UNIT 16
ANALYSIS OF VARIANCE (ANOVA)

1. INTRODUCTION

Analysis of Variance (ANOVA) is a statistical method for testing the equality of three or more population means by partitioning the total variation in a dependent variable into components attributable to different sources of variation.

ANOVA is essentially an advanced -test. If you have just two groups (like Male vs. Female), you use a t-test. If you have three or more groups (like Promotion A, Promotion B, and No Promotion), you must use ANOVA to see if any of those groups have a significantly different average outcome.

- **Null Hypothesis (**H0**):** All population means are equal ($\mu 1=\mu 2=\mu 3$).
- **Alternative Hypothesis (**H1**):** At least one pair of means is significantly different.

2. UNIVARIATE ANOVA (ONE-WAY ANOVA)

Univariate ANOVA (or One-Way ANOVA) involves a single categorical independent variable (called the factor) with three or more levels (groups) and a single continuous dependent variable. It's the simplest form of ANOVA, testing the effect of a single type of experimental treatment.

- **Coffee Shop Example:** We want to test the effectiveness of three different promotional offers on customer spending.
    - **Factor (Independent Variable):** Type of Promotion (3 levels: 1. 10% off, 2. Free Dessert, 3. Buy-One-Get-One).
    - **Dependent Variable:** Average dollars spent per visit.
- **Goal:** Determine if the **average spend** across the three promotion groups is significantly different. If the ANOVA is significant, a **Post-Hoc Test** (like Tukey's HSD) is needed to find *which specific pairs* differ (e.g., 10% off vs. Free Dessert).

3. MULTIVARIATE ANOVA (MANOVA)

Multivariate ANOVA (MANOVA) is an extension of ANOVA used when there are multiple continuous dependent variables to be analyzed simultaneously. MANOVA tests if the treatment groups differ significantly on the multiple dependent variables *collectively*.

If you want to see if your promotions worked on both spending *and* satisfaction at the same time, you use MANOVA.

- **Coffee Shop Example:** We test the same three promotions (the Factor).
    - **Dependent Variables (Multiple):** 1. Average dollars spent per visit, **AND** 2. Customer Satisfaction Score (1-7 scale).
- **Benefit:** MANOVA is more powerful than running two separate ANOVAs because it accounts for the correlation between the two dependent variables (e.g., spending and satisfaction usually increase together).

4. ANOVA FOR RANDOMIZED BLOCKS

This design is a specialized form of ANOVA used when the researcher wants to statistically control for the effects of an extraneous variable (a known covariate) that could otherwise distort the results of the main experimental treatment. The extraneous factor is used to create homogeneous "blocks" within the sample.

If you know a factor, like age, strongly affects the outcome, you don't want it to hide the true effect of your treatment. You divide your participants into blocks (e.g., "Young," "Middle-Aged," "Senior") and then apply all treatments within each block. This isolates the treatment effect from the age effect.

- **Coffee Shop Example:** We test the three Promotions (the Factor). We suspect **Customer Loyalty Tier** (Bronze, Silver, Gold) affects spend.
    - **Blocks:** We create blocks based on Loyalty Tier.
    - **Design:** We ensure all three promotions (10% off, Free Dessert, BOGO) are tested equally within the Gold, Silver, and Bronze blocks. This allows us to statistically isolate the true impact of the **Promotion** itself, separate from how much loyal customers naturally spend.

5. ANOVA FOR LATIN SQUARE DESIGN

The Latin Square Design is an experimental design used to control for two non-interacting extraneous variables (blocks) in addition to the primary experimental factor. It requires that the number of rows, columns, and treatment levels are equal ($n \times n$).

This is used in complex experiments when you have *two* outside factors that could mess up the results, and you need to control for both efficiently without needing a massive sample size.

- **Coffee Shop Example:** We test three **Promotions (**A,B,C**)**. We want to control for two extraneous factors: **Store Location (1, 2, 3)** and **Day of Week (Mon, Tue, Wed)**.
    - The design ensures that each promotion appears exactly once in each store and exactly once on each day. This efficiently separates the effect of the Promotion from the effects of Location and Day.

6. ANOVA WITH INTERACTION (FACTORIAL DESIGN)

In a Factorial Design, two or more factors (independent variables) are analyzed simultaneously. Interaction occurs when the effect of one factor (e.g., Promotion A) on the dependent variable is *dependent* on the level of a second factor (e.g., Store Design).

We don't just ask, "*Does Promotion A work?*" We ask, "*Does Promotion A work BETTER than others, but ONLY if the store has the new modern design?*" When the effect of one variable changes based on the level of another, you have an interaction.

- **Coffee Shop Example:**
    - **Factor 1:** Promotion (Levels: 10% off vs. Free Dessert).
    - **Factor 2:** Store Design (Levels: Classic vs. Modern).

- **Result:** We might find the 10% off promotion works best in the **Modern Design** stores, while the Free Dessert works best in the **Classic Design** stores. This is a significant interaction.

7. FACTORIAL DESIGN

A Factorial Design is an experimental design that includes two or more factors in a single experiment, crossing all levels of the factors to measure the main effect of each factor and the interaction effects between them (e.g., a 2×3 design involves one factor with 2 levels and a second factor with 3 levels).

It's the experimental structure required to test for those interactions mentioned above. It tests every possible combination of treatments.

- **Coffee Shop Example:** A 2×2 factorial design:
    - **Factor A:** New Menu Item (Yes/No)
    - **Factor B:** New Staff Uniforms (Yes/No)
    - **Four Treatments Tested:** (Menu ON / Uniforms ON), (Menu ON / Uniforms OFF), (Menu OFF / Uniforms ON), (Menu OFF / Uniforms OFF).

8. ANALYSIS OF COVARIANCE (ANCOVA)

Analysis of Covariance (ANCOVA) is an extension of ANOVA that includes at least one covariate (a continuous extraneous variable) in the analysis. ANCOVA statistically adjusts the means of the dependent variable to account for the effect of the covariate, thereby increasing the power of the test.

ANCOVA is similar to the Randomized Blocks design, but instead of using a categorical variable (like Loyalty Tier) for blocking, it uses a continuous numerical variable (like Pre-Test Spend). It cleans up the dependent variable by removing the noise caused by this continuous factor.

- **Coffee Shop Example:** We test the effect of our Promotions (the Factor). We want to control for how much the customer spent *before* the experiment began.
    - **Factor:** Promotion (10% off, Free Dessert, BOGO).
    - **Covariate:** Baseline average weekly spend (continuous, interval data).
- **Benefit:** ANCOVA adjusts the final spend results based on the customers' natural spending habits, giving us a purer measure of the promotion's true effect.