

Name: \_\_\_\_\_

1. For each of the two scenarios identify whether the rows of the example data are a statistically independent sample, or if the rows are dependent on each other. Please provide a one or two sentence explanation.

- (a) A web-store has collected sales data of unrelated clients, who are geographically dispersed. Each record contains the total sales made to a single client in the previous fiscal year. Empty registration dates indicate the client created an account before the start of the previous fiscal year.

Client	Sales	Sessions	Shipments	City	Registration	...
A	\$500	25	2	X	June 5	...
B	\$2000	50	5	Y		...
C	\$250	12	1	Z	October 23	...
D	\$1200	61	4	W		...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

- (b) A social-media application has collected the time evolution of the self-reported relationships between pairs of users. Each relationship is represented by a pair of records, each containing a pair of users, the type of relationship, and the time-span of the relationship.

Request User	Relation	Respond User	Start	End	...
A	Daughter	B			...
B	Mother	A			...
A	Friend	C	March 5		...
C	Friend	A	March 5		...
D	Employee	B	May 23	December 15	...
B	Employer	D	May 23	December 15	...
⋮	⋮	⋮	⋮	⋮	⋮

2. A taxi company has collected shift length and number of ride statistics over a fixed time period. Unfortunately the data report has overstated the precision of measurement in an effort to improve data quality, using partial fractional attribution for records overlapping with the boundaries of the time period.

The report stated that there were 2766.45 minutes of driver-shift recorded, but the source data system only records shift lengths in increments of 10 minutes.

Furthermore, the report stated there were 212.85 rides. We know that it is only possible to measure ride counts in increments of 1.

- (a) Using rounding at half away from zero please round each measurement to the reliable digits, and provide the marginal uncertainty.

- (b) Assuming rides are requested randomly and independently from the drivers, find the marginal uncertainty of the ratio of rides over driver-shift length, and provide the units of measurement of the ratio through dimensional analysis. Please show your work.

3. It has been purposed to implement random COVID screening of students using a rapid test. Currently 1 in 1000 Albertans have COVID. Please use the outlined Rapid Test Confusion Matrix in your answer.

Test Result	Actual COVID Infection	
	Absent	Present
Positive	1/100 False Positive	2/3 True Positive
Negative	99/100 True Negative	1/3 False Negative

- (a) If a student tests positive what is the incorrect conclusion reached by the Prosecutor's Fallacy, please specify either of the incorrect probabilities from the confusion matrix.

- (b) What is the correct conclusion and probability of being infected, please show your work.