January 11, 2008

Mine Çağlar

Math 202: Statistics for Social Sciences

Fall 2007 FINAL EXAM

Calculator OK, 120 min.

Instructions: There are five parts to this exam I-V. Please inspect the exam and make sure you have all 5 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: _____

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

Formulas:

1) Confidence interval (CI) : $\overline{X} \mp z\ SE$

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

$(\hat{p}_1 - \hat{p}_2) \mp z\ SE_{2-sample}$

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

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

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

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

2) Binomial formula:

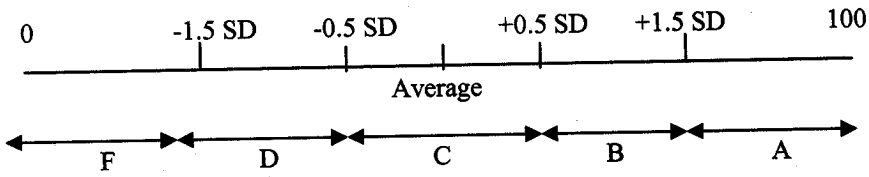
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0,1,2,\ldots,n$$

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

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

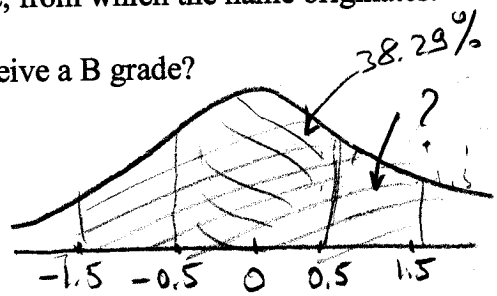
Within group variance: $s_W^2 = \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)}$ and $F = \frac{s_B^2}{s_W^2}$

Part I. (20 points) Grading with a "curve" system with only A, B, C, D, F grades would be as follows: The grades in the interval "the average ∓ 0.5 SD (standard deviation)" would receive a C grade and so on. The grading is described on the scale below.



In this system, the grades are assumed to follow a Normal curve, from which the name originates.

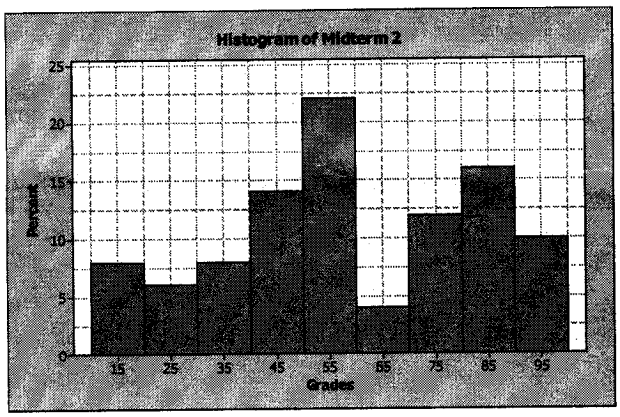
1. (5 points) What percent of the students do you expect to receive a B grade?



From the normal table:

$$\frac{86.64\% - 38.29\%}{2} \cong 24.2\%$$

Consider the following relative frequency histogram and the summary statistics obtained from the MT 2 grades of your class.



Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
C1	50	58.4	3.28	23.2	14.00	42.5	57.0	80.0	98.0

2. (5 points) What are the lowest and the highest ~~grades~~ scores that fall into the B range in this exam?

$58.4 + 0.5(23.2) = 70$ lowest score

$58.4 + 1.5(23.2) = 93.2$ highest score

3. (4 points) What is the percentage of students who receive an A grade from this exam?

10 units

From the histogram 10% of students received in (90, 100)

Since $100 - 93.2 = 6.8$, about 6.8% will receive A.

4. (3 points) What is the inter-quartile range in this exam?

$$Q_3 - Q_1 = 80 - 42.5 = 37.5$$

5. (3 points) Describe the population from which your class can be considered as a "random sample" in one sentence only.

All Math 202 students from past to future

Part II. (20 points) How do makers of Selpak know exactly how many tissues (=kağıt mendil) to put in a box? In a survey conducted by a marketing team, the following data have been collected for the number of times a person cleans his/her nose with a tissue during a cold (=soğuk algınlığı):

	57	62	61	62	52	55	60	74	65
$x_i - \bar{x}$	-4	1	0	2	-9	-6	-1	13	4

$n = 9$

The producers of Selpak put 60 tissues per box believing that the true average is 60.

1. (4 points) Find the mean and the standard deviation of the sample.

$$\bar{x} = \frac{57+62+\dots+65}{9} = 61$$

$$s = \sqrt{\frac{(57-61)^2+\dots+(65-61)^2}{8}} = \sqrt{\frac{324}{8}} \cong 6.4$$

2. (9 points) Perform an appropriate test of significance to decide if Selpak should continue to put 60 tissues per box. Show all steps and state your conclusion in plain English. (Hint: If you did not solve question 1, you can take the standard deviation to be equal to 6 although it is not the exact answer)

$H_0: \mu = 60$

$H_a: \mu > 60$

$$SE = \frac{SD}{\sqrt{n}} = \frac{6.4}{3} \cong 2.13$$

$$t = \frac{61-60}{2.13} \cong 0.47 \quad , \quad d.f. = 9-1 = 8$$

P-value $> 25\%$ from the table, so P-value $> 5\%$

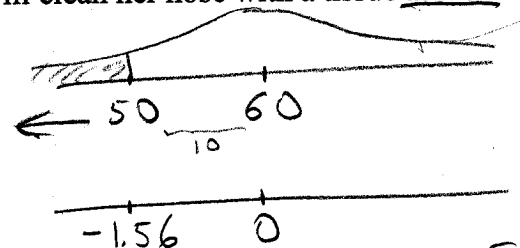
No significant evidence against H_0, do not reject it

Selpak can continue to put 60 tissues per box

3. (2 points) What were your assumptions before making the test in question 2?

1. The data comes from a normal population
2. The data form a simple random sample

4. (5 points) If you select a person randomly from the population, what is the probability that she will clean her nose with a tissue at most 50 times during a cold? Assume normal approximation.



$$\frac{-10}{6.4} = -1.56$$

From normal table 87.89%

the chances of atmost 50 $\Rightarrow \frac{100-87.89\%}{2} \cong 6.1\%$

Part III. (15 points) Many hotel customers fail to cancel their reservations in a timely manner when they are not going to stay. Such customers are called "no-shows". The number of no-shows is recorded in two different hotels, the averages being 15 and ~~20~~ 13 customers per day as shown in the following MINITAB output for Hilton and ~~Movenpick~~ Sheraton, respectively.

Variable	N	Mean	SE Mean	StDev
Hilton	30	15	0.53	2.9
~~Movenpick~~ Sheraton	40	13	0.41	2.6

1. (9 points) Construct a 95% CI for the difference in the number of no-shows per day.

$$(\bar{X}_1 - \bar{X}_2) \mp 2\, SE_{2\text{-sample}} \quad (z \approx 2 \text{ from Normal table for 95\%})$$

$$SE_{2\text{-sample}} = \sqrt{(0.53)^2 + (0.41)^2} = 0.67$$

$$\Rightarrow (15-13) \mp 2(0.67)$$

$$\Rightarrow 2 \mp 1.34 \Rightarrow [0.66, 3.34]$$

2. (6 points) Is there a significant difference in the number of no-shows of the two hotels?

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$$z = \frac{(15-13)-0}{0.67} \cong 2.99$$

$$p\text{-value} = \frac{100\% - 99.73\%}{2} = 0.135\% < 5\%$$

$\Rightarrow$ Reject H_0.

Yes, there is significant difference between the two hotels.

Part IV. (15 points) Recently, it has been found that the bride (=gelin) and the groom (=damat) are at the same age in 7% of a random sample of 65 married couples in Turkey.

1. (9 points) A historian thinks that the same age marriages have been about 5% over the last century based on her previous research. Is there significant evidence that such marriages have increased recently? Answer by showing all steps of a hypothesis test and stating your conclusion in plain English.

$H_0: p = 5\%$

$H_a: p > 5\%$

$SE = \sqrt{\frac{(0.05)(0.95)}{65}} = 0.027 \cong 2.7\%$

$z = \frac{7\% - 5\%}{2.7\%} = 0.74 \xrightarrow{\text{Normal table}} 54.67\%$

$\Rightarrow$ P-value $= \frac{100 - 54.67\%}{2} \cong 22.67\% > 5\%$

Do not reject H_0.

Same age marriages have not significantly increased.

2. (6 points) According to your conclusion in part a), consider the true percentage in the population to be 5% or 7%. If you did not do part a), choose either one you like. In a new random sample of only 4 couples, what is the probability that there are two or more couples of the same age?

$\Rightarrow p = 5\%$ by part a)

Prob. of 0 or 1 couple : $\binom{4}{0}(0.05)^0(0.95)^4 \cong 0.815$

$+ \binom{4}{1}(0.05)(0.95)^3 \cong 0.172$

P(there are two or more couples) $= 1 - 0.815 - 0.172$

$= 0.014 = 1.4\%$

Part V. (15 points) A common core course is taken by a mixed group of students. The following table shows data for a class of 50 students, who can be categorized as qualitative and quantitative majors.

	less than 40	between 40 and 80	greater than 80	Totals
Quantitative	10 10.8	4 5.94	13 10.26	27
Qualitative	10 9.2	7 5.06	6 8.74	23
Totals	20	11	19	50

1. (5 points) What are the variables? What are their values?

Major and grade are the variables.
The values of them are "Qualitative" and "Quantitative" for "major", and "less than 40", bet. 40 and 80" and "greater than 80" for "grade.

2. (10 points) Is the grade distribution independent from the type of major? Test at 1% level of significance by showing all steps and stating your conclusion in plain English.

H_0: independent
H_a: not indep.

$\frac{(20)(27)}{50} = 10.8$ $\quad \frac{(11)(27)}{50} = 5.94$ and so on...

$$\chi^2 = \frac{(10-10.8)^2}{10.8} + \frac{(4-5.94)^2}{5.94} + \frac{(13-10.26)^2}{10.26} + \frac{(10-9.2)^2}{9.2} + \frac{(7-5.06)^2}{5.06} + \frac{(6-8.74)^2}{8.74}$$

$$= 3.088$$

degrees of freedom $= (2-1)(3-1) = 2$

χ^2 table $\Rightarrow$ $10\% < P\text{-value} < 30\%$

$\Rightarrow$ $P\text{-value} > 1\%$ (OR $\chi^2(1\%) = 9.21$ and $3.65 < 9.21$)

Fail to reject the null.
Grade and major are independent

Part VI. (15 points) Starting salaries of graduates of three different majors of Koc University are given in the following table for 2007, where the numbers are simplified by changing each unit to 100 YTL. For example the salary 12 corresponds to 12x100=1200YTL. Each group below is a random sample from the corresponding major. Are the starting salaries of the three majors different on the average? Use the simplified numbers for your analysis.

n : 5 6 8

Major A	Major B	Major C
12	15	9
14	30	14
9	13	9
8	10	8
23	11	11
	8	8
		11
		12
Mean ($\bar{x}$) 13.2	14.5	10.3
Standard Deviation (s) 6	8	2.1
Variance (s^2) 36	64	4.4

$k=3$

$N=19$

Hint: The average of all 19 of the salaries is 12.4 (x100 YTL=1200YTL)

H_0: $\mu_A = \mu_B = \mu_C$

H_a: At least one mean is different from the others.

$$s_B^2 = \frac{5(13.2-12.4)^2 + 6(14.5-12.4)^2 + 8(10.3-12.4)^2}{3-1}$$

$$= 32.47$$

$$s_W^2 = \frac{4(36) + 5(64) + 7(4.4)}{19-3} = 30.925$$

$$\Rightarrow F = \frac{32.47}{30.925} \cong 1.05 \qquad F_{2,16} = 3.63 \text{ from F-table for } \alpha = 0.05$$

Since $1.05 < 3.63$, do not reject H_0.

$\Rightarrow$ The three majors' average starting salaries are the sam