Review Quiz

Here is a short review quiz to prepare you for the midterm exam covering all the content in Modules 1 to 6 (inclusive).

Review Quiz

1. **Which of the statements below are correct?**

   I. Data science involves machine learning and statistics.

   II. A data scientist does not have to communicate results to stakeholders.

   - ○ a. Only I
   - ○ b. Only II
   - ○ c. Neither I nor II
   - ○ d. Both I and II

   ✓ Correct!

2. **What is the mean and median of the given data: 10, 12, 3, 11, 6, 33, 1, 9, 12, 11?**

   - ○ a. 10.8 and 10.5
   - ○ b. 10.8 and 11
   - ○ c. 11.5 and 11
   - ○ d. 10 and 11

   ✓ Correct!

3. **For a bell-shaped distribution, which of the following is true?**

   I. Mean is equal to the mode.

   II. Median is equal to the mode.

   - ○ a. Only I
   - ○ b. Only II

- ○ c. Neither I nor II
- ○ d. Both I and II

✓ Correct!

4. **What are the first and third quartiles for the given dataset: 10, 12, 3, 11, 6, 33, 1, 9, 12, 11??**

- ○ a. 6 and 11
- ○ b. 6 and 12
- ○ c. 5.75 and 11.75
- ○ d. 6.75 and 10.75

✓ Correct!

5. **In a survey, you are asked to rate customer service as bad, good, and best. These ratings are examples of:**

- ○ a. Qualitative data
- ○ b. Quantitative data

✓ Correct!

6. **In a survey, you are asked to rate customer service on a scale of 1-10. This rating is an example of:**

- ○ a. Qualitative data
- ○ b. Quantitative data

✓ Correct!

7. **What is the probability of rolling a die and getting a getting a value less than 3?**

- ○ a. 1/3
- ○ b. 2/3
- ○

c. 1/6

d. 1/2

✓ Correct!

8. **If you create a probability distribution of "rolling two dice", how many outcomes will be there in the table?**

a. 30

b. 15

c. 36

d. 20

✓ Correct!

9. **Which of the statements below are correct?**

I. Population mean and sample mean are computed using the same formula.

II. Population variance and sample variance are computed using the same formula.

a. Only I

b. Only II

c. Neither I nor II

d. Both I and II

✓ Correct!

10. **Under what condition do we reject the null hypothesis?**

a. If the difference between the sample mean and the hypothesized mean is too big.

b. If difference between the sample variance and the hypothesized variance is too big.

c. If difference between the sample median and the hypothesized median is too big.

d. All of the above.

✓ Correct!

11. **If the significance level for a test is 0.001 and the probability value computed for a hypothesis is 0.04, then we can reject the null hypothesis.**

- ○ a. True
- ○ b. False

✓ Correct!

12. **Which one of the following is an evaluation measure for classification models?**

- ○ a. Recall
- ○ b. Accuracy
- ○ c. Precision
- ○ d. All of the above

✓ Correct!

13. **Regression can be used to predict how much money a particular condo in downtown Toronto will sell for.**

- ○ a. True
- ○ b. False

✓ Correct!

14. **Which one of the following gives the accuracy?**

Imagine you are working on a project which is a binary classification problem. You trained a model on training dataset and get the confusion matrix on the validation dataset as follows:

	Predicted Class	

Actual Class	Positive	Negative
Positive	100	10
Negative	15	25

- a. 0.65
- b. 0.83
- c. 0.82
- d. 0.86

✓ Correct!

15. **From Question 14, which one of the following gives the precision?**

- a. 0.70
- b. 0.78
- c. 0.87
- d. 0.84

✓ Correct!

16. **From Question 14, which one of the following gives the recall?**

- a. 0.65
- b. 0.78
- c. 0.82
- d. 0.91

✓ Correct!

17. **Which one of the following algorithms does not require pre-existing labels?**

a. K-means
b. Decision trees

Check Your Answer

18. **What is the Euclidean distance between the two data points P: (10, 11) and Q: (12, 14)?**

a. 3.6
b. 10
c. 10.6
d. 12.4

✓ Correct!

19. **Which of the statements below are correct for cross-validation?**

I. Training data should be mutually exclusive from the test data.
II. The evaluation metrics are computed on both the training dataset and test dataset.

a. Only I
b. Only II
c. Neither I nor II
d. Both I and II

✓ Correct!

20. **Which of the statements below are correct for basic text processing?**

I. We can always use a “space” to identify words in word tokenization.
II. Stemming and Lemmatization mean converting words “trouble” and “troubling” to a meaningful base form “trouble.”

a. Only I
b. Only II

c. Neither I nor II

d. Both I and II

✓ Correct!