**Design and Analysis Using ANCOVA: Combining Categorical and Continuous Predictors**

**Concept:**

ANCOVA (Analysis of Covariance) is a blend of **ANOVA (Analysis of Variance)** and **linear regression**. It allows us to:

* Assess the **effect of a categorical variable** (like treatment groups) on a response.
* Adjust for the **influence of a continuous covariate** (like age, baseline value, etc.).
* Increase statistical power by reducing error variance.

**When to Use ANCOVA?**

* You have a **categorical independent variable** (e.g., treatment group).
* You have a **continuous covariate** (e.g., baseline measure).
* You want to test for **differences in the outcome** across categories **while controlling for** the continuous variable.

**Example Scenario:**

Suppose you're testing the effect of **3 different diets (Diet A, B, C)** on **weight loss**, and you want to control for **initial weight** (continuous variable).

R codes in R file

**Assumptions of ANCOVA**

1. **Linearity** between covariate and dependent variable.
2. **Homogeneity of regression slopes** (no interaction between group and covariate).
3. **Independence**, **normality**, and **homoscedasticity** of residuals.
4. Plus, one additional assumption: the range of the continuous predictor (the covariate; i.e., the values along the x-axis) must be similar for all levels of the categorical predictor (i.e. the lines should overlap along the length of the x-axis).

Example

You're testing how **14 different excipients** affect the **stability of Trastuzumab** after 4 weeks, measured as **percent monomer retention** (response variable). You also collected the **initial protein concentration** (continuous covariate) at time zero, which may influence stability.

You want to see:

* Whether **excipients (categorical)** significantly influence stability,
* After **adjusting for initial protein concentration** (continuous).

NOTES:

###a statistical interaction: where the effect of one predictor variable on the response variable depends on the level of the other predictor variable.

Case 1: regression slopes are parallel = no interaction between categorical and continuous predictor: OK to interpret ‘main’ effect (slide 32-35)

Case 2: regression slopes are notparallel = interaction between categorical and continuous predictor: NOT OK to interpret ‘main’ effect (slide 32-35)

**What Happens When There’s an Interaction?**

In ANCOVA, if you model:

lm(Y ~ A \* X)

Where:

* A = categorical predictor (e.g., excipient)
* X = continuous covariate (e.g., initial concentration)
* Y = response variable (e.g., % monomer retention)

You're telling the model:

"Allow the **relationship between X and Y** (the slope) to vary **depending on the level of A**."

This means:

“The effect of X **depends on which group** (level of A) you're in.”

**Why You *Can’t* Interpret Main Effects Directly**

If the interaction A:X is **significant**, then:

* The effect of X is **not constant across groups**.
* The difference between groups (A) **depends on the value of X**.

So, interpreting the **main effect** of A or X in isolation is **misleading**, because:

* The model includes **different slopes** for each level of A.

**Example**

lm(Y ~ A \* X)

# This is equivalent to:

lm(Y ~ A + X + A:X)

* A is the group effect (e.g., excipient)
* X is the covariate (e.g., initial concentration)
* A:X is the interaction

If the interaction is **significant**, your interpretation must shift to:

“In group A1, the relationship between X and Y is this.  
In group A2, it's different.”

**What Does It Mean in ANCOVA?**

ANCOVA assumes **parallel slopes** (no interaction by default).

But if interaction exists:

* Each group (e.g., excipient) has its **own regression line**, with **different slopes**.
* The **covariate adjustment is no longer uniform** across groups.

So, **standard ANCOVA is not appropriate** in this case.  
Instead, you:

1. **Include the interaction term** in the model.
2. **Visualize or summarize** the regression lines **per group**.
3. Avoid statements like “Excipient B leads to higher stability” unless you specify **at what concentration**.