Midterm

Aug, 2022

- You may not communicate with anyone, other than your instructor and TAs, regarding this exam. Thus, you may not communicate with your classmates.
- In completing this exam, you may consult the course notes, as well as the assigned HW, and posted material on Canvas.
- You may not consult the internet (for example, you may not look up, or post, questions on the internet).
- If you violate these instructions, in addition to other measures, you will receive a zero on the exam, and the zero will not be dropped.
- You must justify your work (legibly) to receive credit.
- If you have any questions, please e-mail *both* your instructor, and the TA, to ensure that you receive a prompt response, and to ensure that we respond consistently to your questions.

1. V1 Researchers studying the relationship between honesty, age, and self-control conducted an experiment on 160 children between the ages of 5 and 15. Participants reported their age, sex, and whether they were an only child or not. The researchers asked each child to toss a fair coin in private and to record the outcome (white or black) on a paper sheet, and said they would only reward children who report white. Half the students were explicitly told not to cheat and the others were not given any explicit instructions. In the "no instruction" group the probability of cheating was found to be the same across groups based on a child's characteristics. In the group that was explicitly told to not cheat, girls were less likely to cheat, and while rate of cheating didn't vary by age for boys, it decreased with age for girls (Ritz et al. 2000). Identify the following:

(a) Identify the main research question of the study. [**4 points**]
(b) Who are the subjects in this study, and how many are included? [**2 points**]
(c) How many variables were recorded for each subject in the study in order to conclude these findings? State the variables and their types. [**4 points**]

Solution: (a) "Does explicitly telling children not to cheat affect their likelihood to cheat?".
(b) 160 children between the ages of 5 and 15.
(c) Four variables: (1) age (numerical, continuous), (2) sex (categorical), (3) whether they were an only child or not (categorical), (4) whether they cheated or not (categorical).

1. V2 Suppose a group of researchers want to study the relationship between age and honesty. They conduct an experiment on 100 children between the age of 10 and 15. Before they enter the study, the children reported their age, sex, and whether they were an only child or not. They were asked roll a die and were told that they would be rewarded if they rolled an even number. The participants self reported the outcome. Half the children were explicitly told not to cheat and the others were not given any explicit instructions. In the group where there were no instructions given, the probability of dishonesty was found to be the same across. In the group that was explicitly told to not cheat, girls were less likely to cheat, and while rate of cheating didn't vary by age for boys, it decreased with age for girls. Identify the following:

(a) Identify the main research question of the study. [**4 points**]
(b) Who are the subjects in this study, and how many are included? [**2 points**]
(c) How many variables were recorded for each subject in the study in order to conclude these findings? State the variables and their types. [**4 points**]

Solution: (a)Does explicitly telling children not to cheat affect their likelihood to cheat?".
(b) 100 children between the ages of 10 and 15.
(c) Four variables: (1) age (numerical, continuous), (2) sex (categorical), (3) whether they were an only child or not (categorical), (4) whether they cheated or not (categorical).

2. V1 The following summary statistics give some information on the exam grades in a math class.

Count	50
Mean	78
Median	76
Std. dev.	8.20
Min	30
Max	95
Range	65
Q1	60.25
Q3	86.5

(a) Do you think this distribution is left skewed, right skewed or symmetric? Explain. [**5 points**]
Solution: This would be slightly left skewed since the mean is slightly less than the median while the 25th percentile is farther away from the median than the 75th percentile.
(b) Are there any outliers? [**5 points**]
Solution: $Q1 - 1.5 \times IQR = 60.25 - 1.5 \times (86.5 - 60.25) = 20.875$;
$Q3 + 1.5 \times IQR = 86.5 + 1.5 \times (86.5 - 60.25) = 125.875$;
There are no outliers since all data are within the fences.

2. V2 The following summary statistics give some information on the exam grades in a math class.

Count	100
Mean	80
Median	84
Std. dev.	5.4
Min	45
Max	99
Range	54
Q1	76
Q3	95

(a) Do you think this distribution is left skewed, right skewed or symmetric? Explain.

Solution: This would be slightly right skewed since the mean is slightly more than the median while the 75th percentile is farther away from the median than the 25th percentile. [**5 points**]

(b) Are there any outliers?

Solution: $Q1 - 1.5 \times IQR = 76 - 1.5 \times (95 - 76) = 47.5$;

$Q3 + 1.5 \times IQR = 95 + 1.5 \times (95 - 76) = 123.5$;

Since the minimum is $45 < 47.5$, there would be at least one outlier. [**5 points**]

3. V1 Suppose X and Y are independent random variables. If we know that $\mathbb{E}[X] = 10, \mathrm{SD}(X) = \sigma_X = 2, \mathbb{E}[Y] = 20, \mathrm{SD}(Y) = \sigma_Y = 3$;

what is the mean and standard deviation of

(a) $X + 3Y$ [**5 points**]

(b) $2X - Y - 6$ [**5 points**]

Solution: (a) $\mathbb{E}[X+3Y] = \mathbb{E}[X]+3\mathbb{E}[Y] = 70$; $\mathrm{Var}(X+3Y) = \mathrm{Var}(X)+9\,\mathrm{Var}(Y) = 85$; hence $\mathrm{SD}(X+3Y) = \sqrt{85}$.

(b) $\mathbb{E}[2X - Y - 6] = 2\mathbb{E}[X] - \mathbb{E}[Y] - 6 = -6$; $\mathrm{Var}(2X - Y - 6) = 4\,\mathrm{Var}(X) + \mathrm{Var}(Y) = 25$

3. V2 Suppose X and Y are independent random variables. If we know that $\mathbb{E}[X] = 40, \mathrm{Var}(X) = \sigma_X^2 = 16, \mathbb{E}[Y] = 20, \mathrm{Var}(Y) = \sigma_Y^2 = 9$;

what is the mean and standard deviation of

(a) $4X - 40$ [**5 points**]

(b) $X - Y$ [**5 points**]

Solution: (a) $\mathbb{E}[4X - 40] = 4\mathbb{E}[X] - 40 = 120$; $\mathrm{Var}(4X - 40) = 16\,\mathrm{Var}(X) = 256$; hence $\mathrm{SD}(4X - 40) = 16$.

(b) $\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] = 20$; $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 25$; hence $\mathrm{SD}(X - Y) = 5$.

3. V3 Suppose X and Y are independent random variables. If we know that $\mathbb{E}[X] = 100, \mathrm{Var}(X) = \sigma_X^2 = 144, \mathbb{E}[Y] = 150, \mathrm{Var}(Y) = \sigma_Y^2 = 169$;

what is the mean and standard deviation of

(a) $X - 5$ [**5 points**]

(b) $\frac{1}{4}X - Y$ [**5 points**]

Solution: (a) $\mathbb{E}[X - 5] = \mathbb{E}[X] - 5 = 95$; $\mathrm{Var}(X - 5) = \mathrm{Var}(X) = 144$; hence $\mathrm{SD}(X - 5) = 12$.

(b) $\mathbb{E}[\frac{1}{4}X - Y] = \frac{1}{4}\mathbb{E}[X] - \mathbb{E}[Y] = -125$; $\mathrm{Var}(\frac{1}{4}X - Y) = \frac{1}{16}\,\mathrm{Var}(X) + \mathrm{Var}(Y) = 178$; hence $\mathrm{SD}(\frac{1}{4}X - Y) = \sqrt{178}$

4. V1 The scores for a given test are approximately normally distributed with a mean of 100 and a standard deviation of 15, what is the (approximate) probability that an individual's score is

(a) above 90? [**5 points**]

(b) between 112 and 132? [**5 points**]

Solution: (a)

$$P(X > 90) = P(\frac{X - 100}{15} > \frac{90 - 100}{15}) \approx P(Z > -0.67) \approx 0.749$$

(b)

$$P(112 < X < 132) = P(\frac{112 - 100}{15} < Z < \frac{132 - 100}{15}) \approx P(0.8 < Z < 2.13) \approx 0.19$$

4. V2 Suppose a given test has its scores approximately normally distributed with a mean of 1500 and a standard deviation of 300. What is the (approximate) probability that a person's score is

(a) below 1600? [**5 points**]

(b) between 1200 and 1700? [**5 points**]

Solution: (a)

$$P(X < 1600) = P(\frac{X - 1500}{300} < \frac{1600 - 1500}{300}) \approx P(Z < 0.33) \approx 0.629$$

(b)

$$P(1200 < X < 1700) = P(\frac{1200 - 1500}{300} < Z < \frac{1700 - 1500}{300}) \approx P(-1 < Z < 0.67) \approx 0.749 - 0.159 = 0.59$$

4. V3 Suppose a given test has its scores approximately normally distributed with a mean of 18 and a standard deviation of 6. What is the (approximate) probability that a person's score is

(a) above 20? [**5 points**]

(b) in between 16 and 21? [**5 points**]

Solution: (a)

$$P(X > 20) = P(\frac{X - 18}{6} > \frac{20 - 18}{6}) \approx P(Z > 0.33) \approx 0.371$$

(b)

$$P(16 < X < 21) = P(\frac{16 - 18}{6} < Z < \frac{21 - 18}{6}) \approx P(-0.33 < Z < 0.5) \approx 0.691 - 0.371 = 0.32$$

5. Suppose a classroom has 20 students, and you know the following information about the students:

- 10 students have brown eyes.
- 8 students are left-handed.
- 3 students have brown eyes and are left-handed.

Define the following events:

- E: A randomly selected student has brown eyes.
- F: A randomly selected student is left-handed.

(a) Are E and F independent? Show your work. [**5 points**]

(b) If a student is randomly selected and found to be left-handed, what is the probability that this student does not have brown eyes? Show your reasoning. [**5 points**]

Solutions

(a) Two events E and F are independent if:

$$P(E \cap F) = P(E) \times P(F).$$

Calculate the probabilities:

$$P(E) = \frac{\text{Number of students with brown eyes}}{\text{Total students}} = \frac{10}{20} = 0.5.$$

$$P(F) = \frac{\text{Number of left-handed students}}{\text{Total students}} = \frac{8}{20} = 0.4.$$

$$P(E \cap F) = \frac{\text{Number of students with brown eyes and left-handed}}{\text{Total students}} = \frac{3}{20} = 0.15.$$

Compute $P(E) \times P(F)$:

$$P(E) \times P(F) = 0.5 \times 0.4 = 0.2.$$

Since $P(E \cap F) = 0.15 \neq 0.2 = P(E) \times P(F)$, the events E and F are **not independent**.

(b) We want to find $P(\overline{E} \mid F)$.

Number of left-handed students without brown eyes:

$$8 \text{ (left-handed)} - 3 \text{ (left-handed with brown eyes)} = 5.$$

Compute $P(\overline{E} \cap F)$:

$$P(\overline{E} \cap F) = \frac{5}{20} = 0.25.$$

Compute $P(F)$:

$$P(F) = 0.4.$$

Calculate the conditional probability:

$$P(\overline{E} \mid F) = \frac{P(\overline{E} \cap F)}{P(F)} = \frac{0.25}{0.4} = \frac{5}{8}.$$

The probability that a left-handed student does not have brown eyes is $\frac{5}{8}$ i.e. 0.625