Elvan Ceyhan November 10, 2010

Math 202: Statistics for Social Sciences

Fall 2010 Midterm 1

Calculator allowed, duration 90 minutes.

Instructions: There are six problems in this exam. Please inspect the exam and make sure you have all 6 pages (5 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 it.

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 and SURNAME: KEY

Formulas:

1) To find SD, sum all (entry-average)^2 values, then divide by n, and then take the square root.

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

Prob 1:	/20
Prob 2:	/20
Prob 3:	/20
Prob 4:	/20
Prob 5:	/12
Prob 6:	/8
Total:	/100

Problem 1: (20 points)

(a) (6 points) The distribution of the Math 100 scores is roughly symmetric and bell-shaped. The average Math 100 score is 75 points and the standard deviation is 10 points. Tamer did really badly on Math 100. Only 2.5% of students did as bad as or worse than him. What is Tamer's Math 100 score? Answer this question using the Empirical Rule.

Average = 75, SD = 10

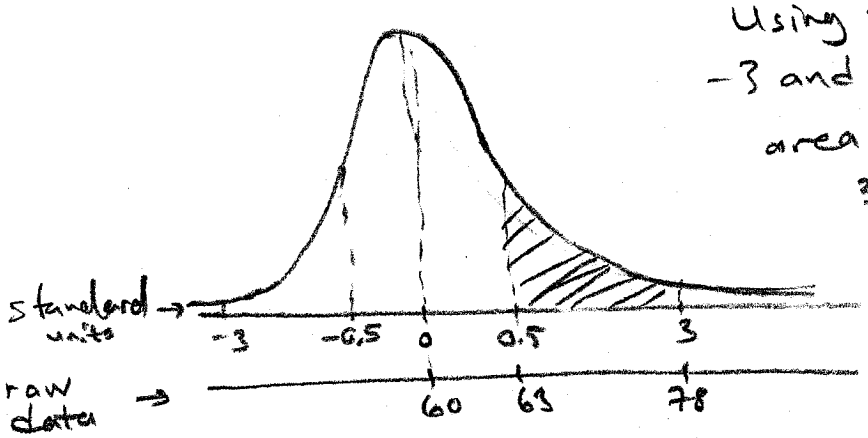
Empirical Rule says that about 95% of the test takers score within 2 SD of the average.

So, $75 \mp 2 \times 10 = 75 \mp 20 = (55, 95)$.

Hence about 2.5% score less than 55 and 2.5% score higher than 95. So Tamer's score is 55.

(b) (6 points) Suppose you know that your data is approximately bell-shaped and the average is 60 with a standard deviation of 6. Approximately how much of the data falls between 63 and 78?

Average = 60, SD = 6 $\quad \frac{63-60}{6} = 0.5 \quad \frac{78-60}{6} = 3$



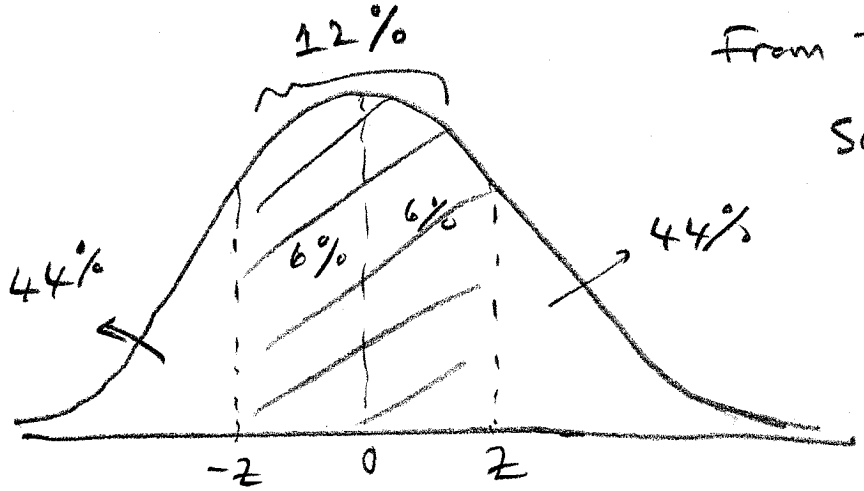
Using the table, the area between -3 and 3 is 99.73% and the area between -0.5 and 0.5 is 38.29%. Hence the shaded area is

$$\frac{99.73 - 38.29}{2} = \frac{61.44}{2}$$

$$= 30.72\ \%$$

(c) (8 points) The average survival time after meningitis (menenjit) is 100 months with a standard deviation of 20 months. The histogram of survival times looks approximately symmetric and bell-shaped. Estimate the 44th percentile of the survival time distribution.

Average = 100, SD = 20 $\quad$ 44th percentile?



From the table $z = 0.15$

So 44th percentile is

$100 - (0.15)(20)$

$= 100 - 3 = 97$ months

Problem 2: (20 points)
The following data are on the number of hours spent studying for a test by 6 randomly chosen Koç students: 12, 6, 2.5, 20, 9.5, 4

sorted data = 2.5, 4, 6, 9.5, 12, 20

(a) (4 points) Find the mean.

$$\text{Mean} = \frac{2.5+4+6+9.5+12+20}{6} = \frac{54}{6} = 9$$

(b) (4 points) Find the median.

median = Average of third and fourth values

$$= \frac{6+9.5}{2} = 7.75$$

(c) (5 points) Find the standard deviation.

$$SD = \sqrt{\frac{(2.5-9)^2+(4-9)^2+(6-9)^2+(9.5-9)^2+(12-9)^2+(20-9)^2}{6}}$$

$$= \sqrt{\frac{42.25+25+9+0.25+9+121}{6}} = \sqrt{\frac{206.5}{6}} = \sqrt{34.42}$$

$$= 5.87$$

(d) (3 points) The first and third quartiles are given to be $Q_1 = 4$ and $Q_3 = 12$. Find the interquartile range.

$$IQR = Q_3 - Q_1 = 12 - 4 = 8$$

(f) (4 points) The data converted into minutes (recall that 1 hour is 60 minutes ☺).
What is the average studying time in minutes? 540

$9 \times 60 = 540$

What is the standard deviation in minutes? 352.2

$5.87 \times 60 = 352.2$

Problem 3 (20 points)
A study monitored men from 1990 to 2000 and found that those who were moderate cigarette smokers (up to 6 cigarettes per day) were at an increased risk to develop heart disease over non-smokers. Please answer the following questions. Explain your answer where it is required.

(a) (2 points) Was this an observational or experimental study? Explain.
Observational study, because researchers did not assign subjects to treatments but only observed cigarette smokers and non-smokers.

(b) (2 points) What is the population of interest?
All men

(c) (2 points) What is the (experimental) unit?
a man

(d) (3 points) What are the treatments? How many treatments are in the study?
two treatments: - smoke cigarettes
- do not smoke cigarettes

(e) (3 points) Is there a control group in this study? If so, what treatment does it receive?
Yes, a group who did not smoke cigarettes.

(f) (3 points) What is the response variable? What values does it take?
heart disease: Yes or No

(g) (3 points) Suggest at least one possible confounding variable. Explain your reasoning.
Age. older men are more likely to have heart disease compared to younger men. (other answers are also possible!)

(h) (2 points) Can we say that smoking cigarettes causes heart disease? Why or why not?
No, observational studies can not establish causation.

Problem 4: (20 points)
A recent research report states that 25% of all employed statisticians are academicians (=working in a university).

$p = 25\%$

(a) (3 points) In a random sample of 1000 employed statisticians, how many of them are expected to be academicians?

$1000 \times 0.25 = 250$

(b) (6 points) Do you agree with the following sentences? Why or why not?

I- The percentage of academicians in a sample of size 100 (i.e., 100 statisticians) will be exactly 25%

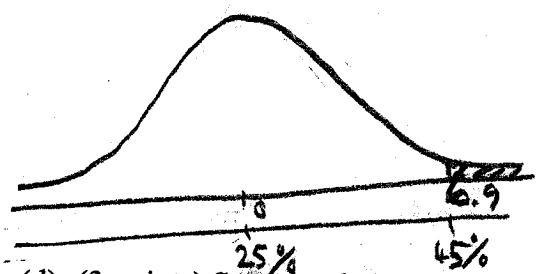
No, because there is chance error.

II- The expected value of the percentage of academicians in a sample of size 100 is 25%, give or take (plus-minus) 4%.

No, because expected value is 25% and is fixed.

(c) (8 points) What are the chances that the percentage of academicians in a random sample of 225 employed statisticians is greater than 45?

$SE = \sqrt{\frac{0.25(0.75)}{225}} = 0.0288 \cong 2.9\%$



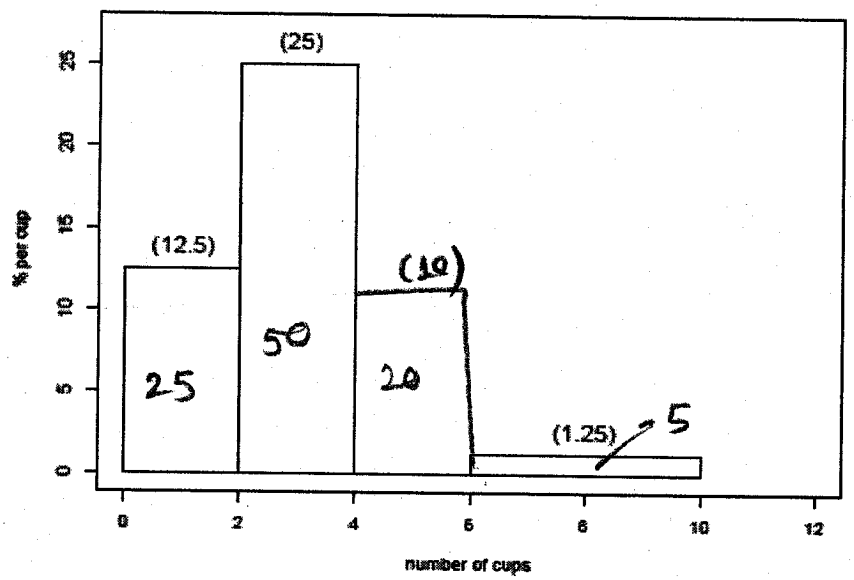
$\frac{45-25}{2.9} = 6.9$

The area between -6.9 & 6.9 is very close to 1. So the shaded area (hence the chances) are approximately 0.

(d) (3 points) Suppose that a random sample of 200 statisticians is taken and it is found that 62 of them work in academia. Fill in the blanks:
For the number of academicians in the sample, the observed value is 62,
but the expected value is 50.

$200 \times 0.25 = 50$

Problem 5: (12 points)
In a study, a density histogram was plotted showing the number of cups of tea per day had by each Koç University student. The density is marked in the parenthesis above the bars. The class intervals include the left endpoint but not the right.



(a) (8 points)The bar over the class interval from 4 to 6 cups is missing. How high must it be?

$12.5 \times 2 = 25$
$25 \times 2 = 50$
$1.25 \times 4 = 5$
80%

100 - 80 = 20 % left for 4 to 6 cups.
Since the class interval has width=2, height should be 10.

(b)(4 points) The percentage who drank 6 or more cups of tea is

$1.25 \times 4 = 5\%$

Problem 6: (8 points)
Cemil is taking Math 202 and Müge is taking Math 201. Cemil got a 72 on the first exam and his class average was 65 with a standard deviation of 5. Müge got an 82 and her class average was a 74 with a standard deviation of 8. The score distributions for the two classes are similar in shape. Who has the better class standing and why? *Answers without explanation will receive no credit.*

Math 202
Average = 65
SD = 5
Cemil : 72
Standard units: 1.4

$\frac{72-65}{5} = 1.4$

Math 201
Ave: 74
SD = 8
Müge: 82
std units = 1

$\frac{82-74}{8} = 1$

So Cemil has a better class standing because his class standing in standard units is higher.