**Part I. (20 points)** The first study on liver cancer had found association between liver cancer and smoking. The data were obtained from the medical records of the patients of a certain hospital.

1. (5 points) Is this an experiment or an observational study? Explain in three sentences at most.

This is an observational study because the subjects naturally assigned themselves to groups, not the investigators. The groups are "smoking" and "nonsmoking" subjects.

2. (6 points) What are the variables in this study?

Condition of smoking.
Having liver cancer or not

3. (5 points) It has been found in a second study that "drinking is associated with smoking, and alcohol causes liver cancer". Explain in two sentences at most why alcohol consumption is a confounding variable for the first study described in the question above.

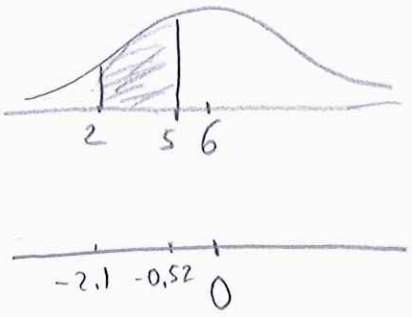
In the first study, both smoking habits and cancer presence are affected by this third variable (most drinkers also smoke). So, the real cause of cancer, namely, alcohol consumption is a confounding variable and the association of smoking with cancer is coming from that.

4. (4 points) How would you control the confounding variable, if you were the investigator in this study? Explain in two sentences at most.

We would group the data according to the confounding variable. We can then analyze as before, but within those who consume alcohol, and separaretely within those who do not.

**Part II. (20 points)** A study of college students nationwide found that the mean hours of sleep students get the night before an exam is 6 hours, with standard deviation 1.9 hours. The sleep distribution is found to be approximately normal.

1. (8 points) What are the chances that a randomly selected student gets between 2 and 5 hours of sleep before an exam?



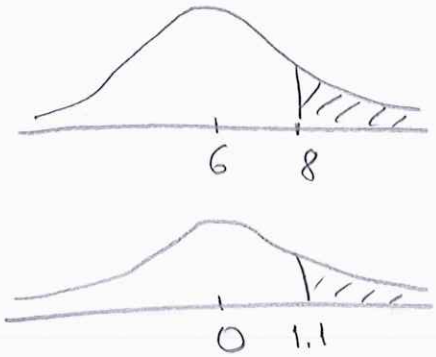

$\frac{2-6}{1.9} = -2.1 \qquad \frac{5-6}{1.9} = -0.52$

$\downarrow$

$\frac{96.43\% - 38.29\%}{2} = \frac{58.14\%}{2}$

$= 29.07\%$

2. (5 points) In this age group, adequate sleep is considered to be at least 8 hours. What percent of college students get adequate sleep?

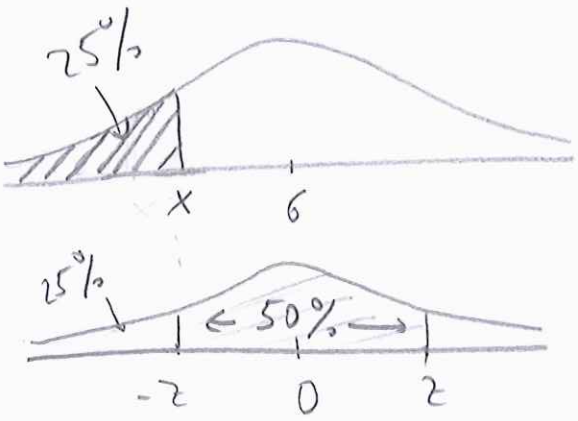



$\frac{8-6}{1.9} \approx 1.1$

$\frac{100 - 72.87\%}{2} = \frac{27.13\%}{2}$

$= 13.56\%$

3. (7 points) Find the 25th percentile of the sleep distribution.




$z = 0.67$ (Area $\approx 50\%$)

$\Rightarrow \frac{x-6}{1.9} = -0.67$

$\Rightarrow x = 4.73$

**Part III. (25 points)** The mean height in a random sample of people is found to be 168cm.

1. (2 points) If the histogram of the data set is approximately bell-shaped, about what percent of the observations fall within three standard deviations of the average?

99.7%

2. (4 points) What is the variable in this question? Is it continuous or discrete?

Height. It's continuous.

3. (6 points) A height of 165cm corresponds to -0.2 in standard units, in the sample of heights mentioned above. Find the standard deviation of the sample using this information.

$$\frac{165-168}{s} = -0.2$$

$$\Rightarrow s = \frac{-3}{-0.2} = 15 \text{ cm.}$$

4. (8 points) Find the standard deviation in the following sample of heights.
185, 140, 162, 170, 155, 178

$$\text{Average} = \frac{185+140+162+170+155+178}{6} = \frac{990}{6} = 165 \text{ cm.}$$

$$SD = \sqrt{\frac{(185-165)^2 + \ldots + (178-165)^2}{6}} = \sqrt{\frac{20^2+25^2+3^2+5^2+10^2+23^2}{6}} =$$

$$= 16.77$$

5. (5 points) Find the interquartile range in the data set of question 4 above.

140, 155, 162 | 170, 178, 185

$Q_1 = 155$ $\quad$ $Q_3 = 178$

$IQR = 178 - 155 = 23$ cm.

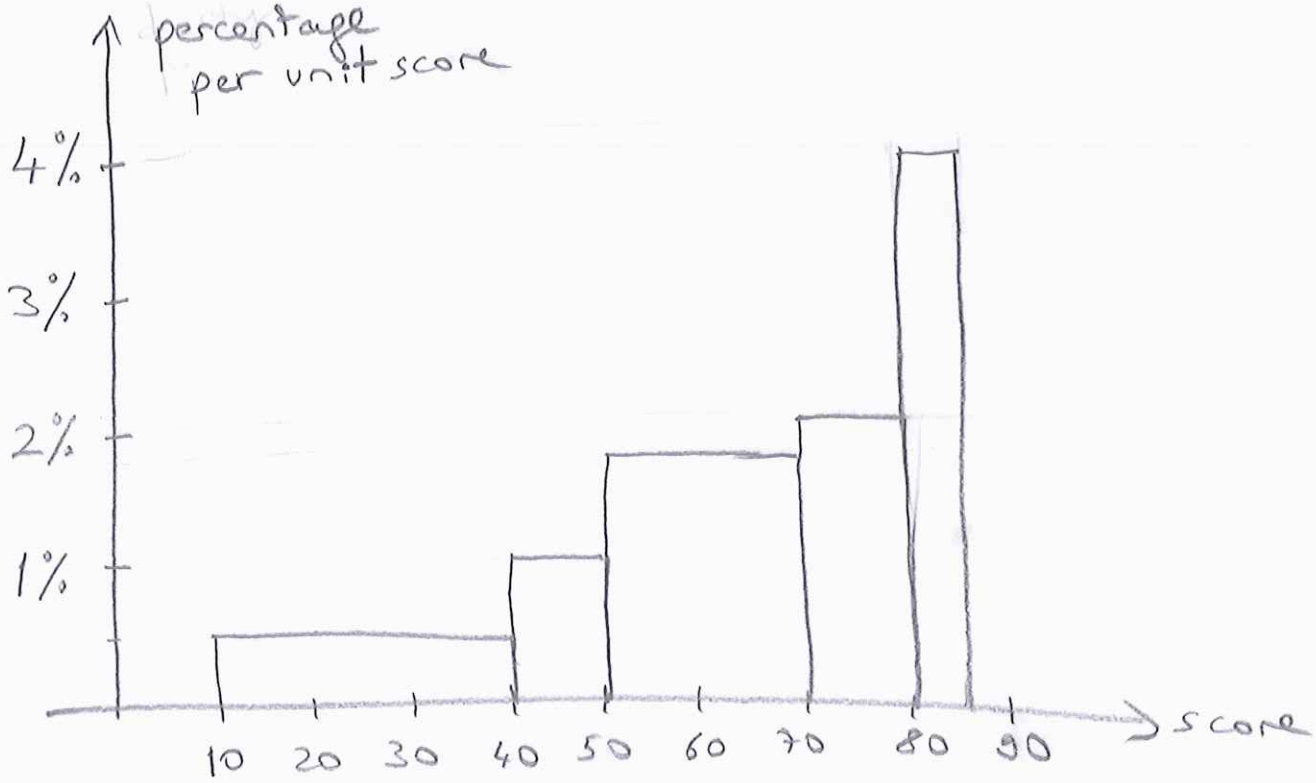
**Part IV. (20 points)** Consider the following data set of scores obtained from a sample of students at Koç University:

11.5 27.2 38.8 47.5 49.6 53.4 54.1 59.3 66.3 68.7 69.3 69.9
71.2 77.5 78.2 79.7 80.1 81.5 82.6 84.3

1. (5 points) Complete the following frequency table, where the left end points are not included and right end points are included for each class interval.

| | 10 - 40 | 40 - 50 | 50 - 70 | 70 - 80 | 80 – 85 |
|---|---|---|---|---|---|
| Frequency | 3 | 2 | 7 | 4 | 4 |
| Rel. freq. | 15% | 10% | 35% | 20% | 20% |
| Density | 0.5% | 1% | 1.75% | 2% | 4% |

2. (10 points) Draw a density scale histogram.




3. (5 points) The data set in the question was obtained as follows: Each college (Humanities, Science, Law, Engineering, etc.) is asked to select a smaller sample of students at random, and then, those samples are combined to form the data set of 20 observations above.

Is this data set a simple random sample? If yes, why? If not, describe how a simple random sample could be obtained.

No. This is cluster sampling. A simple random sample would be a sample where 20 people are randomly picked from the list of all students in the university

**Part V. (25 points)** 64 draws will be made at random from the box

| 12,000 [1]'s 48,000 [0]'s |
|---|

$\Rightarrow p = \frac{12000}{60000} = 20\%$

1. (6 points) Calculate the chance error for the percentage of 1's among the draws.

$$SE = \sqrt{\frac{(0.2)(0.8)}{64}} = 0.05 = 5\%$$

$\sqrt{(6.3)(0.8)}$ +2

2. (6 points) Are the following statements true or false? Explain in 1 sentence only.

T → explanation fine +2

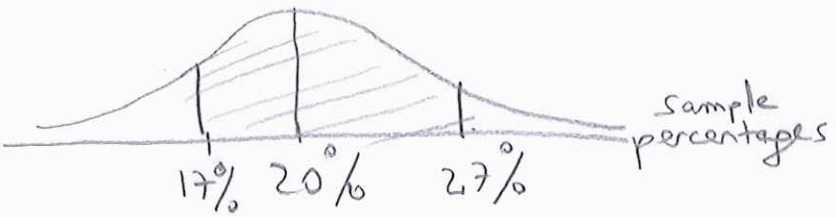
T: The expected value for the percentage of 1's among the draws is exactly 20%. exp^n 4

T - SE is 0.05

F: The expected value for the percentage of 1's among the draws is around 20%, give or take 5% or so.

Expected value is the population percentage of 1's, namely 20% and this is a fixed quantity.

3. (7 points) What is the probability that percentage of 1's among the draws is between 17% and 27%?




$SE = 5\%$

$\frac{17-20}{5} = -0.6$

$\frac{27-20}{5} = 1.4$

$\frac{45.15 + 83.85}{2}\% = 64.5\%$

4. (4 points) Suppose the draws are actually made, and 19 of the 64 draws are found to be 1's. How many standard errors is the percentage of 1's among the draws is away from its expected value?

$\hat{p} = \frac{19}{64} \approx 0.30$ $p = 0.20$

$z = \frac{0.30 - 0.20}{0.05} = 2$ ∴ $\hat{p}$ is about 2 SE above $p$.

5. (2 points) Is there (i) chance error, ii) systematic error or iii) both, in the percentage of 1's in the draws (in the sample) mentioned in Question 4? Explain in one sentence.

Chance error only because this is a simple random sample; draws are made at random.