

**MAT 1371  
Midterm**

**Professor G. Lamothe**

**Duration: 80 minutes**

**Student Number: \_\_\_\_\_**

**Last Name: \_\_\_\_\_ First Name: \_\_\_\_\_**

- **This is a closed book examination. However one sheet is permitted.**
- **Only non-programmable and non-graphic calculators are permitted.**
- **The exam will be marked on a total of 30 points.**

1. [10 points] For the following multiple choice questions, please put your answers in the following table.

Question	a	b	c	d	e
Answer	A	C	B	A	D

- (a) In 2004, Canada produced carbon dioxide (CO<sub>2</sub>) emissions. It was 2.80 in log of millions of metric tons. Re-express this value in millions of metric tons.

A) 630.96    B) 6.96    C) 2.80    D) 398,107.17

**Solution:**  $\log(y) = 2.80 \Rightarrow y = 10^{2.80} = 630.96$

- (b) Environmental researchers have collected data on rain acidity for years. Suppose that a Normal model describes the acidity (pH) of rainwater, and that water tested after last weeks storm had a z-score of -1.8. This means that the acidity of that rain . . .

A) had a pH of -1.8.  
B) had a pH 1.8 times that of average rainwater.  
C) had a pH 1.8 standard deviations smaller than that of average rainwater.  
D) had a pH 1.8 smaller than average rainfall.

- (c) You should use a histogram to display categorical data:

A) True    B) False

- (d) The five-number summary of credit hours for 24 students in an introductory statistics class is:

Min	$Q_1$	Median	$Q_3$	Max
13	15	16.5	18	22

From this we know that

- A) there are no outliers in the data.
- B) there is at least one low outlier in the data.
- C) there is at least one high outlier in the data.
- D) None of the above.

**Solution:**  $Q_1 - 1.5\text{IQR} = 15 - 1.5(3) = 10.5 < 13$  So there are no outliers to the left of the inner fence.

$Q_3 + 1.5\text{IQR} = 18 + 1.5(3) = 22.5 > 22$  So there are no outliers to the right of the outer fence.

Therefore there are no outliers.

- (e) An extreme outlier is a point more than \_\_\_\_\_ IQR from either end of the box in a boxplot.

- A) 0.75    B) 1.5    C) 2.0    D) 3.0

2. [5 points] Consider the following sample of  $n = 45$  observations.

18.9 19.5 19.6 20.0 20.6 20.8 21.6 21.6 21.6 21.6  
 21.7 21.7 22.0 22.0 22.0 22.0 22.0 22.0 22.0 22.0  
 22.0 22.0 22.2 22.2 22.2 22.2 22.2 22.2 22.2 22.2  
 22.2 22.2 22.4 22.4 22.4 22.6 22.6 22.8 22.8 22.8  
 23.1 23.2 23.4 23.6 23.8

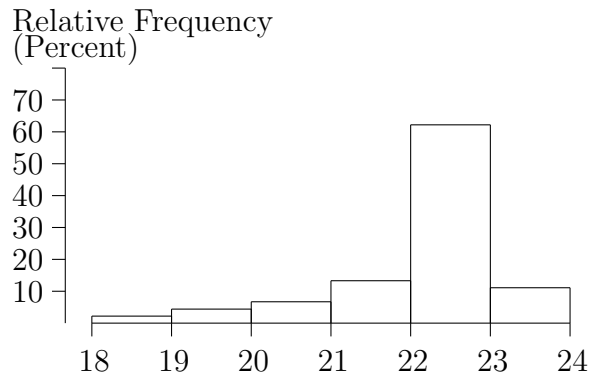
- a) Construct a relative frequency histogram with 6 or 7 bins.
- b) Describe the shape of the distribution. Are there any unusual features?
- c) Using the following sums compute the mean and the standard deviation.  

$$\sum y = 989.1 \quad \text{and} \quad \sum y^2 = 21,784.17$$
- d) Compute the median.
- e) Do you believe that the median or the mean is more appropriate to describe the center of this distribution? Justify.

**Solution:** Below is the distribution for this variable.

$y$	Frequency	Relative Frequency
$18 \leq y < 19$	1	2.2%
$19 \leq y < 20$	2	4.4%
$20 \leq y < 21$	3	6.7%
$21 \leq y < 22$	6	13.3%
$22 \leq y < 23$	28	62.2 %
$23 \leq y < 24$	5	11.1%

Below is the relative frequency histogram.



b) The distribution is unimodal and strongly skewed to the left. Most of the values appear to be clustered between 22 and 23, which represents about 62% of the values. There are no unusual features like gaps or extreme values.

c) The mean is  $\bar{y} = \sum y/n = 989.1/45 = 21.98$  and the standard deviation is

$$s = \sqrt{\frac{\sum y^2 - (\sum y)^2/n}{n - 1}} = 0.997.$$

d) The position of the median is  $50\%(n + 1) = 23$ . Hence the median is 22.2.

e) Since we believe that the distribution is strongly skewed then we should use the median to describe the center of the distribution.

3. [5 points] In Canada, systolic blood pressure readings have a mean of 121 and a standard deviation of 16. A reading above 140 is considered to be high blood pressure.

a) What is the  $z$ -score for a blood pressure reading of 140?

b) Assuming that the Normal model is appropriate to describe systolic blood pressure in Canada, then what percentage of Canadians would be considered to have high blood pressure?

c) Assuming that the Normal model is appropriate to describe systolic blood pressure in Canada and that Bob's reading is 160, what percentage of Canadians have a blood pressure reading smaller than Bob?

**Solution:**

a) The  $z$ -score for 140 is  $z = (140 - 121)/16 = 1.1875 \approx 1.19$ .

b) Equivalently we can find the percentage of  $z$ -scores that are larger than 1.19. From the Table  $Z$ , we get that 88.3% of the scores are smaller than 1.19, thus 11.7% of the scores are larger than 1.19.

Therefore, 11.7% of Canadians would be considered to have high blood pressure.

c) The  $z$ -score for 160 is  $z = (160 - 121)/16 = 2.4375 \approx 2.44$ .

Equivalently we can find the percentage of  $z$ -scores that are smaller than 2.44. From the Table  $Z$ , we get that 99.27% of the  $z$ -scores are smaller than 2.44.

Therefore there are 99.27% of Canadians that have a lower blood pressure reading than Bob.

4. [5 points] On Monday a class of students took a big test, and the highest score was 92. The next day a student who had been absent made up the test, scoring 100. Indicate whether adding that student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:

- a. mean                      increase
- b. median                    stay about the same
- c. range                     increase
- d. IQR                        stay about the same
- e. standard deviation      increase

5. [5 points] In a survey of public access agreements between local authorities and landowners, in and around London, two variables were recorded: whether or not each authority had access agreements; and what type of local authority it was (that is it was a district, a metropolitan district or a London borough). There were 287 authorities surveyed. Here is the corresponding contingency table.

	Agreements present	Agreements absent	Total
Districts	68	169	237
Metropolitan districts	20	7	27
London boroughs	10	13	23
Total	98	189	287

- a) Find each percentage.
- What percentage of the authorities have access agreements?  
*Answer:*  $98/287 \times 100\% = 34.15\%$
  - What percentage of the London borough authorities have access agreements?  
*Answer:*  $10/23 \times 100\% = 48.48\%$
  - What percentage of the authorities without access agreements are classified as a London borough authority?  
*Answer:*  $13/189 \times 100\% = 6.88\%$
- b) Are the distributions of the presence or absence of access agreements and of authority types independent? Explain.

**Solution:** The percentage of the Metropolitan districts authorities that have access agreements is  $20/27 \times 100\% = 74.07\%$ . However the percentage of authorities with access agreements among the London borough authorities is only 48.48%. Thus, the distributions of the presence or absence of access agreements conditional on the type of authority are **not** the same. Hence the variables are **not** independent.