Mine Çağlar

January 15, 2010

Math 202: Statistics for Social Sciences

Fall 2009 FINAL EXAM

Calculator OK, 2 hours and 15 minutes.

Instructions: There are six parts to this exam I-VI. 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.***
SHOW ALL STEPS OF A HYPOTHESIS TEST!!!

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: KEY **Lecture:** 9:30 or 12:30 (circle one)

Part I:	/20
Part II:	/15
Part III:	/15
Part IV:	/15
Part V:	/20
Part VI:	/15
Total:	/100

Formulas:

1) Confidence interval (CI) : $\overline{X} \mp z\,SE$ with $SE = \frac{SD}{\sqrt{n}}$

or $\overline{X} \mp t\,SE$ or $\hat{p} \mp z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or

$(\hat{p}_1 - \hat{p}_2) \mp z\,SE_{2\text{-sample}}$ or $(\overline{X}_1 - \overline{X}_2) \mp z\,SE_{2\text{-sample}}$

or $(\overline{X}_1 - \overline{X}_2) \mp t\,SE_{2\text{-sample}}$

where $SE_{2\text{-sample}} = \sqrt{(SE_1)^2 + (SE_2)^2}$,

and $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for percentages.

2) Binomial formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0,1,2,\ldots,n$$

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

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

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

Here s_i is SD of i^{th} sample and s_i^2 is the variance, namely SD^2 (the square of SD).
To find SD, sum all (entry-average)^2 values, then divide by n-1, and then take the square root

Part I. (20 points) The eating habits of Americans are continuously studied by social scientists who are watching for changing trends and their effects on family life. Some social scientists state the importance of the evening meal as a family meeting. A large survey of adult Americans shows that the distribution of the number of evening meals American adults cook at home in a usual week is given by

0	1	2	3	4	5	6	7
8%	5%	10%	13%	15%	21%	9%	19%

1. (5 points) Find the mean number of days in a week that American adults have evening meals cooked at home. Assume the survey contained 100 American adults

$$\mu = \frac{0\times 8 + 1\times 5 + 2\times 10 + \cdots + 7\times 19}{100} = 4.16 \text{ days}$$

2. (2 points) What percent of American adults cook less than two days per week?

$\underbrace{\text{less than two days}}_{\text{0 days or 1 day}}$

$$8\% + 5\% = 13\%$$

3. (10 points) If a similar survey on 78 adults in Turkey yields the following frequency distribution, can we generalize that the evening eating habits of Turkish adults are similar to American adults or not?

37 observations here

Number of days	0	1	2	3	4	5	6	7
Observed Frequency	9	6	5	7	10	13	12	16
Expected Freq.	6.24	3.9	7.8	10.14	11.7	16.38	7.02	14.82

8% × 78   5% × 78   etc. . . . .   19% × 78

H_0: $p_0 = 8\%$, $p_1 = 5\%$, ..., $p_7 = 19\%$
H_a: at least one p is different.

$$\chi^2 = \frac{(9-6.24)^2}{6.24} + \frac{(6-3.9)^2}{3.9} + \cdots + \frac{(16-14.82)^2}{14.82} = 8.9$$

d.f. $= 8-1 = 7$ $\Rightarrow$ $10\% < \text{P-value} < 30\%$

Since P-value $> 5\%$, we do not reject H_0.

We can say evening eating habits of Americans and Turks are similar

4. (3 points) Estimate the median number of days in a week that Turkish adults cook their meals at home.

Use the sample frequency distribution from 78 adults: median is the average of $\frac{78}{2} = 39^{th}$ and 40^{th} observations. $\Rightarrow \frac{5+5}{2} = 5$ //

Part II. **(15 points)** A genetics institute has developed a technique called MicroSort, which supposedly increases the chances of a couple having a baby girl.

1. (6 points) A random sample of 14 couples who want baby girls is chosen. Assume that a couple has equal chances of having a girl or a boy if the technique was not effective. Find the probability that at least 13 couples have a baby girl after using the MicroSort technique, assuming the technique is not effective.

$\Rightarrow p = \frac{1}{2}$ $\quad n = 14$

$$P(13) + P(14) = \binom{14}{13}\left(\frac{1}{2}\right)^{13}\left(\frac{1}{2}\right)^{1} + \binom{14}{14}\left(\frac{1}{2}\right)^{14}\left(\frac{1}{2}\right)^{0}$$

$$= 0{,}000916$$

2. (2 points) What do you conclude in view of your answer to Question 1 about the effectiveness of MicroSort for having a baby girl, if indeed 13 of the 14 couples had girls and 1 couple had a boy?

If the technique MicroSort was not effective, we would have observed a very unlikely event! Indeed, $0.000916 < 0.05$ it is a very unlikely event that we had so many girls. So, MicroSort is effective

3. (7 points) Since we have a smaller chance error with larger samples, a new sample of size 150 was taken to test the claim that MicroSort increases the chance of having a baby girl. In this sample, 102 couples are found to have a girl. Does MicroSort significantly increase the chance of having a baby girl?

1.) $H_0: p = \frac{1}{2} (= 0{,}50)$

$H_a: p > \frac{1}{2}$

We have $\hat{p} = \frac{102}{150} = 0{,}68$

2) $SE = \sqrt{\frac{(0.50)(0{,}50)}{150}} \cong 0{,}04$

$z = \frac{0.68 - 0{,}50}{0.04} = 4{,}5$

3) P-value $< \frac{100 - 99.9991}{2}\,\% \cong 0.00045 = 0{,}045\,\%$

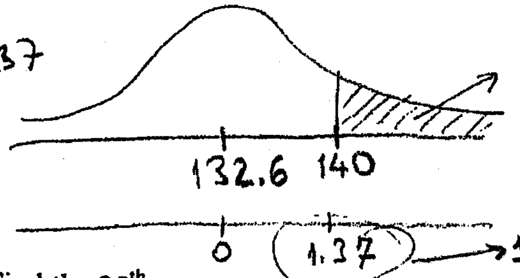
4) Reject H_0

5) Yes, MicroSort significantly increases the chance of having a baby girl.

Part III. (15 points) Anthropologists analyse measurements of human skulls from different periods in history to determine whether they change over time. The largest breadth (=genişlik) is found from skulls of Egyptian males who lived around 3300 B.C. Results show that those breadths are normally distributed with a mean of 132.6 mm and a standard deviation of 5.4 mm.

1. (3 points) Find the probability of getting a value greater than 140 mm if a skull is randomly selected from the period of around 3300 B.C.

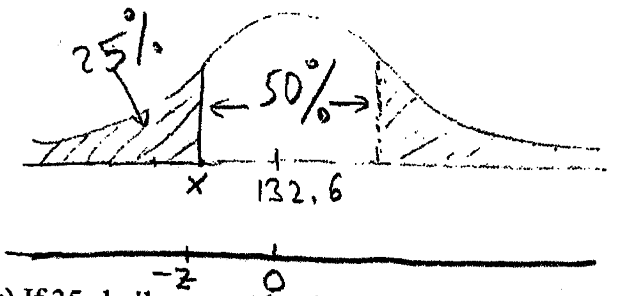
$\frac{140-132.6}{5.4} = 1.37$



table

$\frac{100-82.3}{2}\% = 8.85\%$

2. (4 points) Find the 25^{th} percentile of the breadth distribution.



Normal table ⇒ $z \cong 0.70$

⇒ $132.6 - 0.7 \times 5.4$

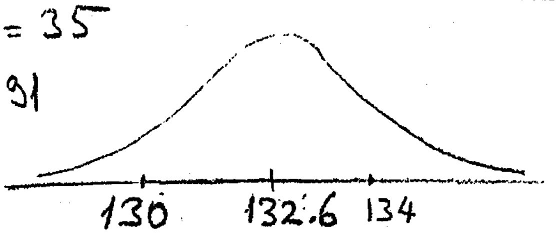
$= 132.6 - 3.78$

⇒ $= 128.82$

3. (4 points) If 35 skulls are randomly selected from the 3300 B.C. population, what is the probability that their mean breadth will be between 130 mm and 134 mm?

$n = 35$

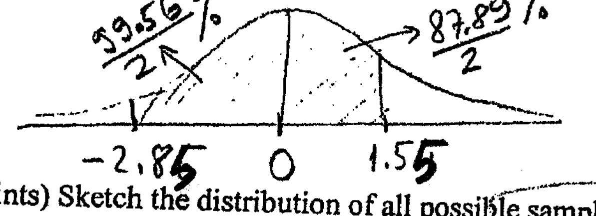
$SE = \frac{5.4}{\sqrt{35}} = 0.91$



$\frac{130-132.6}{0.91} \cong -2.86 \rightarrow -2.85$

$\frac{134-132.6}{0.91} \cong 1.54 \rightarrow 1.55$

From z-table



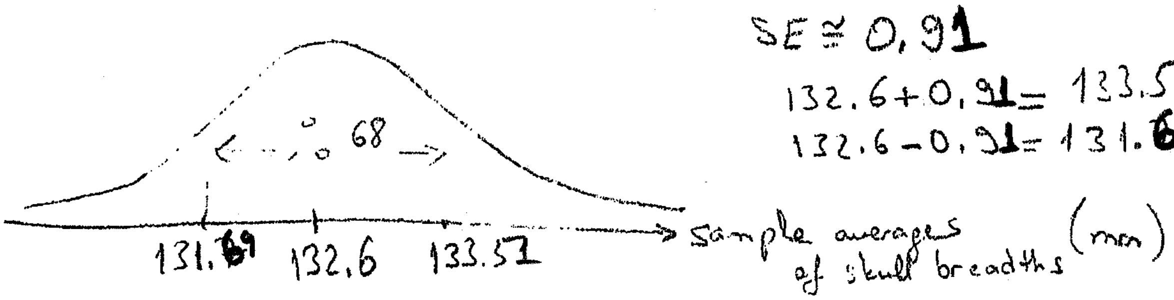
$\frac{99.56+87.89}{2}\% \cong 93.7\%$

4. (4 points) Sketch the distribution of all possible sample averages from the 3300 B.C. population with size 35, by showing at least 3 points on the x-axis. Also label the x-axis.

$SE \cong 0.91$

$132.6 + 0.91 = 133.51$

$132.6 - 0.91 = 131.69$



Part IV. (15 points) Social psychologists investigated the potentially harmful effects of violent music lyrics (=şarkı sözü). In this study, a score from each subject is determined by averaging his/her answers on a seven-point scale (1,2,3, ...,7) after s/he listens to a song with violent lyrics. As a result, each subject receives a final score in [0,7]. The higher the score, the more the subject thinks an ordinary word is aggressive (saldırgan). The following is the data set from 50 subjects:

1.8 1.9 $\rightarrow \frac{2}{50} = 4\%$

2.2 2.4 2.5 2.6 2.7 2.8 2.8 2.9 $\rightarrow \frac{8}{50} = 16\%$

3.0 3.1 3.2 3.2 3.3 3.4 3.4 3.4 3.4 3.5 3.5 3.5 3.6 3.7 3.7 3.8 3.9 3.9 → 18 obs.ns

4.0 4.1 4.1 4.1 4.1 4.2 4.2 4.3 4.4 4.4 4.5 4.6 4.7 4.7 4.8 4.9 $\rightarrow \frac{16}{50} = 32\%$

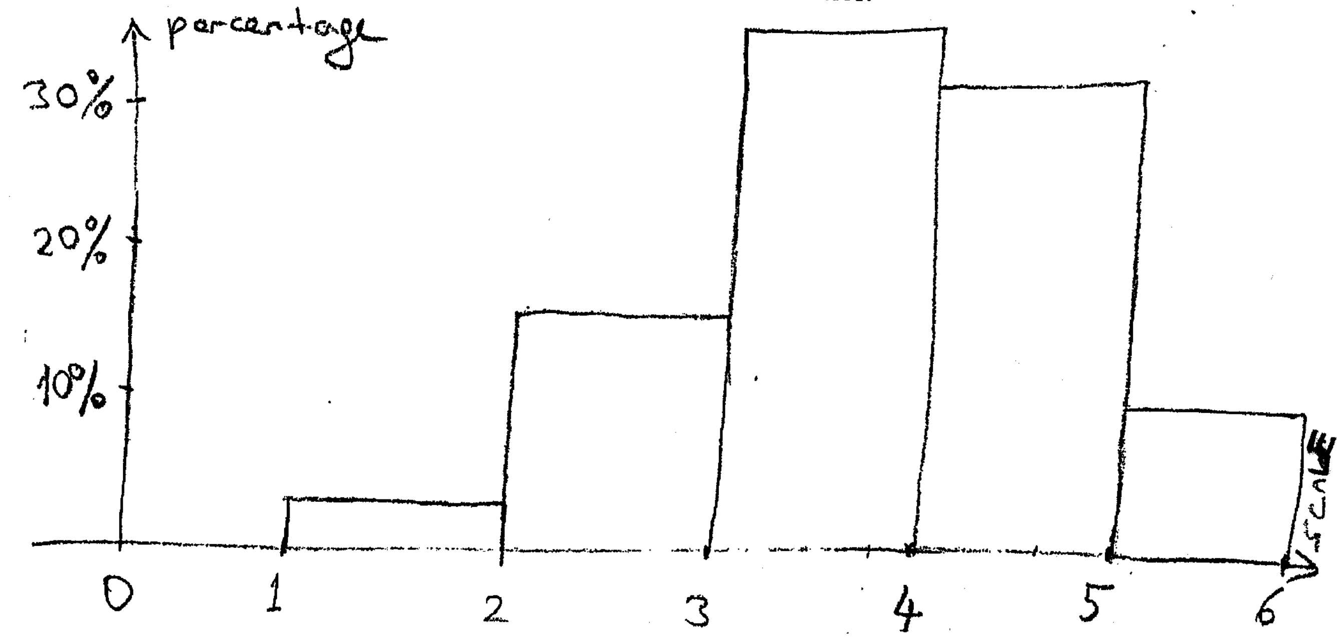
5.0 5.2 5.3 5.7 5.8 5.8 $\rightarrow \frac{6}{50} = 12\%$

Total 50

1. (3 points) What is the relative frequency of the observations that are between 3.0 and 4.0 (3.0 included, 4.0 not included)

$$\frac{18}{50} = 36\%$$

2. (7 points) Draw a histogram of the data set (draw whichever type of histogram you want: density scale or relative frequency). Do not forget to label the axes.



3. (5 points) The investigators found an association (=relationship) between the score given in this part, namely, aggressive cognition (=algı) score and the subjects' habit of listening to songs with violent lyrics in their every day life (the level of this habit was separately determined from each subject). Write a confounding variable for the association found between those two variables by explaining in at most 3 sentences.

"Aggressive cognition score" and "level of violent lyrics listening habit" are found to be associated. The subjects personality, his/her agressiveness in general or the contrary behavior in general affect both the aggressive cognition and habit of listening violent lyrics. So personality is a confounding variable.

Part V. (20 points) Certain foods and wines interact well (=beraber iyi gitmek) while others interact poorly. Taste test scores of 10 randomly selected people are recorded after they are randomly assigned to three food groups below. The higher the score, the better is the taste. In each group, the subjects drank red wine while eating the indicated food.

Food	Red Meat (1)	Chicken (2)	Fish (3)
Taste scores	8,5,5,10	2,4,6	3,2,4
Mean score	7	4	3
s^2	6	4	1

1. (10 points) Is there a significant taste difference when red wine is consumed with different kinds of meat?

$H_0: \mu_1 = \mu_2 = \mu_3$, H_a: at least one mean is different from the others.

$$\bar{X}_{GM} = \frac{8+5+5+10+2+4+6+3+2+4}{10} = 4.9$$

$$s_B^2 = \frac{4(7-4.9)^2 + 3(4-4.9)^2 + 3(3-4.9)^2}{3-1} = 15.45 \rightarrow d.f_N = 2$$

$$s_W^2 = \frac{(4-1)6 + (3-1)4 + (3-1)1}{10-3} = 4 \rightarrow d.f_D = 7$$

$$F = 15.45/4 = 3.86 \qquad F_{critical} = 4.74$$

Since $F < F_{critical}$, we cannot reject H_0 at $\alpha = 5\%$

There is no significant taste difference with different kinds of meat.

2. (4 points) Find the mean and the standard deviation of the "white meat" group, that is, of all scores in Chicken and Fish groups together.

$$\bar{X}_W = \frac{4\times 3 + 3\times 3}{6} = 3.5$$

$$SD_W = \sqrt{\frac{(2-3.5)^2+(4-3.5)^2+(6-3.5)^2+(3-3.5)^2+(2-3.5)^2+(4-3.5)^2}{5}}$$

$= 1.52$ //

3. (6 points) Conduct an appropriate hypothesis test to find out if the choice of red or white meat makes a difference in taste when consumed with red wine, assuming the scores are normally distributed.

1) $H_0: \mu_R = \mu_W$
$H_a: \mu_R > \mu_W$

2) $t = \frac{\bar{X}_R - \bar{X}_W}{SE_{2\text{-sample}}}$

$\Rightarrow t = \frac{7-3.5}{1.37} = 2.55$

$SD_R = \sqrt{6} = 2.45$
$SD_W = 1.52$
$SE_R = \frac{2.45}{\sqrt{4}} = 1.225$
$SE_W = \frac{1.52}{\sqrt{6}} = 0.62$
$SE_{2\text{-sample}} = \sqrt{1.225^2 + 0.62^2} = 1.37$

3) $d.f. = (4-1)+(6-1) = 8 \Rightarrow 1\% < P\text{-value} < 2.5\%$ from t-table

4) Since P-value $< 5\%$, we reject H_0.

5) Yes, consuming red meat with red wine is significantly tastier

Part VI. (15 points) It is well-known that if a ship is sinking, the lifeboats are filled first with women and children. The following frequency table summarizes the ending of the people on Titanic, the ship which sank in 1912.

	Men	Women	Children	Total
Survived	332	318	56	706
Died	1360	104	53	1517
Total	1692	422	109	2223

1. (5 points) Find the conditional distribution of people who died.

Men	Women	Children
$\frac{1360}{1517} \approx 89.6\%$	$\frac{104}{1517} \cong 6.9\%$	$\frac{53}{1517} \cong 3.5\%$

2. (10 points) Were the lifeboats filled first with women and children in Titanic, do you think? Perform your test of significance at $\alpha = 0.01$.

1) H_0: Whether a person survived is independent of whether the person is a men, woman or child

H_a: Survival status and type of person are independent

2) Expected frequencies under H_0

	Men	Women	Children
Survived	537.4	134	34.6
Died	1154.6	288.	74.4

$\frac{706 \times 1692}{2223} \cong 537.4$

$\frac{1692 \times 1517}{2223} \cong 1154.6$

$\frac{706 \times 422}{2223} \cong 134$

etc

$$\chi^2 = \frac{(332-537.4)^2}{537.4} + \frac{(134-318)^2}{134} + \frac{(56-34.6)^2}{34.6} + \frac{(1360-1154.6)^2}{1154.6} + \frac{(422-288)^2}{288} + \frac{(109-74.4)^2}{74.4} \cong 504.65$$

3) $d.f. = (2-1)\times(3-1) = 2 \Rightarrow \chi^2_{critical} = 9.21$ at $\alpha = 0.01$ (or P-value $< 1\%$)

4) Since $16.09 > 9.21$, reject H_0.
(or since P-value < 0.01, reject H_0)

5) Survival status and type of person are not independent. The data indicate that women and children were favored.