MDM4U – Sample Test 3 – PPDAC – October 26, 2023

Name: Solutions

Knowledge	Application	Communication	Thinking	Total	Percent
21	24	19	16	80	%

Knowledge

1. What does PPDAC stand for? /1

Problem	Plan	Data	Analysis	Conclusions

2. Identify the phase of PPDAC where each of the following occurs. /10

(a)	Calculate r	Analysis	(f)	Explain biases	Conclusion
(b)	Collect surveys	Data	(g)	Create graph	Analysis
(c)	Choose research question	Problem	(h)	Write up report	Conclusion
(d)	Decide sampling	Plan	(i)	Find average	Analysis
(e)	Replicate experiment	Data	(j)	Give out placebo	Data

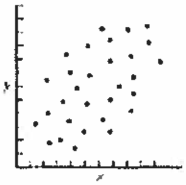
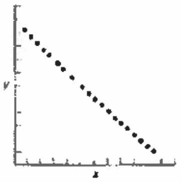
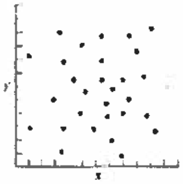
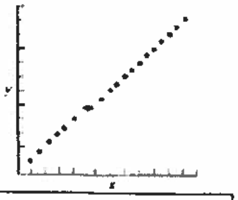
3. Read the following description. Answer the following questions about it. /10

NachoNacho® brand corn tortillas are made at a large facility in Brampton that produces 100,000 tortillas per day on two production lines. A government inspector visited the facility on Oct 22, 2019, to determine if they comply with their advertised diameter of 6 inches. Using a random number generator, a computer selected 328 tortillas to test from both production lines. The mean was 6.012 inches, which was acceptable.

a)	Is it Causal or Descriptive?	Descriptive
b)	How much Replication?	328
c)	Sampling Technique?	Random
d)	Random Assignment?	No
e)	What is the Research Question?	What is the mean size of NachoNacho tortillas?
f)	Identify the variable data was collected about.	size of tortilla
g)	Identify the calculation that occurred.	average size of tortilla
h)	Identify the Problem Unit	NachoNacho tortilla
i)	Identify the Plan Unit	NachoNacho tortilla from Brampton on Oct 22, 2019
j)	What are the Biases or Diversity Limitations?	only on one day

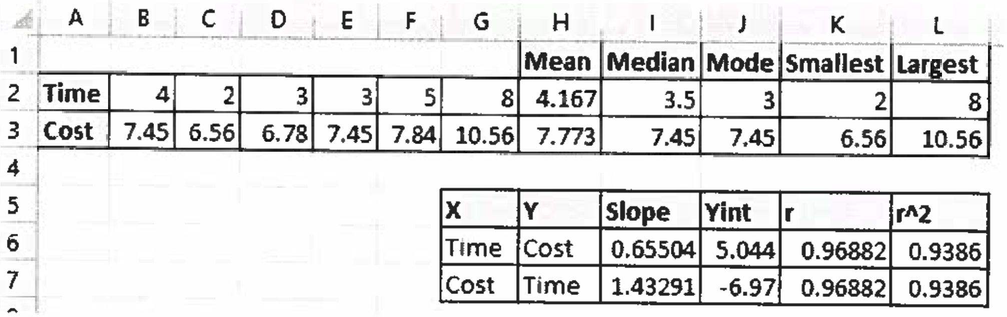
Application

4. For each graph, fill in the chart. /8

r Estimate (1, 0.7, 0, -0.7, -1)	0.7	-1	0	1
Positive/Negative	Positive	Negative	None	positive
Strength of Relationship	Moderate	Strong	None	Strong

5. What are the formulas found in the indicated cells of this spreadsheet? /10



	A	B	C	D	E	F	G	H	I	J	K	L
1								Mean	Median	Mode	Smallest	Largest
2	Time	4	2	3	3	5	8	4.167	3.5	3	2	8
3	Cost	7.45	6.56	6.78	7.45	7.84	10.56	7.773	7.45	7.45	6.56	10.56
4												
5							X	Y	Slope	Yint	r	r^2
6							Time	Cost	0.65504	5.044	0.96882	0.9386
7							Cost	Time	1.43291	-6.97	0.96882	0.9386

H2	=average (B2:G2)
I2	=median (B2:G2)
J2	=mode (B2:G2)
K2	=min (B2:G2)
L2	=max (B2:G2)

I6	y=cost x=time =slope (B3:G3, B2:G2)
I7	y=time x=cost =slope (B2:G2, B3:G3)
J6	y=cost x=time =intercept(B3:G3, B2:G2)
K6	=correl (B3:G3, B2:G2)
L6	=rsq (B3:G3, B2:G2)

6. What 3 things are required to prove causation? /3

7. (a) Calculate the proportion of the variance of Y which cannot be explained by the variance of X if r = 0.5? Show your work. /3

$r^2 = (0.5)^2$

$= 0.25$

∴ 25% of the variance of Y depends on X

75% cannot be thus explained.

(b) What two possible r values will mean that 81% of the variance of X depends on the variance of Y? Show your work.

$r^2 = 0.81$

$r = \pm\sqrt{0.81}$

$= \pm 0.9$

∴ r will be either +0.9 or -0.9 (negative or positive slope) if 81% of Y's variance depends on X.

Communication

8. Write the term indicated in the last column. /10

(a)	A company (e.g., Blue Kai) who finds trends in large data sets.	Data Miner
(b)	r is the ____________ co-efficient. [Fill in the blank]	Correlation (or Pearson)
(c)	r² is the co-efficient of __________. [Fill in the blank]	Determination
(d)	A pioneer in self-tracking. Worked for Facebook's timeline division.	Nicholas Felton
(e)	The method of eliminating spuriousness as a possibility.	Random Assignment
(f)	One variable or Two variable: A histogram.	One variable
(g)	Sampling technique used when you post a poll on social media.	Voluntary / Self selection
(h)	By observing people, you change their behaviour.	Hawthorne Effect
(i)	When Nike funds a study on the quality of running shoes; A bias.	Bias Due To Funding
(j)	When specific conclusions are expanded to other situations; A bias.	Transferring Findings

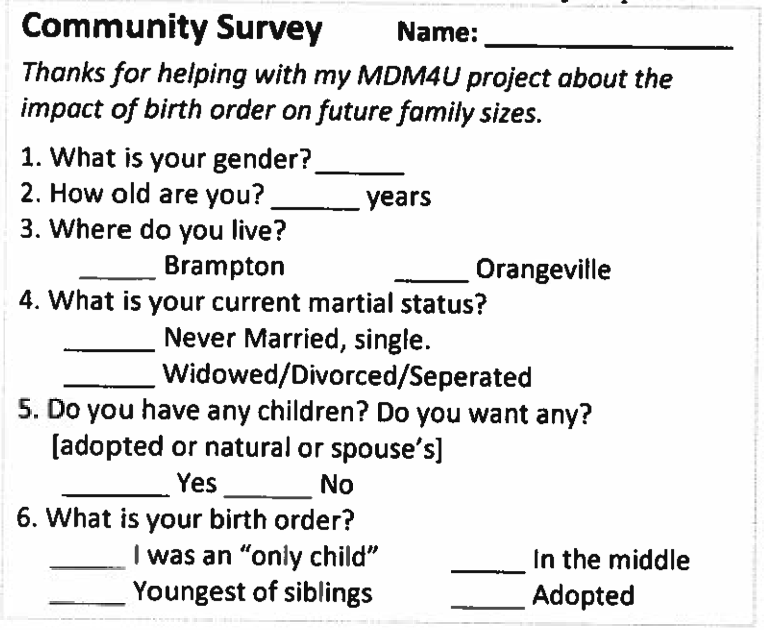
9. A researcher finds the following: ↑ sleep correlates to ↑ test scores. What things could be happening? /3

↑ sleep causes ↑ test scores	↑ test scores causes ↑ sleep	something else causes both ↑ test scores and ↑ sleep

10. Define and explain the importance of the term "double blinding". /2

Double blinding is when neither the subject nor the researcher knows who is in the control group and who is in the test group. This is important because subject and researcher expectations do not enter into the results of the study because neither knows which group applies to which subject.

11. Identify 4 distinct errors with this survey. Explain each briefly. /4



Community Survey **Name:** ________________

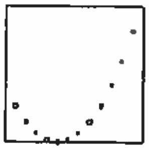
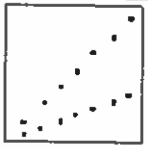
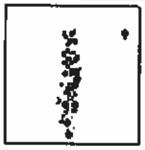
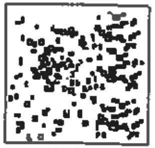
Thanks for helping with my MDM4U project about the impact of birth order on future family sizes.

1. What is your gender? ______
2. How old are you? ______ years
3. Where do you live?
   _____ Brampton _____ Orangeville
4. What is your current martial status?
   ______ Never Married, single.
   ______ Widowed/Divorced/Seperated
5. Do you have any children? Do you want any? [adopted or natural or spouse's]
   ________ Yes ______ No
6. What is your birth order?
   _____ I was an "only child" _____ In the middle
   _____ Youngest of siblings _____ Adopted

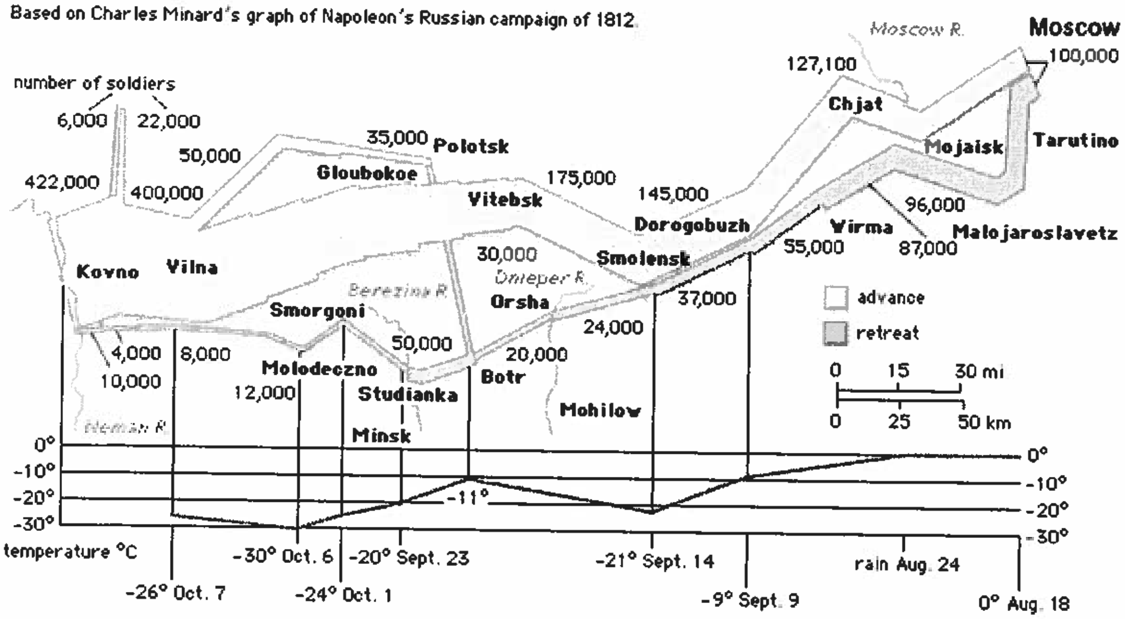
- Not anonymous: asks for name. - People are more likely to answer with a social desirability bias. - Leave off.
- Asks for age. - Especially older subjects are likely to round to nearest 5 or 10. - Ask for birth year instead.
- Missing options for #4 (married) and #6 (oldest) - Proof read better or offer an "other: ___" option
- #5 is a double-barrelled question (a- any children, b- want any) - Break into 2 seperate questions

Thinking

12. Why shouldn't you find a line of best fit for any of these sets of data? /4

 points form a curve or a parabold. Not a line, so no line of best fit	 appears to be 2 lines of best fit. This indicates an error in experiment design.	 many y values for the same x yields an infinite slope. This is not likely to be a real world situation.	 points are too scattered. They don't form a line, they form a blob. There is no correlation

13. This graph is considered the greatest graph ever created. Not limited to a mere one or two variables, Charles Minard graphed six variables all on one graph. What are they? /6



The 6 variables:

advance/ retreat
temperature
number of soldiers
date
latitude
longitude

14. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline: "Study Finds that Smoking Causes Alzheimer's." The article reported that a doctor had reviewed the medical histories from a hospital archive of 21,123 men documenting 23 years. The article also stated that, for those who smoked at least 2 packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

What are the 3 most important errors in the reporting of this study? Explain briefly. /6

① Correlation ≠ Causation Mixup: headline clearly implies causation (they use that word) yet study itself says "risk", which is correlation.

② No Random Assignment: (or other methods to handle this) This means spuriousness was not eliminated, there could be another factor.

③ No Random Selection: While replication is high (21,123), it doesn't say how people were selected. A biased sample yields biased results no matter how large the sample.