about that instructor's effectiveness in cla

Part VI. (15 points) An experiment involves two random samples of newborn rats. They are tested for a specific food called "lecithin". The purpose of the experiment is to demonstrate any deficit in learning performance that results from lecithin depreviation. The score for each animal is the number of errors. The following table summarizes the data.

	Regular Diet	No-Lecithin Diet
Sample size	10	5
Sample mean	25 errors	33 errors
Sample standard deviation	250 errors	140 errors

1. (3 points) If a 90% confidence interval for the difference in the number of errors of regular diet and no-lecithin diet rats were constructed, would that interval be for the difference

   a) in the population or b) in the sample

Choose the correct choice and explain only in one sentence: we know the exact difference in the sample, but we estimate it in the population, so use confi. interv

2. (12 points) Test if the difference in the diets causes significant difference in the number of errors. Show all steps and state your conclusion.

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 < 0$

$SE_1 = \frac{250}{\sqrt{10}} = 79,1$ $SE_2 = \frac{140}{\sqrt{5}} = 62,6$

$SE = \sqrt{(79,1)^2 + (62,6)^2} = 100,8$ we use t-test since sample size<

$t = \frac{(25-33)-0}{100.8} = -0,08$ $d.f. = (10-1)+(5-1) = 13$

critical value: 1,77 $0.08 < 1,77$

we do not reject Ho.

Difference in diets does not cause significant difference in the number of errors.

Part VII. (10 points) The results of a study on persuasion (=ikna etme) are given in the following table. Group A listened to a persuasive (=ikna edici) message that differed only slightly from their original attitudes. For Group B, there was a moderate discrepancy (=fark) between the message and their original attitudes. For Group C, there was a large discrepancy. For each subject, the amount of change in the attitude after reading the message (on a scale of 0 to 6) was measured.

H_0: The amount of change in the attitude does not differ between the three groups.

H_1: At least one of the amounts of change in the attitudes differ.

	Group A	Group B	Group C
	1	3	0
	0	4	2
	0	6	0
	2	3	4
	3	5	0
	0	3	0
	4		1
	2		
Mean	1.5	4	1
Std. dev.	1.51	1.26	1.53

$k = 3 \quad N = 21$

$n_1 = 8$

$n_2 = 6$

$n_3 = 7$

d.f.N = 2

d.f.D = 18

Use an analysis of variance with 5% level of significance to determine whether the amount of change in the attitude differs between the three groups.

$$\bar{x}_{GM} = \frac{(1,5).8 + 6.4 + 7.1}{21} = 2,05$$

$$s_B^2 = \frac{8.(1,5-2,05)^2 + 6.(4-2,05)^2 + 7.(1-2,05)^2}{2} = 16,48$$

$$s_W^2 = \frac{7.(1,51)^2 + 5.(1,26)^2 + 6.(1,53)^2}{18} = 2,11 \quad \text{(also given)}$$

$$F = \frac{16,48}{2,11} \cong 7,81$$

critical value: 3,55

7.81 > 3,55. we reject H_0.

There is significant evidence that the amount of change in the attitude differs between the three groups.

Useful formula and results:

Between group variance: $s_B^2 = \dfrac{\sum n_i(\bar{x}_i - \bar{x}_{GM})^2}{k-1}$ = Please find.

Within group variance: $s_W^2 = \dfrac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)}$ = 2.11 and $F = \dfrac{s_B^2}{s_W^2}$