**Name:** ______________________________

**ID:** ______________________________ **Set:** _______

**British Columbia Institute of Technology**

# MATH 3042 – Midterm Exam

| | |
|---|---|
| **Program**: | Computer Systems Technology |
| **Course Name**: | Applied Probability and Statistics for CST |
| **Course Number**: | Math 3042 |
| **Date**: | October 22, 2025 |
| **Time Allotted**: | 90 min |
| **Exam Pages**: | 12 (including this page) |
| **Total Marks**: | 46 (25% weight for the course) |

## Instructions

1) Do not open the exam or write anything on these pages before you are told to begin.

2) You may use a scientific calculator with statistics functions. No other devices are allowed.

3) If your answer is a probability, round it to four digits after the decimal point. Otherwise, round to three significant digits.

4) A formula sheet is provided separately. No other notes or written materials are allowed.

5) No communication of any sort is allowed with other students or any other person besides your instructor or other exam invigilator.

6) All answers are to be written clearly in this examination booklet.

| **Q1** | **Q2** | **Q3** | **Q4** | **Q5** | **Q6** | **Q7** | **Q8** | **Q9** | **Q10** |
|---|---|---|---|---|---|---|---|---|---|
| **/4** | **/3** | **/4** | **/5** | **/4** | **/8** | **/4** | **/5** | **/4** | **/5** |

**Question 1** [4 marks]

In this question we are considering the variable

$$X = \text{the inflation-adjusted price of gold (\$ per oz)}$$

The value of $X$ on the first day of each month since Oct 2022 is shown below in an R-style stem plot. The sample size is $n = 37$.

In this plot, the decimal point is 2 digits to the right of the |.

```
17 | 7
18 |
19 | 2468
20 | 347888
21 | 022357
22 |
23 | 0599
24 |
25 | 17
26 | 8
27 | 01
28 | 149
29 |
30 |
31 | 5
32 | 78
33 | 016
34 |
35 |
36 |
37 | 4
38 |
39 |
40 | 9
```

a. [2] Determine the median and mode for $X$ based on this sample.

$n = 37$
$\text{Q}_2 = \text{19th value} = 2350$ \$ per oz
$\text{Mode} = 2080$ \$ per oz

b. [1] What is the problem with using the mean as a measure of center for this sample?

The two largest values, 3740 and 4090 (from the two most recent months!), are considerably larger than the others. This will pull up the mean, making it a less accurate measure of center.

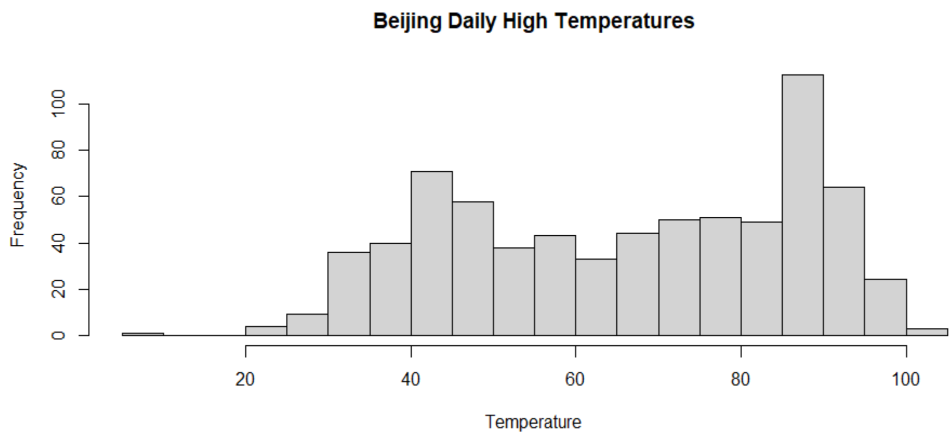
[Pearson's skewness coefficient is $Sk = 0.89$.
Including all 37 values gives $\bar{X} = 2524$ \$ per oz.
Dropping the 2 largest gives $\bar{X} = 2444$ \$ per oz.]

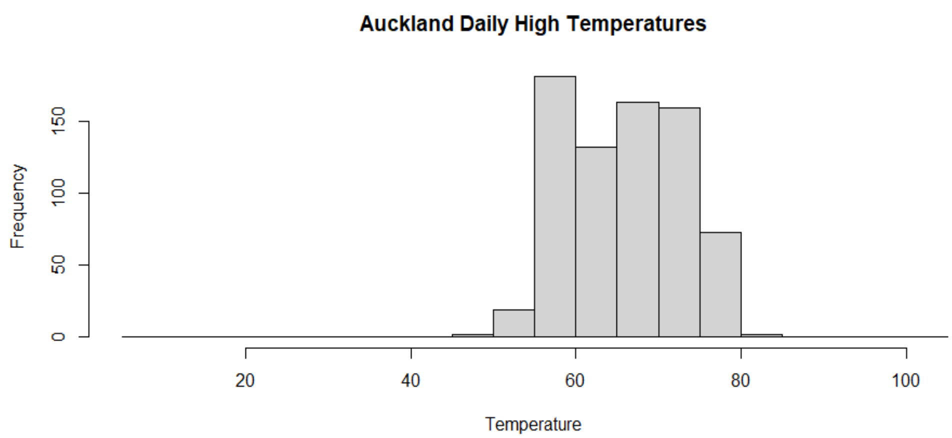
c. [1] Determine the 90th percentile of $X$ based on this sample.

$0.90 \times n = 33.3$
$P_{90} \approx X_{34} = 3310$ \$ per oz

**Question 2** [3 marks]

The daily high temperature readings in the cities of Beijing and Auckland were recorded for 731 days. These data are summarized in the histograms below.




a. [2] Which city likely has the larger standard deviation? Justify your answer with reference to features of the histograms.

Beijing likely has the larger standard deviation.
This is apparent from the much larger width of the distribution for Beijing compared to Auckland.

b. [1] Based on your answer to part a., which city has the more *consistent* daily high temperature?

Consistency is the opposite of variability, so Auckland is the city with more consistent daily high temperatures.

**Question 3** [4 marks]

During a stress test of a web application, memory usage (in MB) was recorded every minute for 11 minutes. The recorded values are:

3200, 3000, 3400, 3350, 3450, 5950, 3250, 3300, 3400, 3350, 3250

a. [2] Calculate the range and the interquartile range.

$R = \max - \min = 5950 - 3000 = 2950 \text{ MB}$
$n = 11$
$Q_3 = X_9 = 3400 \text{ MB}$
$Q_1 = X_3 = 3250 \text{ MB}$
$IQR = Q_3 - Q_1 = 150 \text{ MB}$

The range is 2950 MB, and the interquartile range is 150 MB.

b. [2] Identify any outliers in this data set. You must show the necessary calculations to support your answer.

Lower fence= $Q_1 - 1.5 \times IQR = 3250 - 1.5 \times 150 = 3025 \text{ MB}$
Upper fence= $Q_3 + 1.5 \times IQR = 3400 + 1.5 \times 150 = 3625 \text{ MB}$

There are two outliers: $X = 3000$ and $X = 5950$.

**Question 4** [5 marks]

The ages, in years, of seven randomly chosen BCIT students are recorded below.

| | | | | | | |
|---|---|---|---|---|---|---|
| 22 | 18 | 34 | 19 | 24 | 23 | 21 |

Determine any unusual values in this sample. Justify your answer with the necessary calculation(s).

$n = 7$

$\bar{X} = \frac{1}{n}\Sigma X_i = \frac{1}{7}(22 + 18 + 34 + 19 + 24 + 23 + 21) = \frac{161}{7} = 23$ years

$s = \sqrt{\frac{1}{n-1}\Sigma(X - \bar{X})^2}$

$= \sqrt{\frac{1}{6}((22-23)^2 + (18-23)^2 + (34-23)^2 + (19-23)^2 + (24-23)^2 + (23-23)^2 + (21-23)^2}$

$= \sqrt{\frac{1}{6}(168)} = 5.29$ years

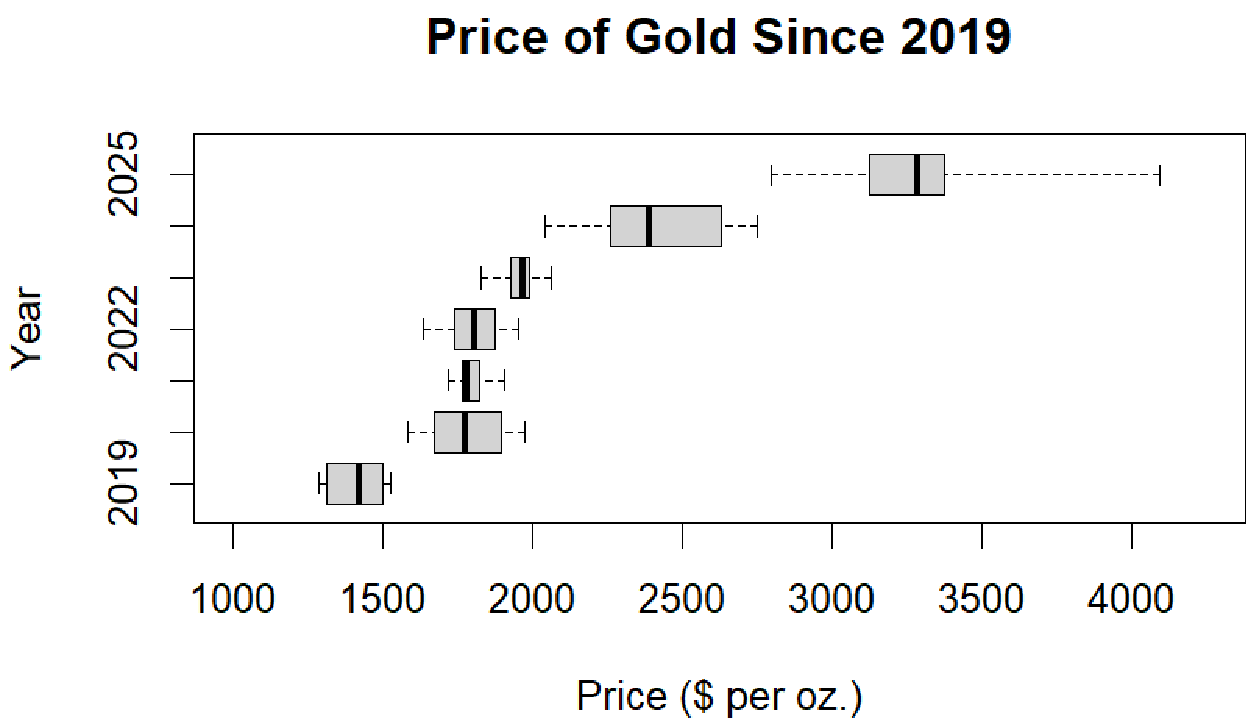
An unusual value would have $Z > 2$ or $Z < -2$. This corresponds to $X$ values being

$$X > \bar{X} + 2s = 23 + 2 \times 5.29 = 33.6$$
$$X < \bar{X} - 2s = 23 - 2 \times 5.29 = 12.4$$

There is one unusual age, which is $X = 34$.

**Question 5** [4 marks]

The chart below shows the inflation-adjusted price of gold ($ per oz) since 2019, grouped by year (2019 to 2025).




For each statement below, write TRUE or FALSE and briefly explain your answer.

a. [2] The median price of gold has more than doubled from 2019 to 2025.

TRUE
The median price in 2019 was below 1500 $ per oz while the median price in 2025 was approximately 3300 $ per oz, which is more than double.

b. [2] The variability (or "volatility" in finance terms) of the price of gold has steadily increased from 2019 to 2025.

FALSE
Variability in this context is best represented by the interquartile range (IQR), which is the width of the box in the box plots. There is no clear trend in the IQR. For instance, 2024 has the largest IQR while 2020 and 2025 are approximately the same.

If you interpret variability as the range, $R = \text{Max} - \text{Min}$, then variability is visibly larger in 2025 and 2024 than in prior years. However, you cannot say that range has *steadily* increased since 2019; furthermore, range is not as suitable as IQR as an indicator of variability, since range reflects only the most extreme values.

**Question 6** [8 marks]

A cybersecurity team monitors login attempts to a company's server and classifies them based on origin and legitimacy. Over a month, they recorded 1100 login attempts that were classified as follows:

| Origin | Legitimacy | | |
|---|---|---|---|
| | **Legitimate** | **Suspicious** | **Malicious** |
| **Internal Network** | 420 | 30 | 10 |
| **VPN** | 250 | 60 | 40 |
| **External IP** | 80 | 90 | 120 |

a. [1] What is the probability that a randomly selected login attempt is **Malicious**?

$$P(\text{Malicious}) = \frac{10 + 40 + 120}{1100} = 0.1545$$

b. [2] What is the probability that the login was flagged as **Malicious**, given that it

i. originated from the **Internal Network**?

$$P(\text{Malicious} \mid \text{Internal Network}) = \frac{10}{420 + 30 + 10} = 0.0217$$

ii. originated from a **VPN**?

$$P(\text{Malicious} \mid \text{VPN}) = \frac{40}{250 + 60 + 40} = 0.1143$$

iii. originated from an **External IP**?

$$P(\text{Malicious} \mid \text{External IP}) = \frac{120}{80 + 90 + 120} = 0.4138$$

c. [2] Are the outcomes **Malicious** and **External IP** independent? Justify your answer with the appropriate mathematical calculation(s).

The events **Malicious** and **External IP** are only independent if

$$P(\text{Malicious}) = P(\text{Malicious} \mid \text{External IP})$$

In this case they are not, since $0.1545 \neq 0.4138$.

The two events are dependent.

d. [3] Suppose that in the subsequent month the cybersecurity team reports that 49% of login attempts came from the internal network, 36% from a VPN and 15% from an External IP. If we assume that the proportion of malicious attempts from each of these origins is the same as you calculated in part b., what is the probability that a given **Malicious** login attempt originated from an **External IP**?

Abbreviate using: M = Malicious, Int = Internal IP, V = VPN, Ext = External IP.

Bayes' Theorem gives:

$$P(\mathrm{Ext} \mid \mathrm{M}) = \frac{P(\mathrm{Ext}) \cdot P(\mathrm{M} \mid \mathrm{Ext})}{P(\mathrm{Int}) \cdot P(\mathrm{M} \mid \mathrm{Int}) + P(\mathrm{V}) \cdot P(\mathrm{M} \mid \mathrm{V}) + P(\mathrm{Ext}) \cdot P(\mathrm{M} \mid \mathrm{Ext})}$$

$$= \frac{0.15 \times 0.4138}{0.49 \times 0.0217 + 0.36 \times 0.1143 + 0.15 \times 0.4138}$$

$$= 0.5452$$

**Question 7** [4 marks]

For an upcoming election, the true voting preferences among potential voters is:

- 47% prefer the Orange party
- 50% prefer the Black party
- 3% are undecided

a. Suppose you randomly select 40 potential voters and ask for their voting preference. Let $X =$ the number of potential voters in the sample who prefer the Orange party.

i. [1] What is the mean and standard deviation of $X$?

he number $X$ of potential voters who prefer the Orange party is a binomial random variable with $n = 40$ and $p = q = 0.47$. Therefore, its mean and standard deviation are:

$\mu = np = 40 \times 0.47 = 18.8$

$\sigma = \sqrt{npq} = \sqrt{40 \times 0.47 \times 0.53} = 3.16$

ii. [2] What is the probability that $X = 19$ exactly?

$P(X = 19) = C(n, x)p^x q^{n-x} = C(40, 19) \cdot 0.47^{19} \cdot 0.53^{21} = 0.1253$

b. [1] Now suppose you randomly select potential voters one at a time, asking each for their voting preferences. What is the mean number of potential voters you will need to ask to find your first undecided potential voter?

The number of trials to get the first undecided potential voter is a geometric variable with $p = 0.03$. Therefore,

$$\mu = \frac{1}{p} = \frac{1}{0.03} = 33.33$$

**Question 8** [5 marks]

On average, a major flood in the Fraser Valley occurs once every 20 years. Assume that floods occur independently and with equal likelihood at any time of the year.

a. [2] What is the probability that *two* major floods occur in the Fraser Valley in the next 10 years?

Let $X =$ the number of major floods in the Fraser Valley in the next 10 years. Under the given assumptions, this is a Poisson variable with mean value

$$\lambda = \frac{1}{20\text{ yr}} \times 10\text{ yr} = 0.5$$

Therefore

$$P(X = 2) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-0.5} \cdot \frac{0.5^2}{2!} = 0.0758$$

b. [3] What is the probability the *next* major flood in the Fraser Valley will occur between 10 and 20 years from now?

This event is equivalent to saying $X = 0$ for the next 10 years and then $X \neq 0$ for the following ten years after that (which is independent of the first 10 years). Therefore

$$P(\text{next flood is 10 to 20 years from now}) = P(X = 0) \cdot \big(1 - P(X = 0)\big)$$

$$= e^{-0.5} \cdot \frac{0.5^0}{0!} \cdot \left(1 - e^{-0.5} \cdot \frac{0.5^0}{0!}\right) = 0.2387$$

[Note: this is equivalent also to a geometric variable problem with $p = 1 - e^{-0.05}$.]

**Question 9** [4 marks]

A lab contains 30 computers. For unknown reasons, 5 of the computers cannot boot-up. Suppose a set of 20 students randomly choose their own computer.

a. [2] Find the probability that 3 students choose computers that cannot boot-up.

$N = 30$
$K = 5$
$N - K = 25$
$n = 20$
Let $X$ = the number of students who choose a computer that cannot boot up.
Then

$$P(X = 3) = \frac{C(5,3) \cdot C(25,17)}{C(30,20)} = \frac{10 \cdot 1081575}{30045015} = 0.3600$$

b. [1] Find the probability that all students get a computer that can boot-up.

Using the same parameters, we get

$$P(X = 0) = \frac{C(5,0) \cdot C(25,20)}{C(30,20)} = \frac{53130}{30045015} = 0.0018$$

c. [1] Determine the *expected* number of students who choose a computer that cannot boot-up.

For a hypergeometric variable,

$$E[X] = n \cdot \frac{K}{N} = 20 \cdot \frac{5}{30} = 3.33$$

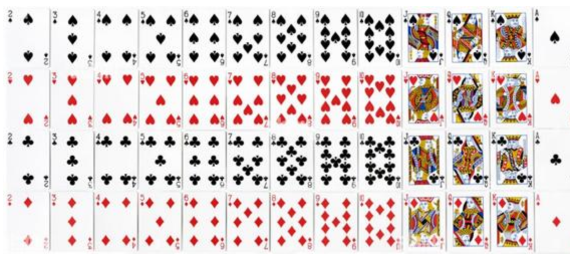
On average we expect 3.33 students to choose a computer that cannot boot-up.

**Question 10** [5 marks]

The rules of a gambling game are as follows:

- You draw 3 cards, without replacement, from a standard deck of cards.
- If all three cards are the same rank (e.g., 3 Kings), then you win $200.
- If all three cards are different ranks (e.g., Seven, King, Ace), then you win $2.
- Otherwise, you win nothing.

After each game the 3 cards drawn are replaced in the deck and the deck is reshuffled.



a. [1] If you play this game once, what is the probability that you win exactly $2?

$$P(\text{win } \$2) = \frac{52}{52} \times \frac{48}{51} \times \frac{44}{50} = 0.8282$$

b. [2] If you play this game repeatedly, how much do you win per game on average?

$$\begin{aligned} &E[\text{Winning}] = \$200 \cdot \mathrm{P}(\$200) + \$2 \cdot \mathrm{P}(\$2) + \$0 \cdot \mathrm{P}(\$0) \\ &= \$200 \cdot 0.002353 + \$2 \cdot 0.828235 + \$0 \\ &= \$2.13 \end{aligned}$$

c. [2] If you play this game twice, what is the probability that you win at least $5 in total?

The only way to win at least $5 is to win $200 on one or both games.

The probability of winning $200 in one game is $P(\$200) = \frac{52}{52} \times \frac{3}{51} \times \frac{2}{50} = 0.002353$.

The outcomes of the games are independent, so the probability we want is:

$$\begin{aligned} P(\text{win} \geq \$5) &= P(\$200 \text{ game } 1) + P(\$200 \text{ game } 2) - P(\$200 \text{ game } 1 \text{ and game } 2)) \\ &= 0.002353 + 0.002353 - 0.002353^2 = 0.0047 \end{aligned}$$