Feature Engineering quiz

8 out of 8 correct

1. Which of the following correlation coefficients measures the strength and direction of a linear relationship between two continuous variables?

- [x] Pearson correlation coefficient
- [ ] Spearman's rank correlation
- [ ] VIF
- [ ] None of the above

Explanation: Pearson correlation coefficient is used to measure the strength and direction of a linear relationship between two continuous variables. It ranges from -1 to +1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and +1 indicates a perfect positive correlation.

2. Which correlation coefficient is appropriate to use when the data is ordinal or non-parametric?

- [ ] Pearson correlation coefficient
- [x] Spearman's rank correlation
- [ ] VIF
- [ ] None of the above

Explanation: Spearman's rank correlation is a non-parametric measure of correlation, which is appropriate to use when the data is ordinal or non-parametric. It measures the strength and direction of the monotonic relationship between two continuous or ordinal variables.

3. What is the range of VIF values?

- ○ 0 to 1
- ○ 1 to 10
- ◉ 0 to infinity
- ○ None of the above

Explanation: VIF values range from 0 to infinity, where a VIF value of 1 indicates no multicollinearity, and values greater than 1 indicate increasing levels of multicollinearity.

4. Which of the following correlation coefficients is more robust to outliers?

- ○ Pearson correlation coefficient
- ◉ Spearman's rank correlation
- ○ VIF
- ○ None of the above

Explanation: Spearman's rank correlation is more robust to outliers because it is a non-parametric measure of correlation that uses ranks instead of actual data values. It measures the strength and direction of the monotonic relationship between two continuous or ordinal variables, which is less affected by extreme values than a linear relationship.

5. What is the range of the Pearson correlation coefficient?

- ◉ -1 to 1
- ○ 0 to 1
- ○ -∞ to ∞
- ○ None of the above

Explanation: The Pearson correlation coefficient is a measure of the linear relationship between two variables, and its value ranges from -1 to 1. A value of

-1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation.

6. What does a Spearman's rank correlation coefficient of -0.85 indicate?

- ◯ A strong positive correlation
- ◯ A moderate negative correlation
- ◉ A strong negative correlation
- ◯ No correlation

Explanation: The Spearman's rank correlation coefficient ranges from -1 to 1, with negative values indicating a negative correlation and positive values indicating a positive correlation. A value of -0.85 indicates a strong negative correlation between the two variables.

7. A researcher wants to study the relationship between two continuous variables in her dataset. Which correlation coefficient should she use if she suspects that the relationship may not be linear?

- ◯ Pearson correlation coefficient
- ◉ Spearman's rank correlation coefficient
- ◯ Both can be used
- ◯ None of the above

Explanation: Unlike the Pearson correlation coefficient, Spearman's rank correlation coefficient is based on the ranks of the data rather than the actual values, making it more suitable for detecting non-linear relationships.

8. A scientist is studying the relationship between the number of hours spent studying and the final exam scores of 50 students. He computes a Pearson correlation coefficient of -0.2. What does this result indicate?

- ◯ There is a strong negative correlation between the two variables
- ◉ There is a weak negative correlation between the two variables
- ◯ There is no significant correlation between the two variables

◯ There is a weak positive correlation between the two variables

Explanation: The Pearson correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. A coefficient of -0.2 indicates a weak negative correlation between the number of hours spent studying and the final exam scores. This means that as the number of hours spent studying increases, the final exam scores tend to decrease, but the relationship is not very strong. Therefore, option b) is the correct answer.

Submit