Elvan Ceyhan

January 18, 2011

# Math 202: Statistics for Social Sciences

## Fall 2010 Final Exam

**Calculator allowed, duration 2 hours and 15 minutes.**

Instructions: There are eight problems in this exam. Please inspect the exam and make sure you have all 7 pages (6 pages of questions and 1 cover page). 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)

| | |
|---|---|
| Prob 1: | /17 |
| Prob 2: | /13 |
| Prob 3: | /8 |
| Prob 4: | /10 |
| Prob 5: | /10 |
| Prob 6: | /12 |
| Prob 7: | /16 |
| Prob 8: | /14 |
| Total: | /100 |

Formulas:

(1) Confidence interval (CI) : $\bar{X} \mp z \times SE$ or $\bar{X} \mp t \times SE$

with $SE = \dfrac{SD}{\sqrt{n}}$ or $\hat{p} \mp z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ or

$(\hat{p}_1 - \hat{p}_2) \mp z \times SE_{difference}$ or $(\bar{X}_1 - \bar{X}_2) \mp z \times SE_{difference}$

or $(\bar{X}_1 - \bar{X}_2) \mp t \times SE_{difference}$

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

and $SE = \sqrt{\dfrac{p(1-p)}{n}}$ 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:

$$\chi^2 = sum\left(\frac{(observed\ frequency - expected\ frequency)^2}{expected\ frequency}\right)$$

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

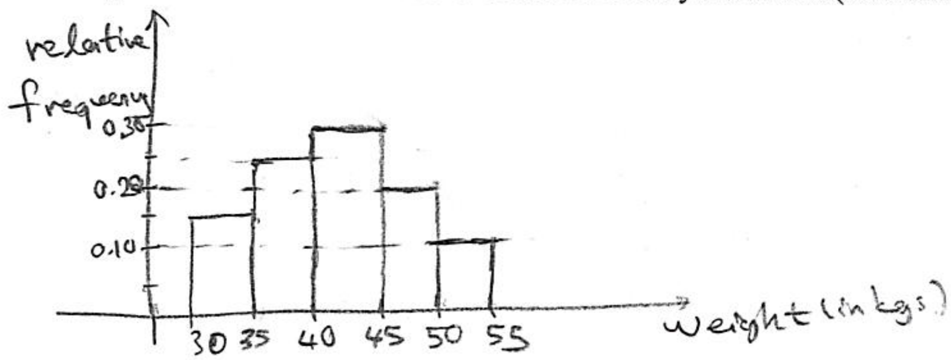
Within group variance: $s_W^2 = \dfrac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)}$ and $F = \dfrac{s_B^2}{s_W^2}$

Here $s_i$ is $SD$ of $i^{\text{th}}$ 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.

**Problem 1:** (17 points)
The weights (in kgs) of a sample of 100 seven-year-old females are given in the Table below. Each interval includes the right endpoint, not the left.

| Interval | # of people | Relative Frequency | Percentage |
|---|---|---|---|
| 30-35 | 15 | 0.15 | 15% |
| 35-40 | 25 | 0.25 | 25% |
| 40-45 | 30 | 0.30 | 30% |
| 45-50 | 20 | 0.20 | 20% |
| 50-55 | 10 | 0.10 | 10% |
| Total | 100 | 1.00 | 100% |

(a) (4 points) Perform the necessary computations in the above table and plot the relative frequency histogram below. Scale and label both axes clearly. Neatness (düzenlilik) is important.




(b) (2 points) What percent of seven-year-old females in the study weighed more than 45 pounds?

$20\% + 10\% = 30\%$

(c) (11 points) The percentages of all seven-year-old females were 20% for all weight groups above in the census two years ago. Is there a significant change in the weight percentages now (compared to the census data)? Use α=5%.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Interval | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 |
| Observed | 15 | 25 | 30 | 20 | 10 |
| Expected | 20 | 20 | 20 | 20 | 20 |

$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = 20\%$ vs $H_a$: at least one percentage is different.

$$\chi^2 = \frac{(15-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(30-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(10-20)^2}{20}$$

$= 12.5$ $\quad df = 5-1 = 4 \quad \chi^2_{crit} = 9.49$

Since $\chi^2 > \chi^2_{crit}$, we reject Ho and conclude that weight percentages are significantly different now compared to the census data.

**Problem 2:** (13 points)

(a) (4 points) Ten people in a lift (asansör) have an average weight of 75 kg. Two persons, one is 65 kg and the other is 89 kg, enter the lift at the third floor (kat). Find the average weight of all 12 people in the lift. (Keep at least one decimal in your final answer).

For $n=10$, $\bar{x}=75$, so sum of the weights of the first ten people is $10\times75=750$ kg

sum of the 12 people is $750+65+89=904$

average of 12 people is $\frac{904}{12}=75.33$ kg

(b) (4 points) Suppose average of a data set is 200 and the standard deviation is 24. Then, 25 is added to each of the numbers in the data set. Next, each value (obtained after addition) is divided by 6. What are the new average and standard deviation?

old: average $= 200$, SD $= 24$

$\frac{(\text{each value}+25)}{6}$ $\Rightarrow$ new: average $= \frac{200+25}{6} = 37.5$

SD $= \frac{24}{6} = 4$

(c) (2 points) Circle the correct answer or answers that complete the following sentence correctly:

The chance error __________________________________

I-goes away after you calibrate (ayarlamak) your measuring equipment.

II- shifts (kaydırmak) all observations in the same direction.

III- occurs at random moments of time: some observations will have it, some not.

IV- is equal to the standard deviation of a sequence of repeated measurements.

V- occurs on either side of the true value.

The subjects did not know which pill they are taking.

(d) (3 points) It is believed that an herbal extract (bitkisel özüt) improves concentration and memory. The following experimental study was designed to check this claim. The subjects were 250 healthy volunteers over 50 years old, with 100 women and 150 men. 50 of the women and 75 of the men were randomly assigned to take the herbal extract pill, while the remaining women and men were assigned to take a placebo pill. All subjects took a series of tests for learning and memory before treatment started and again after six weeks. Is there any element of a good experiment missing from this study? If so what elements? In either case explain your answer.

All the elements of a good experiment are present. (or one might argue that people who give the pills should also be blinded).

**Problem 3**: (8 points)

A box of numbered tickets has an average of 100 and a SD of 20. Four hundred draws will be made at random with replacement from the box.

(a) (4 points) Estimate the chance that the average of the draws will be in the range (i.e., within) 80 to 120.

Average = 100
SD = 20

For $n=400$, $\bar{X}$ is approx. $N\left(100, \frac{20}{\sqrt{400}}\right)$
$= N(100, 1)$

$P(80 < \bar{X} < 120) = P\left(\frac{80-100}{1} < z < \frac{120-100}{1}\right)$
$= P(-20 < Z < 20) \cong 100\%$

(b) (4 points) Estimate the chance that the average of the draws will be in the range 99 to 101.

$P(99 < \bar{X} < 101) = P\left(\frac{99-100}{1} < z < \frac{101-100}{1}\right)$
$= P(-1 < z < 1) = 68.27\%$

**Problem 4:** (10 points)

A recent research report states that 25% of all employed statisticians are academicians (i.e., working at a university). Suppose that a random sample of 200 statisticians is taken and it is found that 62 of them work in academia. Based on this data, do you believe in the recent research? report?

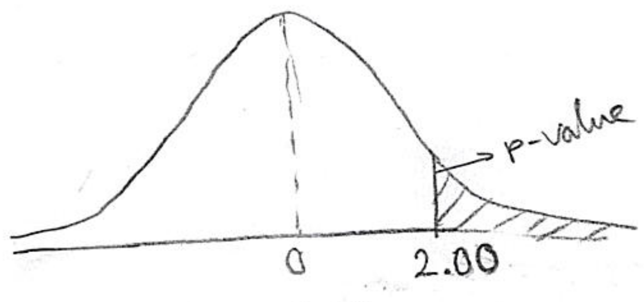
$n = 200$, $\hat{p} = \frac{62}{200} = 31\%$

$H_0: p = 25\%$
$H_a: p > 25\%$

n is large, so use ~~z test~~.

$z = \frac{\text{obs-exp}}{SE} = \frac{31-25}{3} = 2.$

$SE = \sqrt{\frac{0.25(0.75)}{200}}$
$= 0.03$ or $3\%$



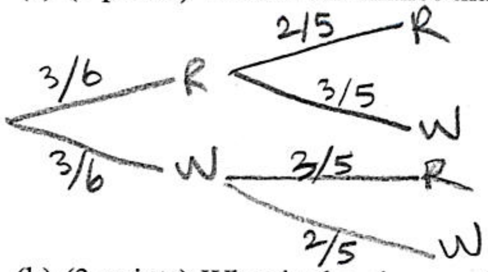

$\text{p-value} = \frac{100-95.45}{2} = \frac{4.55}{2} = 2.275\%$

Since p-value < 5%, we reject Ho and conclude that the report seems to be inaccurate.

**Problem 5**: (10 points)

A box contains 3 red and 3 white balls. Two balls will be drawn at random without replacement.

(a) (4 points) What is the chance that both balls will of the same color?




chance of same color = " " RR + WW

$= \frac{3}{6} \times \frac{2}{5} + \frac{3}{6} \times \frac{2}{5} = \frac{12}{30} = \frac{2}{5}$

(b) (3 points) What is the chance of both balls being red given that both balls are of the same color?

$P(RR \mid \text{same color}) = \frac{P(RR, \text{same color})}{P(\text{same color})} = \frac{P(RR)}{P(\text{same})} = \frac{1/5}{2/5} = \frac{1}{2}$

(c) (3 points) Are the events "both balls are of the same color" and "both balls are red" independent? Why?

$P(RR \mid \text{same color}) = \frac{1}{2} \neq \frac{2}{5} = P(\text{same color})$

**Problem 6**: (12 points)

Two Sections are taking Math 201: Statistics. Section 1 has 35 students who have an average of 65 with a standard deviation of 5 in MT1 (midterm 1). Section 2 has 45 students who have an average of 67 with a standard deviation of 8 in MT1. The MT1 score distributions for the two classes are similar in shape. Is it really the case that Section 2 performs better in MT1 of Math 201, or is it just chance variation? *Answers without proper explanation will receive no credit.*

| Section 1 | Section 2 |
|---|---|
| $n_1 = 35$ | $n_2 = 45$ |
| mean $= 65 = \bar{X}_1$ | mean $= 67 = \bar{X}_2$ |
| SD $= 5$ | SD $= 8$ |

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 < 0$

$n_1, n_2$ both large, so use z-test.

$z = \frac{\bar{X}_1 - \bar{X}_2}{SE_{diff}} = \frac{65 - 67}{1.46}$

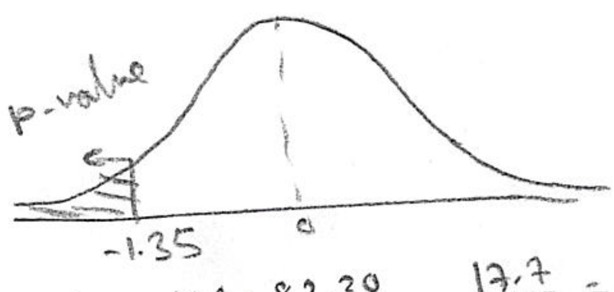
$= -1.36 \rightarrow -1.35$

$SE_1 = \frac{5}{\sqrt{35}} = 0.85$

$SE_2 = \frac{8}{\sqrt{45}} = 1.19$

$\Rightarrow SE_{diff} = \sqrt{0.85^2 + 1.19^2}$

$= 1.46$




p-value $= \frac{100 - 82.30}{2} = \frac{17.7}{2} = 8.8\%$

Since p-value $> 5\%$, we do not reject Ho and conclude that the difference is due to chance variation.

**Problem 7:** (16 points) Ministry of Labor (Çalışma Bakanlığı) wants to know how a new government policy is perceived (algılanmak) among the workers and whether the perception differs with union (sendika) membership. For this purpose, a random sample of 100 blue-collar (mavi yakalı) workers is surveyed about their views on this policy. Below is the breakdown of their proportions in regard to their views and union membership status.

| | Views on government policy | | | |
|---|---|---|---|---|
| Membership status | Supportive | Indifferent | Opposed | Row totals |
| Union | 0.07 | 0.09 | 0.28 | 0.44 |
| Non-union | 0.18 | 0.17 | 0.21 | 0.56 |
| Column totals | 0.25 | 0.26 | 0.49 | 1.00 |

(a) (12 points) Is membership status and views on government policy independent? Test at 5% level.

Out of 100 workers,
obs. frequencies

| | | | |
|---|---|---|---|
| 7 | 9 | 28 | 44 |
| 18 | 17 | 21 | 56 |
| 25 | 26 | 49 | 100 |

exp. frequencies

| | | |
|---|---|---|
| 11 | 11.44 | 21.56 |
| 14 | 14.56 | 27.44 |

$25 \times \frac{44}{100} = 11$, $26 \times \frac{44}{100} = 11.44$ and so on.

$H_0$: union membership & view are independent

$H_a$: " " " dependent

$$\chi^2 = \frac{(7-11)^2}{11} + \frac{(9-11.44)^2}{11.44} + \frac{(28-21.56)^2}{21.56} + \frac{(18-14)^2}{14} + \frac{(17-14.56)^2}{14.56} + \frac{(21-27.44)^2}{27.44} = 6.95$$

$df = (2-1)\times(3-1) = 2$

$\chi^2_{crit} = 5.99$

Since $\chi^2 > \chi^2_{crit}$, we reject $H_0$ and conclude that union membership status and views on government policy are dependent.

(b) (4 points) If a worker is selected randomly, what is the chance that s/he is a union member given that s/he is not opposed?

$P(\text{union member} \mid \text{not opposed}) = \frac{16}{51} = 0.31$ or 31%

**Problem 8:** (14 points) At Yellowshire University, three sections are taking Math 202 from the same instructor. The following statistics are for the final exam scores of these three sections.

| | Section 1 | Section 2 | Section 3 |
|---|---|---|---|
| Sample size | 21 | 21 | 21 |
| Mean | 50 | 60 | 50 |
| SD | 8 | 9 | 11 |

(a) (2 points) What is the variable of interest?

final exam score

(b) (12 points) Are the average final exam scores significantly different for the three sections? Use $\alpha$=0.05.

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_a$: at least one $\mu$ is different

$$\bar{X}_{GM} = \frac{21\times 50 + 21\times 60 + 21\times 50}{21+21+21} = \frac{3360}{63} = 53.33$$

$$s_B^2 = \frac{21\times(50-53.33)^2 + 21\times(60-53.33)^2 + 21\times(50-53.33)^2}{3-1}$$

$$= \frac{232.87 + 934.27 + 232.87}{2} = \frac{1400}{2} = 700$$

$$s_W^2 = \frac{(21-1)\times 8^2 + (21-1)\times 9^2 + (21-1)\times 11^2}{63-3}$$

$$= \frac{1280 + 1620 + 2420}{60} = \frac{4900}{60} = 88.67$$

$$F = \frac{s_B^2}{s_W^2} = \frac{700}{88.67} = 7.89$$

$df_N = 2$, $df_D = 60$ $\quad F_{crit} = 3.15$

Since $F > F_{crit}$, we reject $H_0$ and conclude that final exam scores are significantly different for the three sections.