**Fall 2023**

**Dr. Levkoff**

**BUAN 314 / 370 SAMPLE EXAM 2**

**Directions: The exam consists of 30 multiple choice questions. The first 20 are conceptual. The last 10 require you to analyze data and use software. Choose the *best* answer for each question. You may use your notes, your computer, and the internet, but are not allowed to communicate in real time with anyone else (or AI). You should submit two documents as attachments via email to** [**slevkoff@sandiego.edu**](mailto:slevkoff@sandiego.edu) **with the *exact* emailsubject heading “BUAN 314 370 EXAM 2 SUBMISSION”: 1) a spreadsheet file with a single column - the first column - filled in beginning with the top cell (row 1) with your full name. The next thirty cells (rows 2-31) should contain your answers to the 30 questions using LOWER CASE LETTERS ONLY. DO NOT include a separate column with question numbers or include the question numbers inside the cells with your answers. 2) You should fill out the EXAM 2 TEMPLATE.R file with the code used to solve the relevant exam questions in the space provided. Some code may be partially provided (similar to the DataCamp exercise structure). Once you’ve completed this, save the file as FIRSTNAME\_LASTNAME.R using your full FIRST NAME and LAST NAME. This should be the second attachment in the email. In the body of the email, type the following academic integrity pledge and insert your full name in the blank.**

“I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_understand that it is a violation of the University of San Diego academic integrity policy to provide or receive unauthorized assistance on this examination.”

1. This field allows the user of a table to *uniquely* identify each record in the table.
2. Foreign Key
3. Primary Key
4. House Key
5. DBA Key
6. Which of the following tools is useful for visually describing a business process with respect to how tables in a database are interrelated?
7. Chen Notation
8. Crow’s Foot Notation
9. ERDs
10. All of the above
11. Which of the following joins doesn’t have an explicit clause in SQL?
12. Inner Join
13. Outer Join
14. Self-join
15. Cross Join
16. If two sets are collectively exhaustive, then
17. The union of the sets must be the empty set
18. The intersection of the sets must be the empty set
19. The compliment of the union of the sets must be the empty set
20. None of the other choices are correct
21. Which of the following SQL clauses filters data *after* records have been grouped?
22. GROUP BY
23. WHERE
24. ORDER BY
25. HAVING
26. SQL is a type of
27. IDE
28. Statistical Computing Package
29. Structured Query Language
30. None of the above
31. Another name for the columns of a TIDY database table are
32. Records
33. Fields
34. Rows
35. Observations
36. Forecasting and simulation are considered to be types of \_\_\_\_\_\_\_\_\_\_\_ analytics techniques.
37. Predictive
38. Descriptive
39. Prescriptive
40. None of the above
41. A screenshot of a computer

    Description automatically generatedSuppose you run a national chain of coffee shops. You are given the information below to compare the 12 different products sold across the 20 states in which your business operates over the last year on the right:

Which of the following visualizations are most appropriate for

representing this data?

1. heat map
2. histogram
3. pie chart
4. scatter plot

A chart with different colored squares

Description automatically generatedQuestions 10)-11) reference the plot to the right:

1. Which genre has the lowest mean rating?
2. Science
3. Culture
4. Food and Drink
5. Not enough information provided
6. What alternate plot device would *best* represent the comparison being made in the plot above?
7. Barplots of the mean rating for each group
8. Violin plots of the rating variable facet wrapped by genre
9. Violin plots of the rating variable
10. A histogram plot of the rating variable

12) Practitioners can convert “soft” data into “hard data” is through the use of

1. a soft-hard data compiler
2. conjoint analysis
3. multifactorial design
4. none of the above

13) Which of the following most likely considers the tradeoff faced between the cost of acquiring more observations versus the improved accuracy in the larger data set?

1. sample design
2. sample plan
3. population sampling
4. population control

14)  When each observation has its own row, each variable its own column, and each type of variable/observation class has at most, one table per type, then the data is considered to be

1. raw
2. dirty
3. clean
4. tidy

15)  Which of the following techniques would result in having the largest number of rows in your data after preprocessing and cleaning the data with respect to handling null or missing values/responses?

1. Deletion rows of tidy data when a value is missing
2. Deleting columns of tidy data when a value is missing
3. Deletion only when necessary
4. Bootstrapping the values of missing observations
5. Which of the following is a useful aid for visually describing a cross-sectional relationship in a scatter plot?
6. geom\_smooth()
7. geom\_histogram()
8. geom\_path()
9. geom\_line()

17) This question refers to the dataset below:



Which of the the following best describes the structure of this dataset?

1. Longitudinal
2. Cross-Sectional
3. Time Series
4. Pooled (Cross-Section)

18) Suppose you have a TIDY dataset with thousands of cross-sectional observations on 10 variables. Which of the following cleaning operations is likely to result in the most amount of information lost assuming half of all observations contain null or missing values that are uniformly distributed across the different variables?

1. Deleting columns with any null or missing values
2. Deleting rows with any null or missing values
3. Resampling observations to use in place of null or missing values
4. Deletion of observations only when necessary
5. A graph showing the price of a company

   Description automatically generatedYour company’s analyst was asked to plot

The stock price over time for each month,

resulting in the visualization on the right:

A colleague suggests to “clean up” the

visualization by plotting the data *quarterly*

rather than *monthly*.

Which of the following would be an

implication of making this change?

1. It would distort the drop that occurred

early in the year.

1. The positive trend would no longer be

visible.

1. The data shown would no longer be accurate.
2. The line plot would no longer be the best plot device.

20) Which of the following are alternatives to monofactorial design A/B testing experiments that allow for multiple covariates to be changed at once?

1. Response surface modeling
2. Regression
3. Multifactorial design
4. All of the above