Mine Çağlar January 6, 2019

Math 202: Statistics for Social Sciences

FALL 2018 FINAL EXAM

Calculator OK, 2 hrs. 15 minutes

Instructions: There are six parts to this exam I-VI. Please inspect the exam and make sure you have all six 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. In hypothesis testing questions, show all steps of the test and state your conclusion in plain English.***

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

Part I:	/15
Part II:	/ 20
Part III:	/25
Part IV:	/ 20
Part V:	/15
Part VI:	/15
Total:	/110

Formulas:

1) Binomial formula:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \quad k=1,\dots,n$$

2) Chi-squared: $$\chi^2 = sum\ of\ \frac{(observed\ \ frequency - expected\ \ frequency)^2}{expected\ \ frequency}$$

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

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

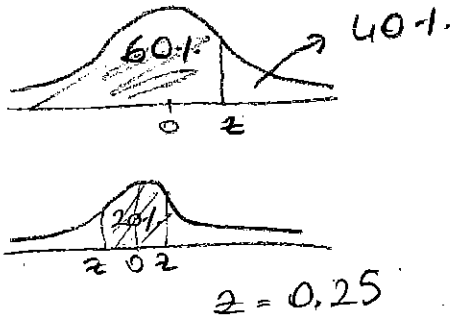
Here s_i is SD of i^{th} sample and s_i^2 is the variance, namely SD^2 (the square of SD).

Part I. (15 points) The weights of all eighteen-year old women follow a normal curve.

1. (6 points) Find the 60th percentile of the weight distribution. Assume that its mean is 59 kg and standard deviation is 6 kg.

$$\frac{x-59}{6} = 0.25$$

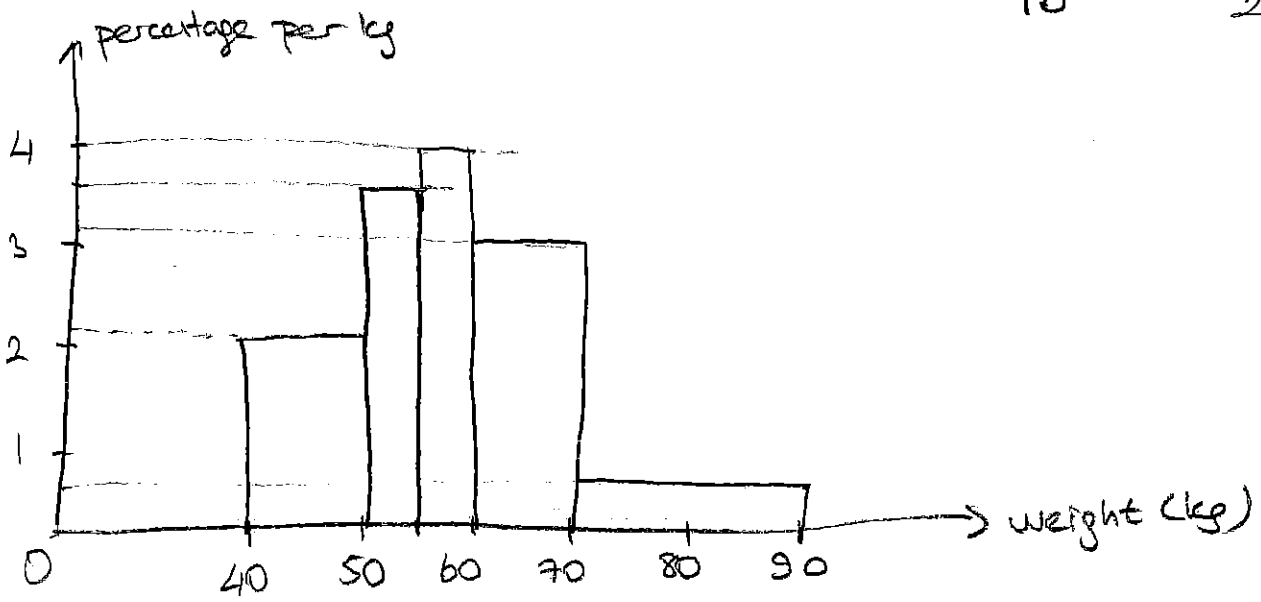
$$x = 60.5$$



2. (9 points) A random sample of size 32 is drawn from the population of eighteen-year old women. The relative frequencies are as given below. Draw a density scale histogram.

Weight	[40,50)	[50,55)	[55,60)	[60,70)	[70,90)
Relative frequency	0.21	0.18	0.20	0.31	0.10
Density	0.021	0.036	0.04	0.031	0.005

$\frac{0.21}{10}$ $\frac{0.18}{5}$ $\frac{0.20}{5}$ $\frac{0.31}{10}$ $\frac{0.10}{20}$



Part II. (20 points) A university is supporting environment friendly initiatives such as carpooling (sharing a car when going back and forth to the university) among students. The university surveyed all 400 students who live outside the campus and reported the following frequency table. That is, the students responded as to whether they own a car or not and stated the location of their home.

Location of Home

Car owner	Inner City	Outside City	Elsewhere in City	Total
Yes	20	70	30	120
No	120	50	110	280
Total	140	120	140	400

1. (6 points) Answer the following true (T) false (F) questions. If TRUE, then explain in one sentence.

T If a student owns a car, he or she is more likely to live elsewhere in the city than to live in the inner city area. prob of elsewhere $= \frac{30}{120} > \frac{20}{120}$ = prob of inner city

F A student is more likely to own a car if s/he lives in the city (inner or elsewhere) than if s/he lives outside the city.

F If a student does not own a car, s/he is more likely to live outside the city than to live in the city (inner or elsewhere).

2. (14 points) Are car ownership and location of home independent variables in this university?

H_0: Car ownership and location of home are independent

H_a: The two variables are not independent

Expected frequencies

$n = 400$

$d.f = (2-1)(3-1) = 2$

$\chi^2_{2,.05} = 5.99$

	Inner city	Outside	Elsewhere	Total
Yes	$\frac{120 \cdot 140}{400} = 42$	120 − 84 = 36	140 − 98 = 42	120
No	140 − 42 = 98	280 − 98 − 98 = 84	$\frac{280 \cdot 140}{400} = 98$	280
Total	140	120	140	

$\chi^2 = \frac{(20-42)^2}{42} + \cdots + \frac{(110-98)^2}{98} = ? > 5.99 \Rightarrow$ Reject H_0

So, car ownership and location of home are not independent variables in this university.

Part III. (25 points) A pharmaceutical company tests whether its drug relieves headache in a randomized controlled experiment. The treatment group is given the drug and the control group is given the placebo. In the treatment group 38% of 50 patients are relieved from a headache, whereas this rate is 25% in the control group formed by 44 patients.

1. (7 points) Construct a 99% confidence interval for the difference in the relief percentage between the treatment and control groups.

$$(38\% - 25\%) \mp 2.6 \underbrace{\sqrt{\underbrace{\frac{38.62}{50} + \frac{25.75}{44}}_{9.47}}\%}_{24.63}$$

z	area
2.6	99.07

$$\Rightarrow [-11.63\%, 37.63\%]$$

2. (10 points) Perform a hypothesis test for the claim that the drug relieves headache by 40%, or not, considering the treatment group only.

$H_0 : p = 40\%$

$H_a : p \neq 40\%$

$$z = \frac{38-40}{\sqrt{\frac{40.60}{50}}} = -0.289 \rightarrow \text{area} : 23.58$$

$$\text{p-value} = 100\% - 23.58\% = 76.42\% > 5\%$$

Do not reject H_0.

Yes, the drug relieves headache by 40%.

3. (8 points) Assume that the drug relieves headache by 40%. Find the probability that at most seven subjects are relieved in a random sample of 9 subjects with headache when all of them take the drug.

$$1 - \binom{9}{8}(0.40)^8(0.60)^1 - \binom{9}{9}(0.40)^9(0.60)^0 = 0.9962$$

Part IV. (20 points) A public policy polling group is investigating whether people living in the same household tend to make independent political choices. They select 200 homes where exactly three voters live. The residents are asked separately for their opinion ("yes" or "no") on a political issue.

If their opinions are formed independently, then the number of residents who say "yes" should have the distribution $p_0 = 0.22, p_1 = 0.43, p_2 = 0.29, p_3 = 0.06$. The following frequency table summarizes the data for the sample of 200 homes.

$n=200$

Number saying "Yes" in a household	Frequency	Expected Freq
0	30	44
1	56	86
2	73	58
3	41	12

1. (12 points) Do the data indicate that people living in the same household tend to make independent political choices?

$H_0: p_0 = 0.22, \quad p_1 = 0.43, \quad p_2 = 0.29, \quad p_3 = 0.06$ 55el?

$H_a:$ at least one p is different

Exp. freqs $= (200)(0.22) = 44$

$(200)(0.43) = 86$

$(200)(0.29) = 58$

$(200)(0.06) = 12$

$$\chi^2 = \frac{(30-44)^2}{44} + \frac{(56-86)^2}{86} + \frac{(73-58)^2}{58} + \frac{(41-12)^2}{12} = 88.88$$

$df = 4-1 = 3$

$\chi^2_{0.01,3} = 9.21 \Rightarrow$ p-value $< 1\%$.

Reject H_0

The political choices in a household are not independent from each other.

2. (4 points) Find the sample average of the number of people in a household who said "Yes".

$$\frac{0\times30 + 1\times56 + 2\times73 + 3\times41}{200} = 1.625$$

3. (4 points) What is the median of the number of people in a household who said "Yes"?

$200 \to 100$, >100

$30+56 = 86 < 100$ but $30+56+73 > 100$

$\frac{1}{0}$ $\frac{1}{1}$ $\frac{1}{2}$

So, median is 2.

Part V. (15 points) Sociologists study the feeling of alienation (=yabancılık) in big cities. They have formulated "mobility index" measuring the changes in the neighborhood, which can contribute to this feeling. The mobility index is higher for cities with more transient (=geçici) populations, and lower in urban areas where neighborhoods are more stable (=kalıcı).

In a specific country, 5 big cities are randomly selected from its western region and 8 cities are randomly and independently selected from its eastern region. The mean mobility index is 51.8 and the standard deviation is 6.3 in the west, whereas the mean and SD are 65.1 and 7.7, respectively, in the east.

1. (12 points) Is the average mobility significantly less in the western region than the eastern region? Report the P-value.

$H_0 : \mu_W = \mu_E$

$H_1 : \mu_W < \mu_E$

Ha ? OK

$$t = \frac{51.8 - 65.1}{\sqrt{\frac{(6.3)^2}{5} + \frac{(7.7)^2}{8}}} = -3.39$$

$SE_1 = \frac{6.3}{\sqrt{5}}$ OK

SE_1^2 ...

4 7

$df = 5 + 8 - 2 = 11 \Rightarrow$ p-value $< 0.5\%$

p-value $< 0.5\%$

$\Rightarrow$ Reject H_0

Avg mobility is significantly less in the western region than the eastern region.

2. (3 points) What are your assumptions to be able to perform the test in Question 1?

The mobility index is normally distributed in both populations.

Part VI. (15 points) For testing the effect of smoking on heart rate (=nabız), the following table gives the heart rate data collected for three independent groups of subjects.

Nonsmokers	Light smokers	Heavy smokers
69	55	91
52	60	72
71	78	81
58	58	67
59	62	95
65	66	84
$\bar{x}$: 62.3	63.2	81.7
s (or SD) : 7.3	8.2	10.8

Does smoking affect the heart rate significantly?

$H_0 = \mu_x = \mu_y = \mu_z$

H_1 = At least one of them is different.

$n_x = 6$ | $\bar{x}_x = 62.3$ | $s_x = 7.3$ | $k = 3$ | $s_x^2 = 53.29$

$n_y = 6$ | $\bar{x}_y = 63.2$ | $s_y = 8.2$ | $N = 18$ | $s_y^2 = 67.24$

$n_z = 6$ | $\bar{x}_z = 81.7$ | $s_z = 10.8$ | | $s_x^2 = 116.64$

$$\bar{x}_{GM} = \frac{n_x\bar{x} + n_y\bar{x}_y + n_z\bar{x}_z}{N = 18} = 69.066$$

$$s_B^2 = \frac{n_x(\bar{x}_x - \bar{x}_{GM})^2 + n_y(\bar{x}_y - \bar{x}_{GM})^2 + n_z(\bar{x}_z - \bar{x}_{GM})^2}{k-1} = 719.42$$

$$s_W^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2 + (n_z-1)s_z^2}{N-k} = 79.096$$

$$F_{stat} = \frac{719.42}{79.056} = 9.1$$

$F_{2,15} = 3.68$ $\quad 3.68 < 9.1 \Rightarrow$ Reject H_0

Smoking affects the heart rate.