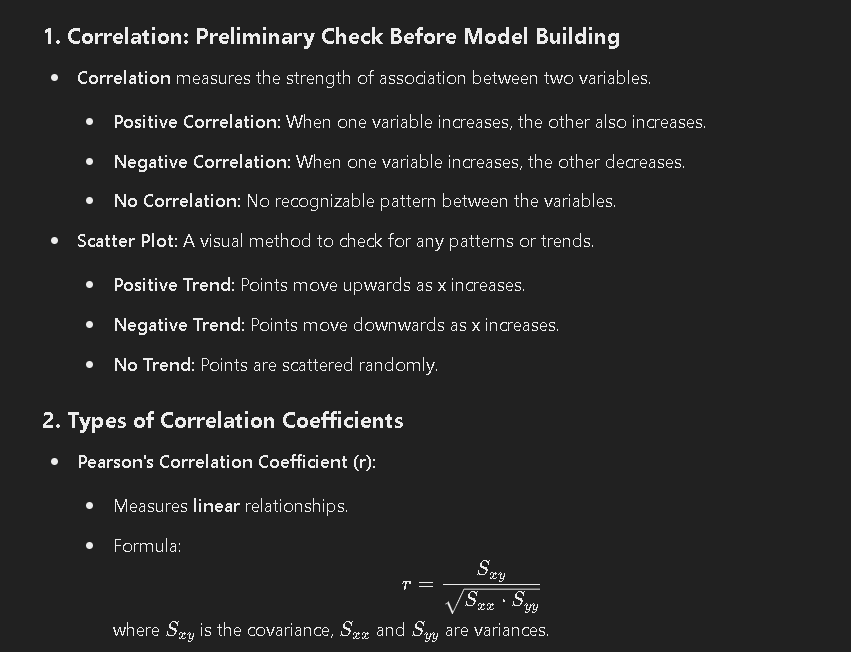
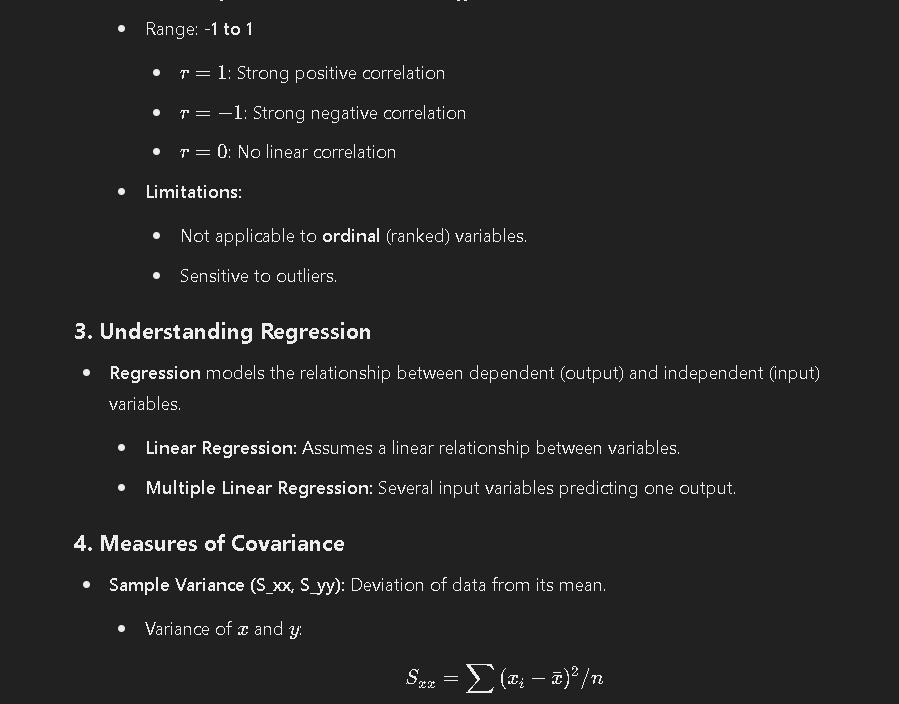
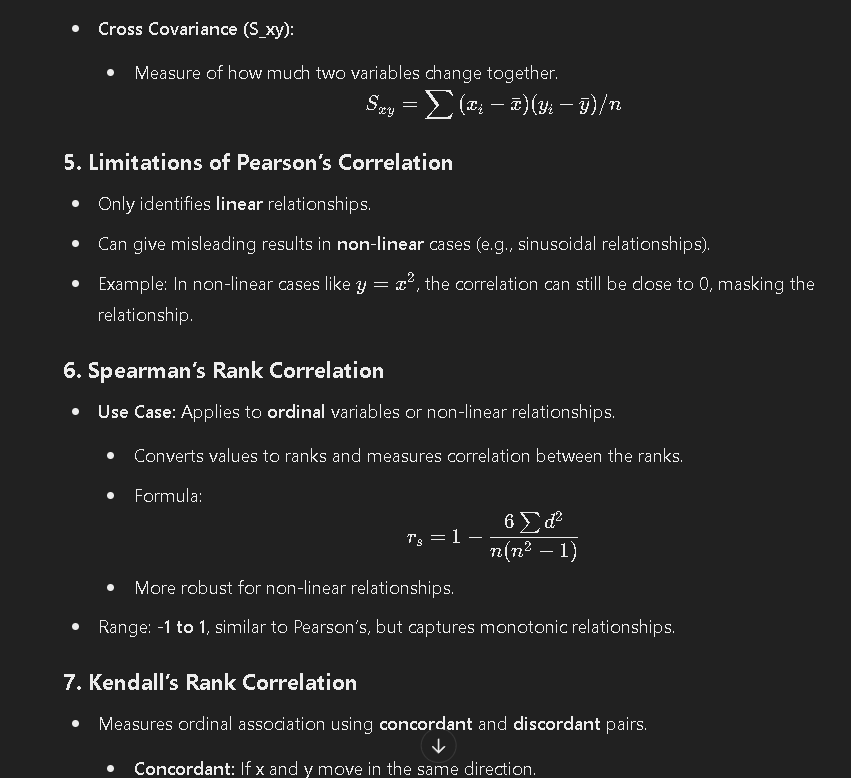
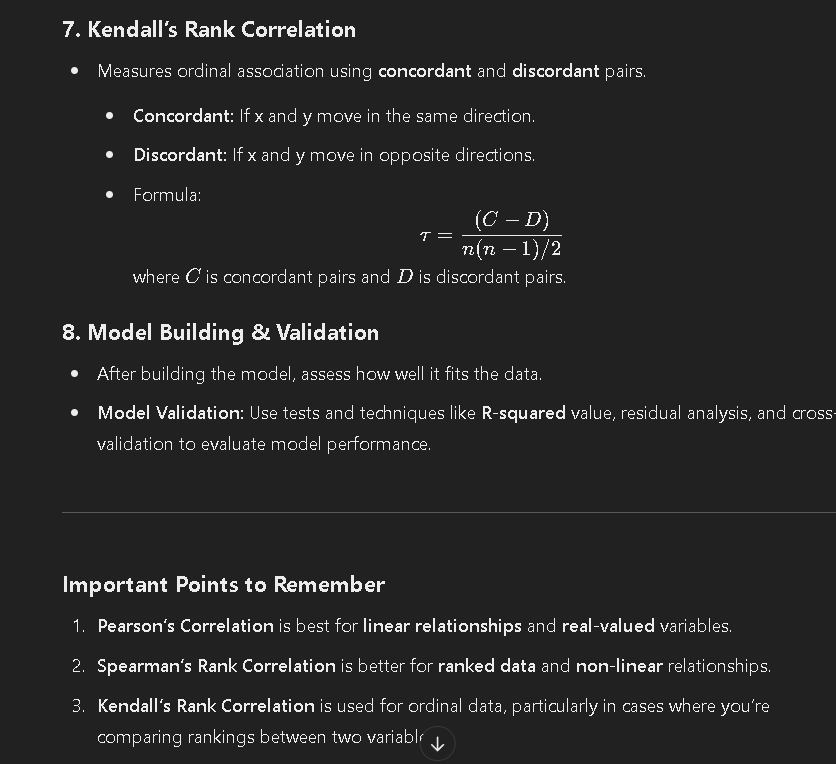
Week 6 Lecture Transcripts notes







Video 2

Pravalikka 6 transcript brief

# Meeting Summary

\*\*Date:\*\* [Insert Date]

\*\*Session:\*\* Week 6 of Data Science for Engineers Problem Solving Sessions

\*\*Instructor:\*\* Saladi Pravallika, PhD Scholar and Prime Minister Research Fellow, Indian Institute of Science

## Agenda

1. Introduction and Instructions

2. Overview of Regression

3. Ordinary Least Squares (OLS) Regression

4. Testing the Goodness of Fit

5. Hypothesis Testing

6. Practical Examples and R Code Demonstrations

## Key Points Discussed

### 1. Introduction and Instructions

- The session is being recorded.

- Breaks will be given every 30 minutes to an hour.

- Participants can request additional topics to be covered in future sessions.

- Reminder about the last day for exam registration.

### 2. Overview of Regression

- \*\*Supervised Learning:\*\* Involves predicting an output based on input data.

- \*\*Regression:\*\* Predicts continuous output values.

- \*\*Classification:\*\* Categorizes outputs into classes.

- \*\*Types of Regression:\*\*

- \*\*Univariate Regression:\*\* Involves one independent variable.

- \*\*Multivariate Regression:\*\* Involves multiple independent variables.

- \*\*Linear Regression:\*\* Relationship between variables is a straight line.

- \*\*Nonlinear Regression:\*\* Relationship is not a straight line.

### 3. Ordinary Least Squares (OLS) Regression

- \*\*Definition:\*\* A statistical method to estimate unknown parameters in a linear regression model.

- \*\*Equation:\*\* \( Y = \beta\_0 + \beta\_1X + \epsilon \)

- \( \beta\_0 \): Intercept

- \( \beta\_1 \): Slope

- \( \epsilon \): Error term

- \*\*Objective:\*\* Minimize the sum of squared errors (SSE).

- \*\*Normal Equations:\*\* Derived by taking partial derivatives of SSE with respect to \( \beta\_0 \) and \( \beta\_1 \) and setting them to zero.

### 4. Testing the Goodness of Fit

- \*\*R-Squared (R²):\*\* Measures the proportion of variance explained by the model.

- \*\*Formula:\*\* \( R^2 = 1 - \frac{\sum (Y\_i - \hat{Y\_i})^2}{\sum (Y\_i - \bar{Y})^2} \)

- \*\*Interpretation:\*\* Higher R² indicates a better fit.

- \*\*Adjusted R-Squared:\*\* Adjusts R² for the number of predictors in the model.

- \*\*Formula:\*\* \( \text{Adjusted } R^2 = 1 - \left( \frac{(1 - R^2)(n - 1)}{n - p - 1} \right) \)

- \*\*Interpretation:\*\* Higher adjusted R² indicates a better fit, especially when adding more variables.

### 5. Hypothesis Testing

- \*\*Objective:\*\* Test the significance of regression coefficients.

- \*\*Null Hypothesis (H₀):\*\* \( \beta\_1 = 0 \) (No relationship between X and Y)

- \*\*Alternative Hypothesis (H₁):\*\* \( \beta\_1 \neq 0 \) (There is a relationship)

- \*\*T-Test:\*\* Used to determine if the null hypothesis can be rejected.

### 6. Practical Examples and R Code Demonstrations

- \*\*Univariate Linear Regression:\*\*

- Example data: X = [1, 2, 3, 4, 5, 6], Y = [3, 5, 7, 9, 11, 13]

- Calculated mean, deviations, covariance, and variance.

- Derived coefficients \( \beta\_0 \) and \( \beta\_1 \).

- Plotted the regression line.

- \*\*Multivariate Linear Regression:\*\*

- Example data with two independent variables X1 and X2.

- Created a 3D scatter plot and regression plane.

- \*\*Outlier Detection:\*\*

- Demonstrated how outliers affect the regression line.

- Showed how to remove outliers and fit a new regression line.

- \*\*Residual Plots:\*\*

- Explained how to interpret residual plots to validate the model.

### Additional Notes

- \*\*Questions and Answers:\*\*

- Addressed various questions from participants regarding regression concepts, calculations, and interpretations.

- \*\*Future Topics:\*\*

- Degrees of freedom and confidence intervals will be covered in upcoming sessions.

## Conclusion

The session provided a comprehensive overview of regression analysis, focusing on OLS regression, goodness of fit, and hypothesis testing. Practical examples and R code demonstrations helped in understanding the application of these concepts. Participants were encouraged to practice the problems discussed and reach out with any questions.