

Unit #3 Test – Displaying and Analyzing DataK/U: _____
20Application: _____
19Communication: _____
11TOTAL: _____
50**Knowledge/Understanding: 20 marks****Multiple Choice:** (Circle the appropriate answer)

<12 marks>

1. Which of the following is NOT a measure of central tendency?

- a) Mode
- (b) Standard Deviation**
- c) Median
- d) Mean

2. Which of these statements can be concluded from the pictograph to the right:

- a) The population of Quebec is 700 000
- b) Nova Scotia is the least populated province in Canada
- (c) The population of Ontario is approximately double that of B.C.**
- d) The population of Quebec is increasing

Province	Population
Ontario	7 people
Quebec	5 people
B.C.	3 people
Nova Scotia	1 person

3. The variable in the question “what is the average height of grade 12 males?” is:

- a) Continuous**
- b) Discrete
- c) Open
- d) Closed

4. A population refers to:

- a) A portion of a group about which data are being collected
- b) The number of individuals in a cluster
- c) The entire group about which data are being collected**
- d) The number of individuals within a given category

5. A sample selected simply because it is easily accessible is known as:

- a) Voluntary-response Sampling
- (b) Convenience Sampling**
- c) Cluster Sampling
- d) None of the above

6. Which of the following scenarios will NOT result in a sampling bias?

- a) The regular meeting time of a group is determined by people who attend a Monday morning meeting.
- b) American food preferences are determined by a random sample of California residents.
- c) Hourly traffic flow is measured by a man who counts cars on his lunch break.
- d) None of the above

7. How can you avoid a non-response bias?

- a) Post a survey on your Facebook page because all your friends have access to it.
- b) Hand a survey out to homeroom classes and wait for them to be completed.
- c) Attach comment cards to the bill at the restaurant you work at.
- d) Send a questionnaire through email so people can easily respond.

8. What type of closed survey question is the following: "How would you describe your satisfaction with your current mark in Data Management on a scale of 1 – 5, with 5 being extremely satisfied?"

- a) Information
- b) Ranking
- c) Checklist
- d) Rating

9. A box-and-whisker plot is a graphical representation of:

- a) Stratified random samples
- b) Median, upper and lower quartiles
- c) Mean, median and mode
- d) Standard deviation

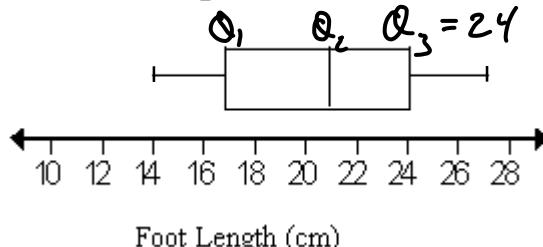
10. What is the "1.5 Rule" for determining outliers?

- a) $Q_3 + 1.5(IQR)$
- b) $Q_1 - 1.5(IQR)$
- c) $Q_2 + 1.5(IQR)$
- d) Both a) and b)

11. For the box-and-whisker plot the right, Q_3 is:

- a) 17
- b) 21
- c) 24
- d) 26

Foot Lengths of Grade 9 Students



12. A question on a survey where the respondent can answer in his/her own words is:

- a) A qualitative question
- b) An information question
- c) An open question
- d) A closed question

Problems:

13. Determine the mean, median, and mode for the following set of bowling scores:

$$\{181, 177, 154, 181, 177, 184, 262\} \quad n=7$$

<4 marks>

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{n} \\ &= \frac{181+177+154+181+177+184+262}{7} \\ &= 188 \quad \therefore \text{The mean is } 188\end{aligned}$$

Sorted:

$$154, 177, 177, 181, 181, 184, 262$$

$$\begin{aligned}med &= \frac{n+1}{2} \\ &= \frac{7+1}{2} \\ &= 4 \quad \therefore \text{The median is } 181.\end{aligned}$$

$\therefore 177$ and 181 appear the most frequently (twice),
the mode is 177 and 181 .

14. A student's marks for his science course are weighted as follows:

$$K/U = 35\% \quad App = 30\% \quad Thinking = 15\% \quad Comm = 20\%$$

His marks in the four categories are as follows:

$$K/U = 87\% \quad App = 91\% \quad Thinking = 78\% \quad Comm = 88\%$$

- a) Calculate his term mark before the exam.

<2 marks>

$$\begin{aligned}Mark_k &= \frac{\sum w_i x_i}{\sum w_i} \\ &= \frac{(35)(87) + (30)(91) + (15)(78) + (20)(88)}{35+30+15+20} \\ &= 87.05 \quad \therefore \text{His term mark before the exam is } 87.05\% \text{ or } 87\%\end{aligned}$$

- b) What mark must this student achieve on the final exam to earn a final grade of 90% if his term marks are worth 70% of his course mark?

<2 marks>

Using the term mark from a),

$$w_T m_T + w_E m_E = f$$

$$(0.70)(87.05) + (0.30)m_E = 90$$

$$m_E = \frac{90 - (0.70)(87.05)}{0.3}$$

$= 96.88\% \quad \therefore \text{The student must achieve a } 96.88\% \text{ (or a } 97\%) \text{ to earn a final grade of } 90\%.$
 $\approx 97\%$

Application: 19 marks

Name: _____

1. Calculate the mean, the median and modal intervals for the following data set:

<4 marks>

Age	Frequency	Median	$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$
0-4	8	2.5	
5-9	6	5	
10-14	16	7.5	$= \frac{8(1.5) + 6(5) + 16(7.5) + 15(10) + 10(12.5)}{8+6+16+15+10}$
15-19	15	10	
20-24	10	12.5	≈ 8.091
			$med = \frac{\sum f_i + 1}{2}$
			$\therefore \text{The mean is about } 8.091.$

2. For the following data:

Midterm Marks	54	80	12	61	71	66	92	81	86	61	72	74	72	71	91	68	50	84
---------------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

range,
∴ The median is within the
15-19 range.

∴ 33 falls in the 15-19

range,
∴ The median is within the
15-19 range.

Sort:

12 12 41 50 54 61 61 63 66 71 72 74 80 80 81 84 91 92
 $n = 18$

- a) determine the Quartile values:

$$Q_1 \text{ at } \frac{n+1}{2} = \frac{18+1}{2} = 9.5$$

$$Q_1 \text{ at } \frac{n+1}{2} (1^{\text{st}} \text{ half}) = \frac{9+1}{2} = 5$$

$$Q_2 = \frac{66+71}{2} = 68.5 \quad Q_1 = 54$$

$$Q_3 \text{ at } \frac{n+1}{2} (2^{\text{nd}} \text{ half}) = \frac{9+1}{2} = 5 \quad Q_4 = 100\% \text{ data} = 92.$$

$$Q_3 = 80$$

∴ The quartiles Q_1, Q_2, Q_3, Q_4 are: 54, 68.5, 80, 92, respectively.

- b) determine the Interquartile Range:

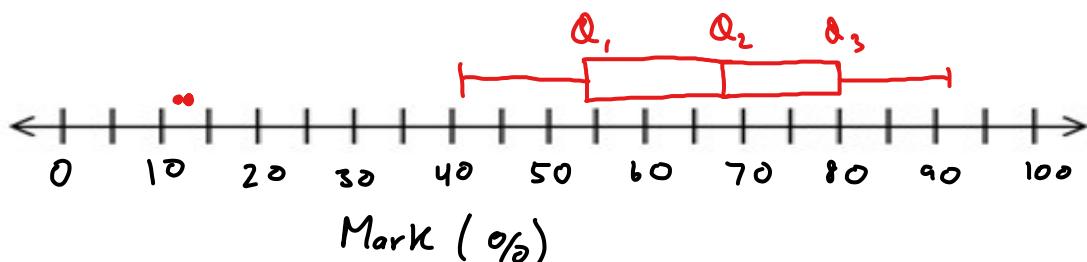
$$IQR = Q_3 - Q_1 \\ = 80 - 54 \\ = 26$$

∴ The interquartile range is 26.

- c) sketch a box and whisker plot, be sure to include all key features:

<3 marks>

Midterm Marks



3. Listed below are the number of tech-support questions successfully answered each day by Brittney and Jenn over a one-week period.

Brittney: 23, 20, 24, 21, 23

$$n=5$$

Jenn: 22, 28, 25, 19, 21

$$n=5$$

- a) Compare measures of central tendency – who has the better average? Explain your decision. <4 marks>

Brittney:

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} \\ &= \frac{23+20+24+21+23}{5} \\ &= 22.2 \text{ support calls / day}\end{aligned}$$

Jenn:

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} \\ &= \frac{22+28+25+19+21}{5} \\ &= 23 \text{ support calls / day}\end{aligned}$$

∴ Over this one-week period, Jenn successfully answered an average of 23 support calls a day while Brittney successfully answered an average of 22.2 support calls a day. For this week, Jenn has the higher (and thus better) average.

- b) Determine the standard deviation for both. Who is more consistent? How can you tell? <4 marks>

Brittney:

Day	Successfully Answered Support Questions	Deviation ($x_i - \bar{x}$)	$(x_i - \bar{x})^2$
Monday	23	0.8	0.64
Tuesday	20	-2.2	4.84
Wednesday	24	1.8	3.24
Thursday	21	1.2	1.44
Friday	23	0.8	0.64
Total	111	6.8	10.8

$$\bar{x} = 22.2$$

↳ Intentionally
only considered
magnitude for display

Jenn:

Day	Successfully Answered Support Questions	Deviation ($x_i - \bar{x}$)	$(x_i - \bar{x})^2$
Monday	22	1	1
Tuesday	28	5	25
Wednesday	25	2	4
Thursday	19	-4	16
Friday	21	2	4
Total	115	14	50

$$\bar{x} = 23$$

$$\begin{aligned}s &= \sqrt{\frac{(x_i - \bar{x})^2}{n-1}} \\ &= \sqrt{\frac{10.8}{5-1}} \\ &= 1.64\end{aligned}$$

∴ Brittney has a standard deviation of 1.64 in her daily support calls.

$$\begin{aligned}s &= \sqrt{\frac{(x_i - \bar{x})^2}{n-1}} \\ &= \sqrt{\frac{50}{5-1}} \\ &= 3.54\end{aligned}$$

∴ Jenn has a standard deviation of 3.54 in her daily support calls.

∴ Standard deviation measures a distribution's spread, a smaller standard deviation implies greater consistency. Brittney's standard deviation is smaller than Jenn's standard deviation. ∴ Brittney is more consistent than Jenn in their support calls.

Communication: 11 marks

1. Describe the differences between a bar graph and a histogram and explain what type of data you would display with each one. <3 marks>

Bar graphs compare categories of data in separate bars. The categories of data can be qualitative (red), but must map to a value.

Histograms show frequency of ranges of the original continuous data in connected bars. The data is continuous (ex. height of all gr. 12 students) and is organized into ranges (ex. 150cm - 159cm, 160cm - 169cm, etc.) for each range's frequency to be displayed in the histogram.

2. a) What are two different measures of spread discussed in class? <2 marks>

Standard deviation and Interquartile range are 2 measures of spread discussed in class.

- b) What do measures of spread tell us about a set of data?

Measures of spread tell us how closely the data is packed around the mean.

A high measure of spread would suggest the data is not closely packed around the mean (the data is "spread out") while a low measure of spread would suggest the data is close to the mean.

3. Describe the difference between sampling bias and response bias. Give an example of each. <3 marks>

Sampling bias: Occurs when your sampling technique results in different probabilities of selecting groups from the population. Ex: A survey to see if the school should upgrade their computers. You interview the computer science club. The CS club is predisposed to answering yes.

Response bias: Occurs when participants are more likely to answer questions incorrectly. Ex: A teacher wants to know how many times students have been in detention. Students are unlikely to tell the teacher and the teacher's authority will encourage students to lie. Response bias

Mathematical Form	Level R	Level 1	Level 2	Level 3	Level 4
Proper Use of formulas and notation.	Never uses formulas and/or notation properly	Rarely uses formulas and/or notation properly	Sometimes uses formulas and/or notation properly	Often uses the formulas and/or notation properly	Always uses formulas and/or notation properly

Completed within time given: 1C

Occurs when a participant is more likely to answer a survey incorrectly while sampling bias doesn't choose a representative or proportional sample to conduct a survey on.