Midterm Exam (Solutions)

October 19, 2021 2:57 PM

Name: Answer Key **Set:**

BRITISH COLUMBIA INSTITUTE OF TECHNOLOGY – Midterm Examination

Program:	Computer Information Technology
Course Name:	Statistics for CIT
Course Number:	MATH 1350
Date:	October 21, 2021
Time Allotted:	120 minutes
Exam Pages:	12 (including this page)
Total Marks:	~~55~~ 54 (30% of this course)

Special Instructions:

1.) All answers are to be written in this examination booklet. **Only work in this booklet will be considered when calculating marks.**

2.) Students are permitted the use of an electronic calculator and RStudio. No other equipment or software is allowed.

3.) If you are asked to **include supporting details**, you must write out the calculation in detail or write out an R command that achieves the same result. **Otherwise**, you may write just the final answer.

4.) Textbooks and notes are NOT allowed!

5.) A formula sheet is provided.

6.) Answer all probability problems rounded to 4 decimal places.

7.) Turn off your phone and put it away.

8.) Before we begin, download the file "MATH_1350_Midterm_Data.xlsx" from Learning Hub (in Content→Midterm and Final).

Question 1 [16 marks total]

Suppose data has been collected for a sample of $n = 50$ students in MATH 1350 at BCIT. Half of these students are in set D and half are in set E. (This is not real data!)

The raw data are available in the file "MATH_1350_Midterm_Data.xlsx" which you must download from Learning Hub.

a) [1 mark] Which of the following could *not* be the population of interest? (Circle the correct answer.)

(1) (since it is a subset)

i. All students in set D
ii. All students in MATH 1350
iii. All BCIT students
iv. All post-secondary students in Canada

b) [1 mark] Which of the following is *not* a statistic pertaining to this data? (Circle the correct answer.)

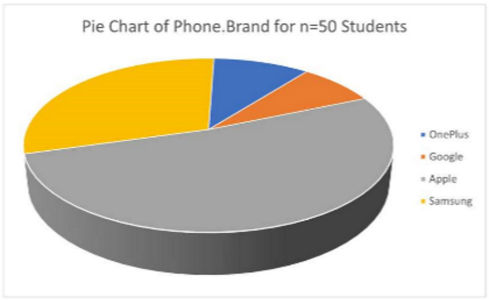
(1) not from sample data

i. n = sample size
ii. IQR = sample inter-quartile range of Age
iii. s^2 = sample variance of Siblings
iv. μ = population mean of Height

c) [1 mark] What is the mode of the variable *Wears.Glasses* for this sample?

(1) mode = "Yes"

d) [1 mark] What is one problem with presenting *Phone.Brand* using a chart like the following? Give a one-sentence answer.



Two answers:

(1) • It's hard to accurately compare slices.

or (1) • The 3-D perspective artificially makes some segments look larger.

(**Question 1** continued…)

e) [2 marks] Find the mean and standard deviation of *Siblings* for this sample. Use the correct symbol for each and round your answers to two decimal places.

$\bar{x} = 1.28$ (1)

$s = 1.13$ (1)

f) [2 marks] Calculate P_{86} (the 86th percentile) for *Siblings*. **Include supporting details.** (You may use R's algorithm or the simplified algorithm presented in class.)

$L = \left(\frac{k}{100}\right)n = \frac{86}{100} \times 50 = 43$ (1)

The 43rd value is 2 and the 44th value is 3,

so $P_{86} = \frac{2+3}{2} = \boxed{2.5}$ (1)

Using R: quantile(df$Siblings, 0.86) = 2.14 (1) (1)

g) [2 marks] Would a student for whom *Siblings* = 0 be considered statistically *unusual* for this sample? Circle "Yes" or "No". **Include supporting details.**

(1) Yes (No) (Circle one)

$z = \frac{x-\mu}{\sigma} = \frac{0-1.28}{1.13} = -1.14.$

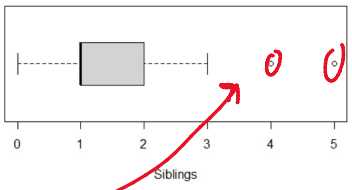
Since z is between −2 and +2, it is not unusual. (1)

(**Question 1** continued…)

h) [1 mark] List the value of all outliers for *Siblings*. [*Hint: You can read the answer off the appropriate boxplot.*]

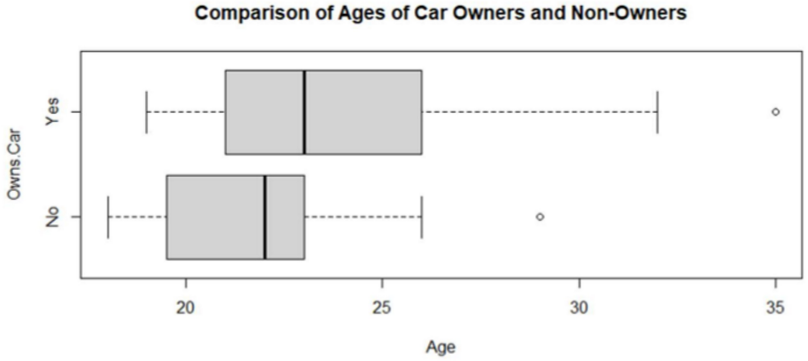
Siblings = 4, 5

(1)



from boxplot

i) [2 marks] Shown below are boxplots for *Age* of students, grouped by the variable *Owns.Car* into car owners and non-owners.



Does the graph imply that students who own cars are typically older than students who do not own cars?

Yes No (Circle one)

Justify your answer by providing the numerical values of one measure of center (one number for each group). ~~**Include supporting details.**~~

Median (car owners) = 23

Median (non-owners) = 22

or

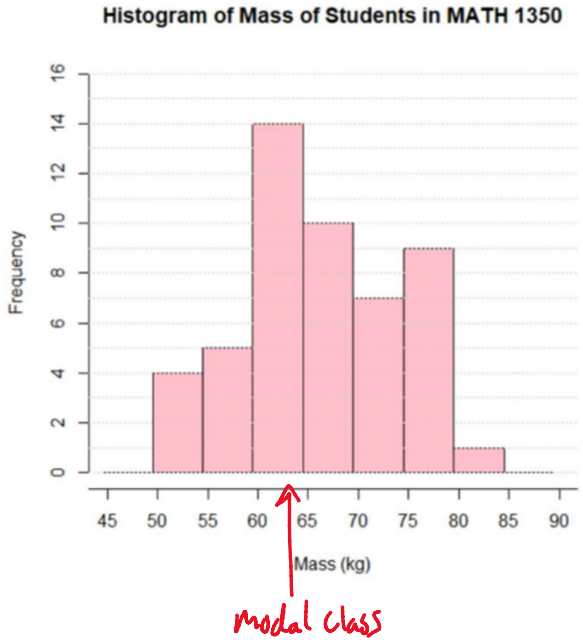
Mean (Car-Owners) = 24.1

Mean (non-owners) = 21.9

(**Question 1** continued…)

j) Assume that the mass of each student was also measured and recorded as the variable *Mass* (in kilograms, rounded to the nearest kilogram). Note that you have not been given the raw data for this variable.

To plot a histogram of *Mass*, frequencies were tallied for the following classes (in kg): 45-49, 50-54, 55-59, 60-64, and so on. The resulting histogram is shown below.



i. [1 mark] What are the *class boundaries (lower and upper)* for the modal class?

lower boundary = 59.5 (0.5)
upper boundary = 64.5 (0.5)

ii. [1 mark] What percentage of students have *Weight* < 74.5 kg?

9+1 = 10 are ≥ 75, so 50 − 10 = 40 are < 74.5

Answer: 40/50 = 80%

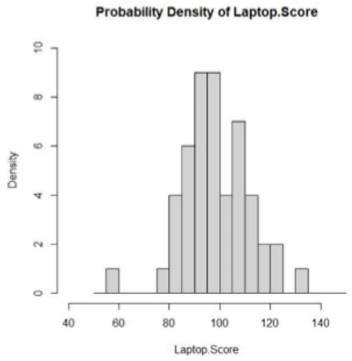
iii. [1 mark] In which class must the 30th percentile of *Weight* be located? State its lower and upper class limits.

30% of 50 = 15.

The 15th value is in the class 60-64 kg

Question 2 [10 marks total]

Suppose that each student's laptop was benchmark tested using a standard computational task. The variable *Laptop.Score* contains the results (in mips = millions of instructions per second). Use the provided data set to answer the questions:



a) [2 marks] What is the *z*-score for a student whose benchmark result was *Laptop.Score* = 93.2 mips? **Include supporting details.**

$\bar{X} = 98.532$ mips (1)

$s = 13.475$ mips

$z = \frac{x - \bar{x}}{s} = \frac{93.2 - 98.532}{13.475} = -0.396$ (1)

or (1)

b) [1 mark] If the *z*-score for a student's laptop is $z = 2.5$, what was its *Laptop.Score* value? **Include supporting details.**

$x = \bar{x} + z \cdot s = 98.532 + 2.5 \times 13.475$

$= 132.2$ mips (1)

c) [2 marks] Use Chebyshev's Theorem to find an interval in which *Laptop.Score* must lie for at least 75% of individuals in the sample. Round to ~~two~~ one decimal places. **Include supporting details.**

Chebychev says $P(-k \le z \le k) \ge 1 - \frac{1}{k^2}$.

So let $k = 2$ to get $P(-2 \le z \le 2) \ge 1 - \frac{1}{4} = 75\%$. (1)

$z = -2 \Rightarrow x = 71.6$

$z = +2 \Rightarrow x = 125.5$

at least 75% of individual values must lie in the interval [71.6 mips, 125.5 mips] (1)

(**Question 2** continued…)

d) [1 mark] The variable *Laptop.Score* is approximately normally distributed. According to the Empirical Rule, what percentage of individuals should lie within one standard deviations of the mean?

68% (1)

e) [2 marks] What is the *actual* percentage of individuals for which *Laptop.Score* is within one standard deviations of the mean? **Include supporting details.**

```
xbar ← mean(df$Laptop.Score)
S ← sd(df$Laptop.Score)
Sum((xbar − s < df$Laptop.Score) &
    (Xbar + s > df$Laptop.Score))/n
```

= 0.72 = 72%

f) Let Z denote the z-score derived from *Laptop.Score* for individuals in this sample.

i. [1 mark] What are the mean and standard deviation of Z?

$\bar{Z} =$ 0

$s_Z =$ 1

ii. [1 mark] If you randomly select one individual, what is $P(0 \leq Z \leq 2)$? **Include supporting details.**

$P(0 \leq Z \leq 2) =$ 0.42

Question 3 [6 marks total] Consider the variable *Gender* for students in this sample. Suppose you randomly select two students *with* replacement. Define the events:

F_1 = *Gender* is "Female" for first student
F_2 = *Gender* is "Female" for second student

a) [1 mark] Calculate $P(F_1)$ using the data provided.

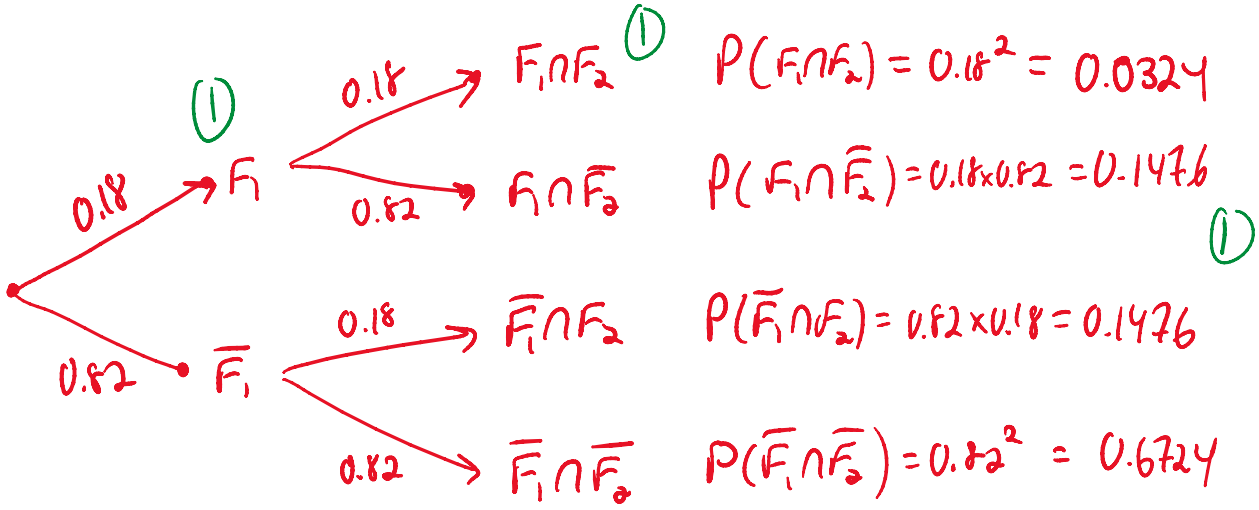
$P(F_1) = 0.18$

b) [2 marks] Are events F_1 and F_2 independent? Explain.

Yes, since the selection is done with replacement.

c) [3 marks] Draw a tree diagram appropriate for this random experiment. Ensure that:

- each branch is labelled with a numerical probability
- the tree gives probabilities for all four final outcomes



Question 4 [8 marks total] A two-way table generated in R for the variables *Owns.Car* and *Phone.Brand* is shown below.

Owns.Car	Phone.Brand: Apple	Google	OnePlus	Samsung	total
No	9	2	2	10	23
Yes	17	2	3	5	27
total	26	4	5	15	50

Suppose you randomly select one student from the sample. Define the events:

C = student owns a car
A = student's phone brand is Apple

For each of the following, **include supporting details.**

a) [1 mark] Calculate $P(A)$ = $\frac{|A|}{|S|} = \frac{9+17}{50} = 0.52$ (1)

b) [1 mark] Calculate $P(C)$ = $\frac{|C|}{|S|} = \frac{27}{50} = 0.54$ (1)

~~c) [2 marks] Calculate $P(A \mid C)$ =~~

d) [~~2~~ 1 mark] Calculate $P(A \cap C)$ = $\frac{|A \cap C|}{|S|} = \frac{17}{50} = 0.34$ (1)

e) [2 mark] Are A and C *independent*? Explain.

(1) No. Since $P(A \cap C) = 0.34$
but $P(A) \cdot P(C) = 0.52 \times 0.54 = 0.2808$.
(1)

or $P(A \mid C) = \frac{17}{27} = 0.6296 \neq 0.52 = P(A)$
(1)

Question 5 [7 marks total] Suppose you flip 4 fair coins, meaning heads (H) and tails (T) are equally likely for each flip. Then the sample space can be represented as the following set of *equally likely* outcomes:

$$SS = \{\, HHHH,\ HTHH,\ THHH,\ TTHH,\ HHHT,\ HTHT,\ THHT,\ TTHT,\ HHTH,\ HTTH,\ THTH,\ TTTH,\ HHTT,\ HTTT,\ THTT,\ TTTT \,\}$$

Define the events:

$A = \{\, HHHH \,\}$ (or "all heads")
$B = \{\, TTTT \,\}$ (or "all tails")

For each of the following, **include supporting details (i.e., show the calculation)**.

a) [1 mark] Calculate $P(A \cap B) =$ $\frac{|A \cap B|}{|SS|} = \frac{0}{16} = 0$ (1)

b) [1 mark] Calculate $P(A \cup B) =$ $P(A)+P(B)-P(A \cap B)$
$= \frac{1}{16} + \frac{1}{16} - 0 = 0.1250$ (1)

c) [1 mark] Calculate $P(\bar{B}) =$ $1 - P(B) = 1 - \frac{1}{16} = \frac{15}{16} = 0.9375$ (1)

d) [1 mark] Calculate $P(\bar{B} \mid \bar{A}) =$ $\frac{|\bar{B} \cap \bar{A}|}{|\bar{A}|} = \frac{14}{15} = 0.9333$ (1)

e) [1 mark] Are $\bar{A}$ and $\bar{B}$ independent? Explain.

No, since $P(\bar{B}) \neq P(\bar{B} \mid \bar{A})$ (1)

f) [1 mark] Calculate $P(\bar{A} \cap \bar{B}) =$ $\frac{|\bar{A} \cap \bar{B}|}{|SS|} = \frac{14}{16} = 0.8750$ (1)

g) [1 mark] Calculate $P(\bar{A} \cup \bar{B})$ $= P(\bar{A})+P(\bar{B})-P(\bar{A} \cap \bar{B})$
$= \frac{15}{16} + \frac{15}{16} - \frac{14}{16} = 1.000$ (1)

Question 6: [8 marks total] Data for the variable *Eye* is summarized in the frequency table below:

Black	Blue	Brown	Green
10	5	32	3

$n = 50$

a) [2 marks] In how many ways can you select an ordered *sequence* of 3 students from this sample (without replacement)? **Include supporting details.**

$n = 50$, $r = 3$

$nPr = 50P_3 = 117600$ (1) (1) (Calculator)

b) [2 marks] If you randomly select an ordered sequence of 3 students (without replacement), what is the probability that all three students have green eyes? **Include supporting details.**

$$\frac{3P_3}{50P_3} = \frac{6}{117600} = 5.102 \times 10^{-5} = 0.00005102 = \boxed{0.0001}$$

or $\frac{3C_3}{50C_3} = \frac{1}{19600}$ = Same answer

c) [2 marks] If you randomly select an ordered sequence of 3 students (without replacement), what is the probability of obtaining *at least* one student with black eyes? **Include supporting details.**

$$P(\text{at least one Black}) = 1 - P(\text{no Black}) = 1 - \frac{40C_3}{50C_3} = \boxed{0.4959}$$

d) [2 marks] If you randomly select an ordered sequence of 3 students (without replacement), what is the probability of obtaining 2 students with brown eyes and 1 student with blue eyes? **Include supporting details.**

$$P(2 \text{ Brown } 1 \text{ Blue}) = \frac{32}{50} \times \frac{31}{49} \times \frac{5}{48} + \frac{32}{50} \times \frac{5}{49} \times \frac{31}{48} + \frac{5}{50} \times \frac{32}{49} \times \frac{31}{48}$$

$$= \boxed{0.1265}$$

(**Question 6** continued…)

e) [3 marks] If you randomly select 3 students, what is the probability that at least two of them have the *same* eye colour? Show enough detail to make it clear what method you used. (You may use simulation in R.)

Final answer:

$p = 0.837$