Data Analysis: IA02 (75 points)

Overall Instructions

1. Assignments are graded for accuracy after the deadline. During the assignment, you will not get feedback on whether your answer is correct or not.
2. Follow instructions here and enter your answers into the Blackboard assignment.
3. This assignment covers:
    a. summer probability series
    b. week 1: descriptive analytics
    c. week 2: distribution, z scores, visualization
4. For some questions, you will be asked to use the class dataset of real estate properties. For others, you will be asked to use a built-in dataset available in JASP.
5. Write probability & percentages in decimal format only (no % sign or fractions), with up to 3 decimal places.
6. When using JASP: Always copy values exactly from JASP.
7. When using Excel/Google Sheets: Always round numeric values to 3 decimal places.
8. You are allowed to make an unlimited number of submissions attempts before the deadline. Only the most recent submission will be used for grading purposes.
9. Be sure to hit the submit button to complete the submission process.

Section 1: 3 points per question

Answer the following questions based on the FROZEN class dataset.

(1) Which of the following are two mutually exclusive events?

A. being from Nigeria and being from Virginia
B. being a woman and having active duty status
C. having a two-car garage and having four bedrooms
D. having purchased a bicycle and a swimming pool

(2) If **SQFT** is recoded into categories (Small/Medium/Large), the measurement scale changes from:

A. Ratio to Ordinal
B. Interval to Nominal
C. Ratio to Nominal
D. Ordinal to Ratio

(3) Which of the following is **NOT ratio-scaled**?

A. Age of property

B. SQFT
C. WalkScore (0–100 index)
D. Zipcode

(4) Using the class dataset, calculate the probability that a **property is a condo**, given that we already know that this property is in **ZIP code 23188**.

(5) The variable Zipcode is a/an _____ variable, and its scale of measurement is _____.
A. Quantitative; Ratio
B. Quantitative; Interval
C. Categorical; Nominal
D. Categorical; Ordinal

(6) Bath is counted in whole numbers. Therefore, it is _____ (Continuous/Discrete) and measured on a/an _____ scale.
A. Continuous; Interval
B. Continuous; Ratio
C. Discrete; Interval
D. Discrete; Ratio

(7) If we convert WalkScore into these ranks: Top 10%, Middle 50%, Bottom 40%, the WalkScore variable has been transformed into a _____ scale.
A. Nominal
B. Ordinal
C. Interval
D. Ratio

Use JASP to answer the following questions based on the FROZEN class dataset.

(8) Which of these variables has the largest **skewness**?
A. BED
B. Zestimate
C. Assessment
D. Price

(9) What's the **Standard Deviation** of the **age** of real estate properties included in the class dataset?

(10) What's the **Mean** of the log (base10) of the **OfferPrice** variable?

(11) What is the percentage of **property type "House" in ZIP code 23185**?

Use Excel/Google Sheets to answer these questions based on the FROZEN class dataset.

(12) The smallest Z score in the **SQFT** variable is _____.

(13) How many records from the class dataset have a Z score greater than 3 in the **Miles** variable?

Answer this question using Excel/Google Sheets and/or JASP

(14) Create a new variable (with any name) and compute it by adding 5 to the class data of **ZillowDays**. Find the **Variance** of this new variable.

Answer the following questions using the "Big Five Personality Traits" dataset from "4. Regression" in JASP's Data Library.

(15) If Bhavesh has an Openness score of 4, what is his Z score?

(16) If Abdullah has a Z score of -2 on the Conscientiousness scale, his Conscientiousness score would be_____.

Answer the following questions without using any dataset.

(17) For a corporate finance analysis, if the variable Dividends has a **variance** of 400, what is the base 10 log of its standard deviation value?

(18) When it comes to the extraversion dimension, Kendall's Z score is 2. Jessie has calculated the class average to be 40 and the class variance to be 81. Kendall's actual extraversion score should be ___.

(19) On a scale of 1 to 100 for extraversion, the class average is 60 and variance is 9. This suggests that the class is:

A. extremely extraverted.
B. extremely introverted (i.e., the opposite of extraverted).
C. not very extraverted nor very introverted.
D. full of students who score 100 on extraversion.

(20) Gabby analyzed client accounts and noticed that the average balance is 1000, with a variance of 2500. Which of the following customers would have a balance of 900?

A. David with a Z score of +2
B. Connor with a Z score of -100
C. Agda with a Variance of -2
D. Holden with a Deviation of +100

Section 2: 1 point

Ask at least one question about materials we have covered so far in class.

Section 3: 14 points

Using the FROZEN dataset, create these 4 visualization worksheets in a single Excel (.xlsx) file, or Tableau (.twbx) file. You only need to submit one, not two, files.

Sheet 1: Numeric: Visualize a single numeric variable

Sheet 2: Categorical: Visualize a single categorical variable

Sheet 3: NumericCategorical: Visualize a numeric variable by a categorical variable

Sheet 4: NumericNumeric: Visualize a numeric variable by a numeric variable

Save your Excel as a .xlsx file with your full name (e.g., IA02 RachelJones.xlsx).

Save your Tableau work as a .twbx file with your full name (e.g., IA02 RachelJones.twbx).

Upload the .xlsx or .twbx file on the blackboard for submission in the "Additional Content" box. Please make sure the file is uploaded correctly before submitting this assignment.

Note: Blackboard shows a preview of your Excel file which may appear to be poorly formatted. If you can see a preview, your file is fine and there's no need to worry about file distortion.