# STA220 Midterm 1, Feb. 7, 2023

**(80 minutes; 5 questions; 7 pages; total points = 48)**

**Do not open this test until told to do so. Try to answer as many questions as you can.** Write down your name, email and student number at each page of the assignment. Aids allowed: a single non-programmable calculator, and course notes (both printed and handwritten). No devices are permitted. Please write your answers in the space provided, only. You should explain all of your solutions clearly. A full mark will not be given for an answer without explanation. Point values for each question are indicated in [square brackets].

**Academic Integrity Acknowledgement Form**

The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic misconduct, the processes for addressing academic offences, and the penalties that may be imposed. Potential offences include (but are not limited to):

- Looking at someone else's answers during the exam.
- Letting someone else look at your answers during the exam.
- Misrepresenting your identity or having someone else complete your exam.
- Sharing or posting the exam questions during or for 24 hours after the exam.

Prior to beginning this exam, you must affirm that you will follow the Code of Behaviour on Academic Matters in the completion of this exam, by completing the following statement:

I, ________________________________________ (print your full name), agree to fully abide by the Code of Behaviour on Academic Matters. I promise not to commit academic misconduct, and am aware that significant penalties may be imposed if I do.

Signature: ________________________________________

**1.** You decided to study the influence of exercising on sleep quality. You asked six of your friends how many times they did exercise last week. Unfortunately, you forgot to record one of your measurements, so you have an incomplete list of six measurements

$$\texttt{exercise} = 0, 4, 1, ?, 1, 5$$

**(a)** **[2]** Given that the mean is equal to 2, can you recover the missing measurement? If yes, what is the value?

Yes, you can.

$$\text{mean} = \frac{0+4+1+?+1+5}{6} = 2 \Rightarrow 11+? = 12 \Rightarrow \boxed{? = 1}$$

**(b)** **[2]** Given that the minimum is equal to 0 and maximum is equal to 5, can you recover the missing measurement? If yes, what is the value?

No, you cannot.

? can be any number between 0 and 5.

**(c)** **[2]** Given that the median is equal to 1.5, can you recover the missing measurement? If yes, what is the value?

Yes, you can. Let's sort the values 0 1 1 4 5

If $? \le 1$, e.g. $0\ ?\ \boxed{1\ 1}\ 4\ 5 \Rightarrow \text{median} = \frac{1+1}{2} = 1$ ✗

If $? \ge 4$, e.g. $0\ 1\ \boxed{1\ 4}\ 5\ ? \Rightarrow \text{median} = \frac{1+4}{2} = 2.5$ ✗

Thus $0\ 1\ \boxed{1\ ?}\ 4\ 5$ and $\text{median} = \frac{1+?}{2} = 1.5 \Rightarrow \boxed{? = 2}$

**(d)** **[2]** Given that the first quartile is equal to 1 and the third quartile is equal to 3.75, can you recover the missing measurement? If yes, what is the value?

Yes, you can.

If $n = 6 \Rightarrow$ positions for $Q_1$ and $Q_3$ are 2.25 and 4.75

$?\ \boxed{0\ 1}\ 1\ 4\ 5 \Rightarrow Q_1 = 0.25$ ✗

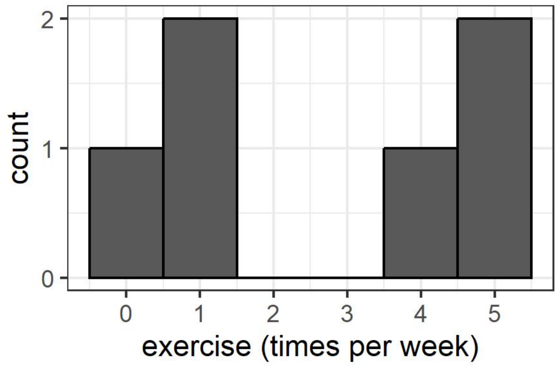
$0\ ?\ 1\ \boxed{1\ 4}\ 5$ or $0\ 1\ ?\ \boxed{1\ 4}\ 5 \Rightarrow Q_3 = 3.25$ ✗

$0\ 1\ 1\ \boxed{4\ ?}\ 5 \Rightarrow Q_3 = 4 + (?-4) \cdot 0.75 > 4$ ✗

$0\ 1\ 1\ \boxed{4\ 5}\ ? \Rightarrow Q_3 = 4.75$ ✗

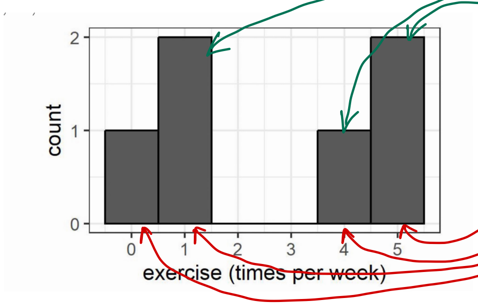
Thus $0\ \boxed{1\ 1}\ \boxed{?\ 4}\ 5 \Rightarrow Q_1 = 1$ and $Q_3 = ? + (4-?) \cdot 0.75 = 3.75 \Rightarrow \boxed{? = 3}$

**(e)** **[2]** Given the histogram, can you recover the missing measurement in `exercise`? If yes, what is the value?



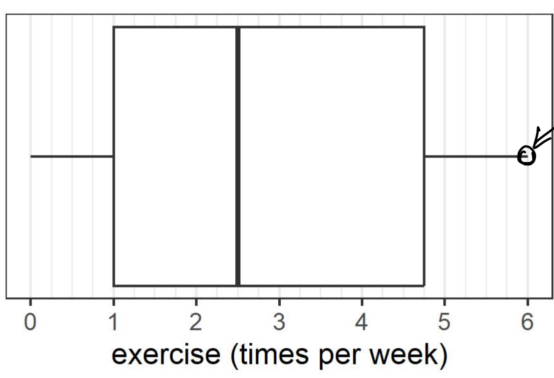

Yes, you can.
From the histogram:
0, 1, 1, 4, 5, 5 $\Rightarrow$ $\boxed{? = 5}$

**(f)** **[2]** How would you use this histogram to find the proportion of your friends who exercised last week? Explain your steps.




$$\frac{\text{Sum the heights}}{\text{sum the heigts}} = \boxed{\frac{5}{6}}$$

**(g)** **[2]** Given the boxplot, can you recover the missing measurement in `exercise`? If yes, what is the value?




Yes, you can.
This is whisker's end, which should be an observation from the data!
Thus $\boxed{? = 6}$

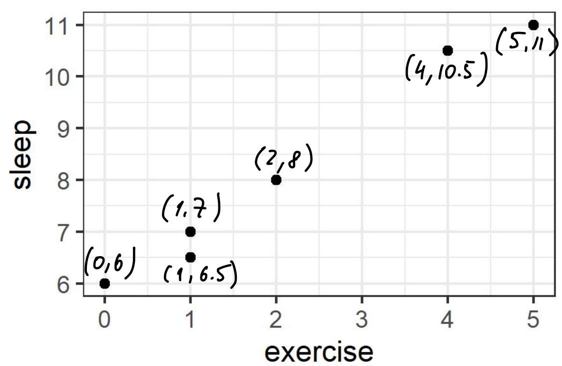
**(h)** **[2]** Given the same boxplot, can you recover the lower and upper fences of `exercise`?

Yes, you can.
From the boxplot $Q_1 = 1$ $Q_3 = 4.75$ $\Rightarrow$ $IQR = 3.75$
$LF = 1 - 1.5 \cdot 3.75 = \boxed{-4.625}$
$UF = 4.75 + 1.5 \cdot 3.75 = \boxed{10.375}$

**2.** Now you asked the same six friends (in the same order) how many hours on average did they sleep last week. Again, you forgot to record three of these measurements

$$\texttt{sleep} = 6, 10.5, ?, ?, 6.5, ?$$

**(a)** [**2**] Given the scatterplot, can you recover the missing measurement in `exercise`? If yes, what is the value?




| | | | | | | |
|---|---|---|---|---|---|---|
| exercise | 0 | 4 | 1 | ? | 1 | 5 |
| sleep | 6 | 10.5 | ? | ? | 6.5 | ? |

From the plot, the observations in exercise are 0, 1, 1, 2, 4, 5 $\Rightarrow$ $\boxed{? = 2}$

**(b)** [**2**] Given the same scatterplot, can you recover the missing measurements in `sleep`? If yes, what are the values?

| | | | | | | |
|---|---|---|---|---|---|---|
| exercise | 0 | 4 | 1 | 2 | 1 | 5 |
| sleep | 6 | 10.5 | ? | ? | 6.5 | ? |

From the plot,
exercise = 5 $\Rightarrow$ sleep = 11
exercise = 2 $\Rightarrow$ sleep = 8
exercise = 1 $\Rightarrow$ sleep = 6.5 or 7

$\Rightarrow$ sleep = 6, 10.5, $\boxed{7, 8}$, 6.5, $\boxed{11}$

**(c)** [**2**] Use the same scatterplot to answer the following question. Is the correlation coefficient between `exercise` and `sleep` close to 1, -1 or 0? What does the correlation coefficient say about the relationship between these variables?

Positive trend, close to linear $\Rightarrow$ $\boxed{\text{cor} \simeq 1}$
On average, increase in exercise implies increase in sleep.

**(d)** [**2**] What will happen to the correlation coefficient if you replace average hours of sleep per week by the total number of hours of sleep per week?

This is equivalent to considering sleep_total = sleep · 7.
Correlation is scale invariant $\Rightarrow$ it will $\boxed{\text{stay the same}}$

**3.** You roll two dice and consider two events (both about the same experiment!):

$A$ = you get five exactly in one roll; $B$ = the product of the two scores is divisible by 10.

**(a)** [**2**] What is the probability of A?

Two options : 1) first roll is 5 & second roll is 1, 2, 3, 4, 6
2) first roll is 1, 2, 3, 4, 6 & second roll is 5

$$P(A) = \frac{1}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{1}{6} = \frac{10}{36} = \boxed{\frac{5}{18}}$$

**(b)** [**2**] What is the probability of B?

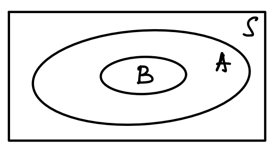
Two options : 1) first roll is 5 & second roll is 2, 4, 6
2) first roll is 2, 4, 6 & second roll is 5

$$P(B) = \frac{1}{6} \cdot \frac{3}{6} + \frac{3}{6} \cdot \frac{1}{6} = \frac{6}{36} = \boxed{\frac{1}{6}}$$

**(c)** [**2**] List all the outcomes from $A \cap B$.

$A \cap B = \{(5,2), (5,4), (5,6), (2,5), (4,5), (6,5)\}$

Actually, $A \cap B = B$




**(d)** [**2**] Are events A and B independent?

$$P(A \cap B) = \frac{1}{6} \neq P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{5}{18} \Rightarrow \boxed{\text{dependent}}$$

**4.** It is estimated that 20% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims:

(1) it can detect 99% of spam emails;

(2) the probability for a non-spam email to be detected as spam is 5%.

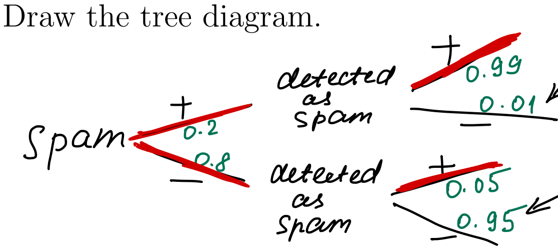
**(a)** [**2**] Restate (1) and (2) in terms of conditional probabilities. (Hint: use events $A$ = an email is spam; $B$ = an email is detected as spam)

$$P(\text{detect as spam} \mid \text{spam}) = P(B \mid A) = 0.99$$

$$P(\text{detect as spam} \mid \text{not spam}) = P(B \mid \bar{A}) = 0.05$$

**(b)** [**2**] Draw the tree diagram.

$$P(\bar{B} \mid A) = 1 - P(B \mid A) = 1 - 0.99$$




$$P(\bar{B} \mid \bar{A}) = 1 - P(B \mid \bar{A}) = 1 - 0.05$$

**(c)** [**2**] Use this diagram to compute the probability that an email is detected as spam.

$$P(B) = P(\text{detect as spam}) = P(B \cap A) + P(B \cap \bar{A}) =$$

$$= P(B \mid A) \cdot P(A) + P(B \mid \bar{A}) \cdot P(\bar{A}) = (\text{see red branches})$$

$$= 0.2 \cdot 0.99 + 0.8 \cdot 0.05 = \boxed{0.238}$$

**(d)** [**2**] Given an email is detected as spam, what is the probability that it is in fact a non-spam email? (Hint: express this question in terms of conditional probability first)

$$P(\bar{A} \mid B) = \frac{P(\bar{A} \cap B)}{P(B)} = \frac{P(B \mid \bar{A}) \cdot P(\bar{A})}{P(B)} = \frac{0.8 \cdot 0.05}{0.238} \approx \boxed{0.168}$$

**5.** A standard 52-card deck comprises 13 ranks in each of the four suits: clubs (♣), diamonds (♦), hearts (♥) and spades (♠). You repeat the following experiment four times:

(1) You pick one card at random from the deck.
(2) If the suit of your card is hearts, you win 5 dollars; otherwise, you win nothing.
(3) Then you return the card back to the desk and shuffle it.

As a result of this game you can win $0, 5, 10, 15$ or $20$ dollars.

**(a)** **[2]** The probability to win $x$ dollars is 0.2109. Find $x$.

$Y$ = number of victories $X$ = cash prize

$Y$ has Binomial distribution with $n = 4$ $p = 0.25$

$P(Y=k) = 0.2109 \Rightarrow k = 2$ (from the table)

If we won $Y$ rounds, we won $X = 5 \cdot Y$ in total $\Rightarrow \boxed{x = 10\,\$}$

**(b)** **[2]** Write down the distribution table for you cash prize.

| victories | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| prize | 0 \$ | 5 \$ | 10 \$ | 15 \$ | 20 \$ |
| probability | 0.3164 | 0.4219 | 0.2109 | 0.0469 | 0.0039 |

**(c)** **[2]** What is the expected value for you cash prize?

$$E(X) = \sum_x x \cdot P(X=x) = 0 \cdot 0.3164 + 5 \cdot 0.4219 + 10 \cdot 0.2109 + 15 \cdot 0.0469 + 20 \cdot 0.0039 = \boxed{5}$$

**(d)** **[2]** How will the distribution table change if we replace (2) by "If the suit of your card is hearts, you win 5 dollars; otherwise, you lose 5 dollars?" What will be the new expected value for the cash prize?

| victories | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| prize | −20 \$ | −10 \$ | 0 \$ | 10 \$ | 20 \$ |
| probability | 0.3164 | 0.4219 | 0.2109 | 0.0469 | 0.0039 |

$$EX = -20 \cdot 0.3164 - 10 \cdot 0.4219 + 0 \cdot 0.2109 + 10 \cdot 0.0469 + 20 \cdot 0.0039 = \boxed{-10}$$

**[END OF EXAMINATION; total points = 48]**

**Binomial Probability Table**

$n$=Number of trials, $k$=Number of successes and $p$=Probability of success

| | | $p$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | 0.01 | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 |
| 2 | 0 | 0.9801 | 0.9025 | 0.8100 | 0.7225 | 0.6400 | 0.5625 | 0.4900 | 0.4225 | 0.3600 | 0.3025 | 0.2500 |
| | 1 | 0.0198 | 0.0950 | 0.1800 | 0.2550 | 0.3200 | 0.3750 | 0.4200 | 0.4550 | 0.4800 | 0.4950 | 0.5000 |
| | 2 | 0.0001 | 0.0025 | 0.0100 | 0.0225 | 0.0400 | 0.0625 | 0.0900 | 0.1225 | 0.1600 | 0.2025 | 0.2500 |
| 3 | 0 | 0.9703 | 0.8574 | 0.7290 | 0.6141 | 0.5120 | 0.4219 | 0.3430 | 0.2746 | 0.2160 | 0.1664 | 0.1250 |
| | 1 | 0.0294 | 0.1354 | 0.2430 | 0.3251 | 0.3840 | 0.4219 | 0.4410 | 0.4436 | 0.4320 | 0.4084 | 0.3750 |
| | 2 | 0.0003 | 0.0071 | 0.0270 | 0.0574 | 0.0960 | 0.1406 | 0.1890 | 0.2389 | 0.2880 | 0.3341 | 0.3750 |
| | 3 | | 0.0001 | 0.0010 | 0.0034 | 0.0080 | 0.0156 | 0.0270 | 0.0429 | 0.0640 | 0.0911 | 0.1250 |
| 4 | 0 | 0.9606 | 0.8145 | 0.6561 | 0.5220 | 0.4096 | 0.3164 | 0.2401 | 0.1785 | 0.1296 | 0.0915 | 0.0625 |
| | 1 | 0.0388 | 0.1715 | 0.2916 | 0.3685 | 0.4096 | 0.4219 | 0.4116 | 0.3845 | 0.3456 | 0.2995 | 0.2500 |
| | 2 | 0.0006 | 0.0135 | 0.0486 | 0.0975 | 0.1536 | 0.2109 | 0.2646 | 0.3105 | 0.3456 | 0.3675 | 0.3750 |
| | 3 | | 0.0005 | 0.0036 | 0.0115 | 0.0256 | 0.0469 | 0.0756 | 0.1115 | 0.1536 | 0.2005 | 0.2500 |
| | 4 | | | 0.0001 | 0.0005 | 0.0016 | 0.0039 | 0.0081 | 0.0150 | 0.0256 | 0.0410 | 0.0625 |
| 5 | 0 | 0.9510 | 0.7738 | 0.5905 | 0.4437 | 0.3277 | 0.2373 | 0.1681 | 0.1160 | 0.0778 | 0.0503 | 0.0312 |
| | 1 | 0.0480 | 0.2036 | 0.3281 | 0.3915 | 0.4096 | 0.3955 | 0.3602 | 0.3124 | 0.2592 | 0.2059 | 0.1562 |
| | 2 | 0.0010 | 0.0214 | 0.0729 | 0.1382 | 0.2048 | 0.2637 | 0.3087 | 0.3364 | 0.3456 | 0.3369 | 0.3125 |
| | 3 | | 0.0011 | 0.0081 | 0.0244 | 0.0512 | 0.0879 | 0.1323 | 0.1811 | 0.2304 | 0.2757 | 0.3125 |
| | 4 | | | 0.0005 | 0.0022 | 0.0064 | 0.0146 | 0.0284 | 0.0488 | 0.0768 | 0.1128 | 0.1562 |
| | 5 | | | | 0.0001 | 0.0003 | 0.0010 | 0.0024 | 0.0053 | 0.0102 | 0.0185 | 0.0312 |