

MATH 203: Principles of Statistics 1

Fall 2024

Practice Midterm Exam

MATH 203: Specific Instructions.

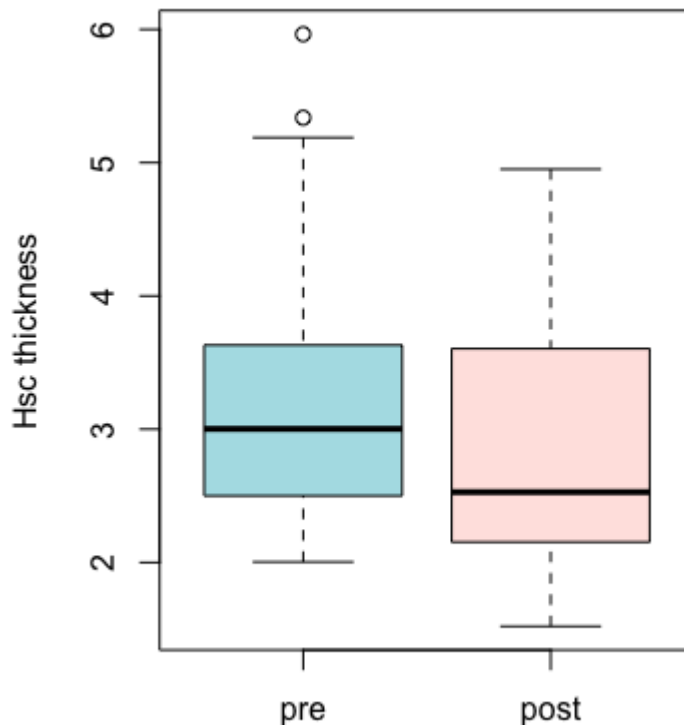
Please write your answers on the exam in the boxes provided for each question; only answers written inside the correct boxes will be marked. Please write clearly.

This paper contains four questions. Each question carries 10 marks. Credit will be given for all questions attempted. The total mark available is 40 but rescaling of the final mark may occur.

A formula sheet is provided. No crib sheet is allowed.

Only non-programmable and non text-storing calculators may be used. Dictionaries and Translation dictionaries are permitted.

1. **Randomized controlled trial of the immediate and long-term effect of massage on adult postburn scar.** (Nedelec et al, Burns, 2019 Feb;45(1):128-139). The following boxplots show the distribution of skin thickness (in mm) in hypertrophic scars (HSc) before (pre) and after (post) treatment consisting of a 12-week course of scar massage sessions.



- (a) Describe either one similarity or one difference between the scar thickness for pre and post observations in terms of centre, spread or variability, shape, and presence of outliers 4 MARKS
- (b) Is it appropriate here to say that approximately 95% of the pre-treatment observations fall in the interval $(\bar{x} - 2s, \bar{x} + 2s)$? 2 MARKS
- (c) Report the approximate values of the medians of the pre and post treatment scar thicknesses. 2 MARKS
- (d) Assuming that less scar thickness is better, make an (informal) conclusion about the effectiveness of the massage therapy for HSc scars. 2 MARKS

Justify your answers in one or two sentences each.

Q1

- (a)
- Centre: Median of pre is higher than median of post (this is the line in the middle of the box)
 - Spread: Spread is broadly similar across both, although the interquartile range is greater for post (this is measured via the box)
 - Shape: Both pre and post are right skewed
 - Outliers: Pre has a couple of outliers, post does not have any

4 MARKS

- (b) It is probably not the case, due to the skewness mentioned above; however, the skewness is not huge, so the approximation might hold. For the Empirical Rule to apply, you need symmetry and mound shape.

2 MARKS

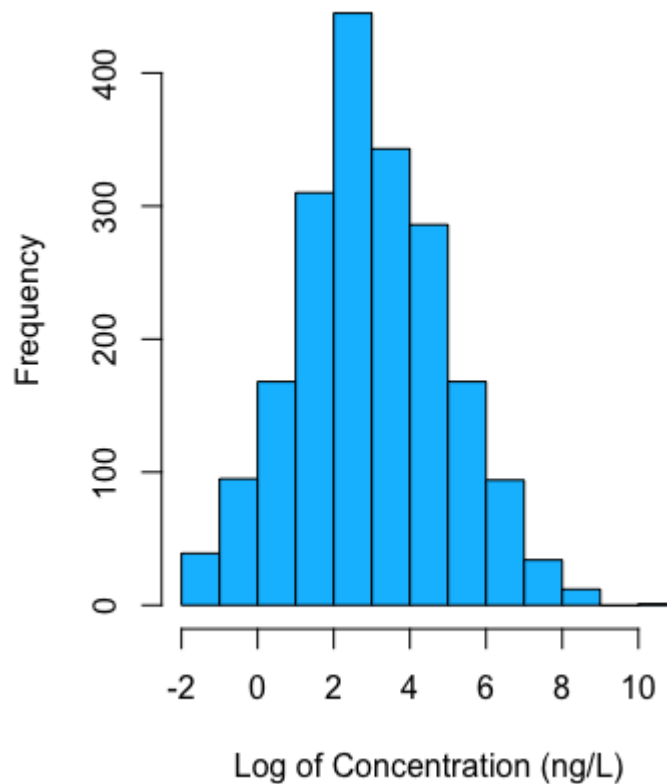
- (c) Pre: 3.0mm, post: 2.5mm (these are the horizontal lines in the middle of the boxes). 2 MARKS

- (d) The treatment appears effective, as the median post treatment is lower than the median pre treatment.

2 MARKS

NB: This comparison might not in fact be completely valid, as if the pre and post measurements are made on the same individuals, you should really look at the difference post - pre. However this information is not given in the question.

2. **Global occurrence of anti-infectives in contaminated surface waters: Impact of income inequality between countries** (Segura et al. Environment International Volume 80, July 2015, Pages 89-97). The following is a histogram of 1995 concentrations of antibiotics from pooled data from different regions in the world (values are log-transformed).



- (a) Compute the sample mean and sample standard deviation for the concentrations of antibiotics if

$$\sum_{i=1}^{1995} x_i = 5956.149, \quad \sum_{i=1}^{1995} x_i^2 = 25517.14.$$

4 MARKS

- (b) Compute the interval $(\bar{x} - 2s, \bar{x} + 2s)$.

2 MARKS

- (c) Approximately what percentage of the observations fall within the interval in (b)? Justify your answer in one sentence.

2 MARKS

- (d) Find the z -score for a log-concentration of 8. Interpret the z -score.

2 MARKS

Q2

(a) We have that

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{5956.149}{1995} = 2.9855$$

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \\ &= \frac{1}{1994} \left(25517.14 - \frac{1}{1995} (5956.149)^2 \right) = 3.8791 \end{aligned}$$

so that $s = 1.9695$.

4 MARKS

(b) We have by direct calculation

$$(\bar{x} - 2s, \bar{x} + 2s) = (-0.9535, 6.9246)$$

2 MARKS

(c) Approximately 95% should lie in this interval, as the Empirical Rule applies to this symmetric distribution. 2 MARKS

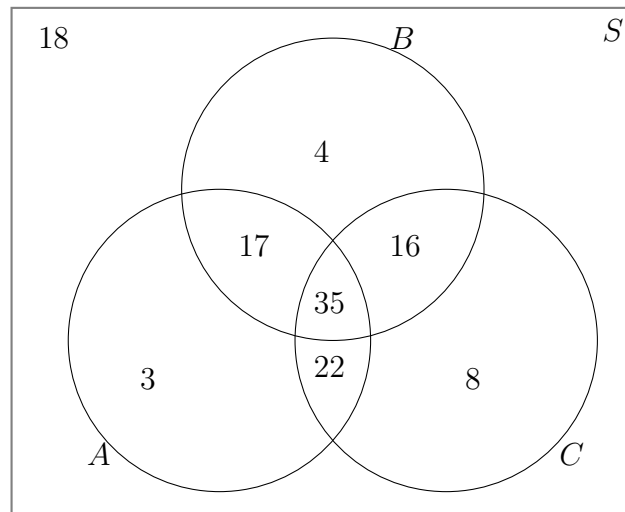
(d) We have that

$$z = \frac{8 - \bar{x}}{s} = \frac{8 - 2.9855}{1.9695} = 2.5460.$$

The z -score tells you how many standard deviations the observation is from the mean; in this case, 8 is 2.5460 standard deviations above the mean. 2 MARKS

3. (a) A total of 123 people were interviewed as part of a transportation study.
- some people live more than 10km from city centre (event A),
 - some people regularly drive to work (event B),
 - some people would use a public transportation if it was available (event C).

The Venn diagram below gives the count breakdown for each partition element:



One person is selected from the 123 people in the study population, with all people equally likely to be selected. Compute the probabilities of the following events:

- the selected person drives to work; 2 MARKS
- the selected person drives to work and would use a public mass transportation system if it were available; 2 MARKS
- the selected person lives more than 10km from the city centre, drives to work, but would not use a mass transportation system. 2 MARKS

Formulate your answers using probability notation, and leave your answers as fractions.

- (b) In a much larger study of a different city, only information concerning events A and B was collected. In this second study, it was determined that

$$P(A \cap B^c) = 0.42 \quad P(A \cap B) = 0.20 \quad P(A^c \cap B) = 0.15.$$

For this city, compute the probability that a person

- lives more than 10km away from the city centre. 2 MARKS
- lives more than 10km away from the city centre, **given** that the person regularly drives to work. 2 MARKS

Q3

- (a) (i) *the selected person drives to work;*

This is $P(B) = (4 + 17 + 35 + 16)/123 = 72/123 = 0.5853$. 2 MARKS

- (ii) *the selected person drives to work and would use a public mass transportation system if it were available;*

This is $P(B \cap C) = (35 + 16)/123 = 51/123 = 0.4146$. 2 MARKS

- (iii) *the selected person lives more than 10km from the city centre, drives to work, but would not use a mass transportation system.*

This is $P(A \cap B \cap C^c) = 17/123 = 0.1382$. 2 MARKS

- (b) (i) *The probability that a person lives more than 10km away from the city centre.*

This is $P(A) = P(A \cap B) + P(A \cap B^c) = 0.20 + 0.42 = 0.62$. 2 MARKS

- (ii) *The probability that a person lives more than 10km away from the city centre, **given** that the person regularly drives to work.*

This is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A \cap B) + P(A^c \cap B)} = \frac{0.20}{0.20 + 0.15} = 0.5714$$

2 MARKS

4. A student responds to a multiple choice question in an exam. The question has four choices and only one of the choices is the correct answer. Suppose that

- the probability that the student knows the correct answer is 0.8;
- if the student **knows the correct answer** they **always** answers the question correctly;
- if the student **does not** know the correct answer, they guess one of the four choices, and is equally likely to pick any one of the four possible answers.

(a) What is the probability that the student does not know the answer, but still chooses the correct answer? 3 MARKS

(b) What is the probability that the student responds correctly? 3 MARKS

(c) What is the probability that the student did not know the answer given that the student answered it correctly? 4 MARKS

Formulate your answers using probability notation.

Q4

Let

- K denote the event that the student knows the answer
- C denote the event that the student answers correctly.

We have from the question

$$P(K) = 0.8 \quad P(C|K) = 1 \quad P(C|K^c) = \frac{1}{4} = 0.25$$

(a) This is

$$P(K^c \cap C) = P(K^c) \times P(C|K^c) = (1 - P(K)) \times P(C|K^c) = (1 - 0.8) \times 0.25 = 0.0500$$

3 MARKS

(b) This is

$$\begin{aligned} P(C) &= P(C \cap K) + P(C \cap K^c) \\ &= P(C|K)P(K) + P(C|K^c)P(K^c) \\ &= (1 \times 0.8) + (0.25 \times (1 - 0.8)) = 0.85 \end{aligned}$$

3 MARKS

Q4

(c) This is

$$\begin{aligned} P(K^c|C) &= \frac{P(K^c \cap C)}{P(C)} \\ &= \frac{0.0500}{0.85} \\ &= 0.0588 \end{aligned}$$

4 MARKS