Do NOT write in the area above this line.

Practice Final Exam

Course: DS 1000B Winter 2026

Sections: 002, 003

Date:

180 minutes

Instructors:

Marieke Mur

Pavel Shuldiner

Allowed aids:

A calculator (non-programmable, non-graphing)

Formula sheet (provided).

Standard Normal Table (provided).

Full Name (print) *(e.g. Tom Marvolo Riddle):*	
Western ID *(e.g. baldemort13):*	
Student Number *(e.g. 251123456):*	

1. Legibly **print** your Western User ID, full name, and student number in the spaces provided above.
2. Do **not** detach the pages of the exam. You may ask for scrap paper if needed.
3. The space at the top of each page is reserved for the scanner. Do not write on or near the barcode.
4. The exam has 24 pages. The last three pages may be used for additional workspace or scrap paper.

Section	**Marks**
Multiple Choice (1 mark each)	10
Exploring Data	6
Study Hours	10
Height and Shoe Size	10
Exercise and Sleep Quality	13
Fair and Biased Coins	11
Wine Bottle Volumes	17
Reverse Curving Grades	12
Student Preferences and Activities	8
Educational Pathways	8
Total	**106**

Multiple Choice (1 mark each) **(10 marks)**

Each of the following multiple-choice questions is worth 1 mark. Circle the correct answer for each question. No justification is required and no partial marks will be awarded.

MC1. (1 mark) A regression model predicting final exam score from hours studied has $R^2 = 0.64$. The best interpretation is:

A. 64% of the variability in exam scores is explained by the linear relationship with hours studied.

B. The correlation between hours studied and exam score is 0.64.

C. 64% of students passed the exam.

D. For each additional hour studied, the exam score increases by 64%.

E. 64% of students studied enough hours to do well.

MC2. (1 mark) A data set has a mean of 50 and a median of 48. If the largest value in the data set is replaced by a value ten times as large, which of the following is true?

A. Both the mean and median increase.

B. The median increases but the mean stays the same.

C. Both the mean and median stay the same.

D. Both the mean and median decrease.

E. The mean increases but the median stays the same.

MC3. (1 mark) A histogram of household incomes in a city is strongly right-skewed. Which of the following is most likely true?

A. The mean is less than the median.

B. The mean equals the median.

C. The mean is greater than the median.

D. The standard deviation equals zero.

E. The distribution is symmetric.

MC4. (1 mark) A random variable X takes values 0, 1, 2, and 3 with probabilities 0.10, 0.25, 0.40, and 0.25 respectively. What is $P(X \geq 2)$?

A. 0.25 **B.** 0.40 **C.** 0.65 **D.** 0.75 **E.** 0.90

MC5. (1 mark) A researcher measures the heights of 50 randomly selected students and calculates a mean of 170 cm. The value 170 cm is best described as:

A. A population

B. A statistic

C. A variable

D. A sample space

E. A parameter

MC6. (1 mark) In a regression analysis, a data point has a residual of -5. This means:

A. The observed value is 5 units above the predicted value.

B. The slope of the regression line is -5.

C. The predicted value is 5 units above the observed value.

D. The correlation coefficient is -5.

E. The regression model has an error of 5%.

MC7. (1 mark) As the number of coin flips increases, the proportion of heads observed tends to get closer to 0.5. This phenomenon is explained by:

A. The Central Limit Theorem

B. Bayes' theorem

C. The empirical rule

D. The Law of Large Numbers

E. The addition rule

MC8. (1 mark) Which of the following best describes what the standard deviation measures?

A. The middle value of the data set.

B. The difference between the largest and smallest values.

C. The most frequently occurring value.

D. The average of all data values.

E. The typical distance of data values from the mean.

MC9. (1 mark) A data set has $Q_1 = 20$ and $Q_3 = 40$. Using the $1.5 \times$ IQR rule, a value is considered an outlier if it is:

A. Below 10 or above 50.

B. Below 0 or above 60.

C. Below 5 or above 55.

D. Below -10 or above 70.

E. Below 15 or above 45.

MC10. (1 mark) You interview 10 randomly selected workers about their commute distance and compute the sample mean. If you repeat this survey many times, recording each sample mean, what does a histogram of these sample means represent?

A. The bias present in the sampling method

B. The true population mean commute distance

C. A simple random sample

D. The sampling distribution of the sample mean

E. The confidence interval for the population mean

Exploring Data **(6 marks)**

Q1. A student council conducts a survey of 200 students about their study habits and preferences. Answer the following questions based on their findings.

(a) (2 marks) The survey found that 35% of students prefer studying at the library, 30% at home, 25% at a coffee shop, and 20% in study rooms. Can a pie chart be drawn to display these results? Explain.

(b) (2 marks) The survey also recorded each student's weekly study hours. Is this variable categorical or quantitative? Name one appropriate visualization for displaying the distribution of this variable.

(c) (2 marks) What does correlation measure?

Study Hours **(10 marks)**

Q2. The following data represent the number of hours studied per week by 12 randomly selected university students:

8, 13, 15, 17, 18, 19, 21, 22, 24, 26, 28, 30

(a) (2 marks) Draw a stemplot for this data.

(b) (1 mark) Identify the shape of the distribution.

(c) (5 marks) Calculate the five-number summary for this data.

(d) (2 marks) Based on the shape you identified in part (b), which measures of center and spread would be most appropriate to summarize this data? Justify your answer.

Height and Shoe Size **(10 marks)**

Q3. A researcher studies the relationship between height (in inches) and shoe size (US men's sizing) among college students. After collecting data, they obtain the following least squares regression line:

$$\widehat{\text{Height}} = 50 + 2.5 \times (\text{Shoe Size})$$

(a) (2 marks) Suppose the residual for a particular observation is 1 inch and the observed height is 74 inches. What is the shoe size of this student? Show your work.

(b) (2 marks) If the mean height of the sampled students is 66.67 inches, what is the mean shoe size? Show your work.

(c) (3 marks) Using the same data, the researchers found that the least squares regression line for predicting shoe size from height is

$$\widehat{\text{Shoe Size}} = -4.0 + 0.16 \times (\text{Height})$$

Calculate the correlation coefficient (r) between height and shoe size. Show your work.

(d) (3 marks) If we measure height in centimetres instead of inches (where 1 inch = 2.54 cm), would the correlation coefficient change? Would the slope of the regression line change? Explain.

Exercise and Sleep Quality **(13 marks)**

Q4. The table below shows data from a survey about exercise habits and sleep quality among 500 adults.

	Poor	**Fair**	**Good**	**Total**
Exercises regularly	30	x	160	
Does not exercise regularly	y	100	z	
Total				500

where x, y, and z represent unknown cell counts:

- x = number of adults who exercise regularly and report "Fair" sleep
- y = number of adults who do not exercise regularly and report "Poor" sleep
- z = number of adults who do not exercise regularly and report "Good" sleep

(a) (2 marks) Suppose it is known that 200 of the adults surveyed exercise regularly. Find x.

(b) (4 marks) The relative risk of having good sleep quality for those who exercise regularly versus those who do not is 2. Determine z, the number of adults who do not exercise regularly and report "Good" sleep quality. Show your work.

(c) (3 marks) Calculate the odds ratio of having poor sleep quality for those who do not exercise regularly compared to those who exercise regularly. Show your work.

(d) (2 marks) Interpret the odds ratio from part (c) in the context of this problem.

(e) (2 marks) Is this study an experiment or an observational study? Can we conclude that exercise causes better sleep quality? Explain.

Fair and Biased Coins **(11 marks)**

Q5. A bag contains two coins: one is a fair coin, and the other is a biased coin with heads on both sides. One coin is selected at random from the bag, flipped once, and the result is observed.

(a) (2 marks) Draw a tree diagram to represent all possible outcomes of this experiment. Label all branches with their probabilities.

(b) (1 mark) What is the probability of selecting the fair coin?

(c) (1 mark) Given that the biased coin was selected, what is the probability of observing heads?

(d) (2 marks) What is the probability of observing heads?

(e) (3 marks) Given that heads was observed, what is the probability that the biased coin was selected? Show your work.

(f) (2 marks) What is the probability that either tails is observed or the biased coin is selected?

Wine Bottle Volumes **(17 marks)**

Q6. The volume of wine in a set of bottles is known to follow a normal distribution with standard deviation $\sigma = 10$ mL.

(a) (5 marks) A random sample of 25 bottles is taken, and the sample mean volume is $\overline{x} = 750$ mL. Construct a 95% confidence interval for the true mean volume μ. Show your work.

(b) (5 marks) For the same data, construct a 99% confidence interval. Compare its width[1] to the 95% interval and explain the difference.

[1]The width of a confidence interval is the difference between its upper and lower bounds.

(c) (7 marks) How many bottles must be sampled to obtain a 95% confidence interval for μ with total width equal to 1 mL? Show your work.

Reverse Curving Grades **(12 marks)**

Q7. A statistics professor's midterm was too easy. The raw scores for 200 students are normally distributed with mean $\mu = 82$ and standard deviation $\sigma = 8$. The department requires that the class average align with a target of 70. The professor decides to apply a linear "reverse curve" of the form

$$Y = aX + b$$

where X is the unadjusted score and Y is the adjusted score.

(a) (2 marks) What proportion of students had an unadjusted score above 90? Show your work.

(b) (2 marks) Find the 50th percentile of the unadjusted score distribution.

(c) (3 marks) If the standard deviation of the curved scores is 12, find the value of a. Show your work.

(d) (2 marks) Using the requirement that the curved class average is 70, find the value of b. Show your work.

(e) (3 marks) What raw score does a student need to earn at least an A (80 or above) on the curved exam? Show your work.

Student Preferences and Activities **(8 marks)**

Q8. A survey is conducted among students about their preferences and activities. Consider the events A, B and C where,

A : Owns a bicycle

B : Prefers tea over coffee

C : Member of chess club

- A and B are independent
- A and C are mutually exclusive
- $P(A) = 0.30,\ \ P(B) = 0.40,\ \ P(C) = 0.20$
- $P(B \text{ and } C) = 0.10$

(a) (2 marks) Calculate the probability that a student owns a bicycle or prefers tea over coffee. Show your work.

(b) (1 mark) Determine the probability that a student owns a bicycle and is a member of the campus chess club.

Recall that a survey is conducted among students about their preferences and activities. Consider the events A, B and C where,

A : Owns a bicycle B : Prefers tea over coffee C : Member of chess club	• A and B are independent • A and C are mutually exclusive • $P(A) = 0.30,\ P(B) = 0.40,\ P(C) = 0.20$ • $P(B \text{ and } C) = 0.10$

(c) (3 marks) Calculate the probability that a student prefers tea given that they are in the chess club. Show your work.

(d) (3 marks) Calculate the probability that a student owns a bicycle or prefers tea over coffee, or is a member of the campus chess club; that is, find $P(A \text{ or } B \text{ or } C)$.

Hint : Use the general addition rule for three events:

$$P(A \text{ or } B \text{ or } C) = P(A)+P(B)+P(C)-P(A \text{ and } B)-P(B \text{ and } C)-P(A \text{ and } C)+P(A \text{ and } B \text{ and } C)$$

Educational Pathways **(8 marks)**

Q9. Assume the educational path is sequential: finishing high school is required before starting undergraduate studies; an undergraduate degree is required before pursuing a master's degree; a master's degree is required before a Ph.D. Each stage depends on completing the previous one.

Suppose 90% of Canadian students finish high school. Of those who finish high school, 50% complete an undergraduate degree. Of those who complete undergraduate, 40% earn a master's degree. Of those who earn a master's, 10% earn a Ph.D.

(a) (2 marks) Draw a tree diagram to represent the educational pathways and label all branches with their probabilities.

(b) (2 marks) What percentage of students earn a Ph.D.?

(c) (2 marks) Find the probability that a student finished high school but did not earn a Ph.D.

(d) (2 marks) Given that a student completed a Ph.D., find the probability that they earned an undergraduate degree.

Additional Workspace

Use this space for additional work if needed. Clearly indicate which question you are working on and reference the page number.

Additional Workspace

Use this space for additional work if needed. Clearly indicate which question you are working on and reference the page number.

Additional Workspace

Use this space for additional work if needed. Clearly indicate which question you are working on and reference the page number.