B.Sc (Statistics) Sem. IV CC 409 (Linear Model and Regression Analysis)

# Topic- Analysis of Covariance

**Dr. Saurabh**
**Asstt. Professor**
**Dept. of Statistics**
**(email- shaurabh.bhu@gmail.com**

# ANCOVA

Analysis of covariance (**ANCOVA**) is a general linear model which blends ANOVA and regression.

# ANCOVA

- ANCOVA by definition is a general linear model that includes both ANOVA (categorical) predictors and Regression (continuous) predictors.
- ANCOVA examines the influence of an independent variable on a dependent variable while removing the effect of the covariate factor.
- ANCOVA first conducts a regression of the independent variable (i.e., the covariate) on the dependent variable.
- The residuals (the unexplained variance in the regression model) are then subject to an ANOVA.

## ANCOVA in other words…

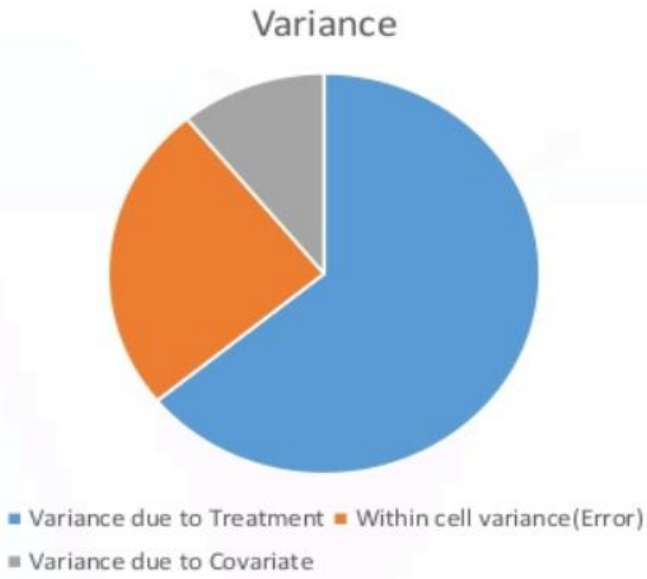
- Analysis of Covariance (ANCOVA) is a statistical test related to ANOVA
- It tests whether there is a significant difference between groups after controlling for variance explained by a covariate
- A covariate (CV) is a continuous variable that correlates with the dependent variable (DV)
- This is one way that you can run a statistical test with both categorical and continuous independent variables

## Purposes of ANCOVA

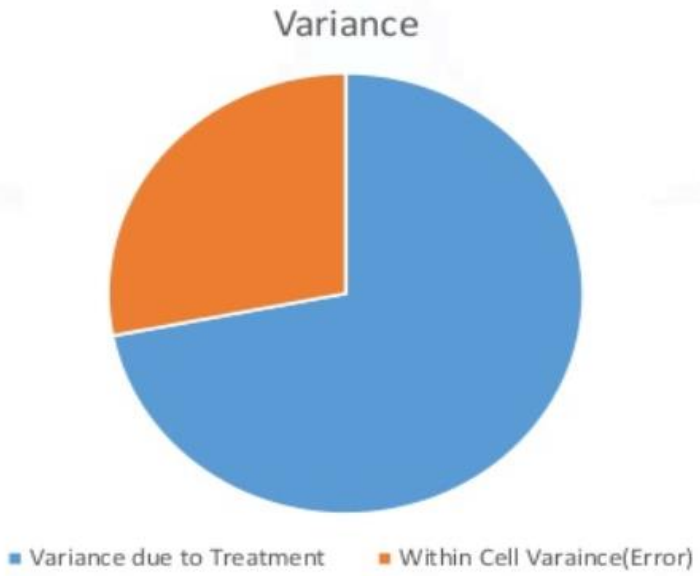
- Increase sensitivity of F test
  - Removes predictable variance from the error term
  - Improves power of the analysis

# Adjustment of Covariate Effect

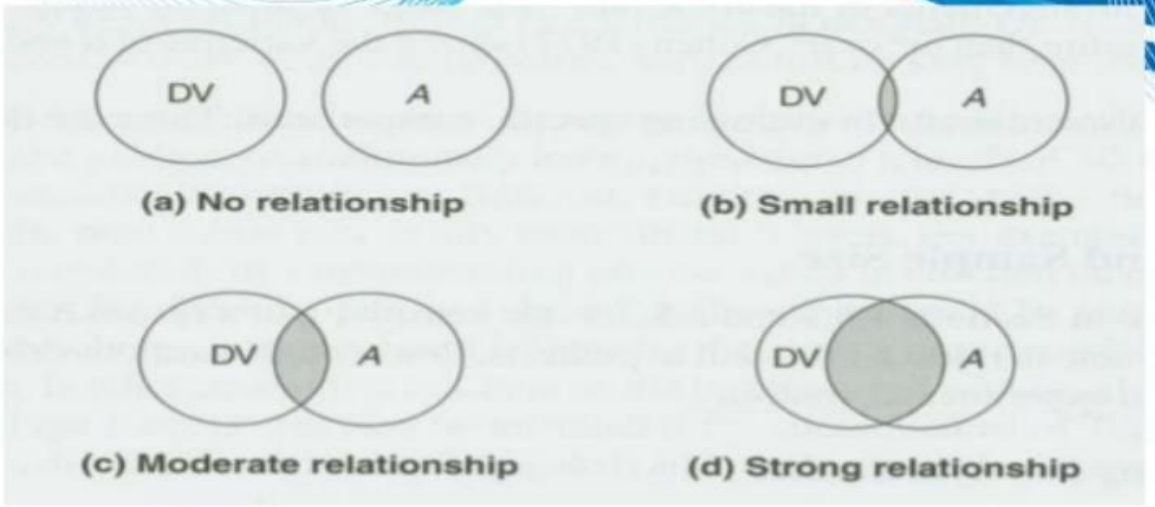
**Partitioning variance in ANOVA**




**Partitioning variance in ANCOVA**



## Relationship between CV and DV




## **Hypotheses** for **ANCOVA**

$H_0$: the group means are equal after controlling for the covariate

Vs

$H_1$: the group means are not equal after controlling for the covariate

# Assumptions

- **linearity of regression:** The regression relationship between the dependent variable and concomitant variables must be linear.
- **homogeneity of error variances:** The error is a random variable zero mean and equal variances for different treatment classes and observations
- **independence of error terms:** The errors are uncorrelated. That is that the error covariance matrix is diagonal.
- **normality of error terms:** The residual (error terms) should be normally distributed $\varepsilon_{ij} \sim N(0, \sigma^2)$
- **homogeneity of regression slopes:** The slopes of the different regression lines should be equivalent, i.e., regression lines should be parallel among groups.

## Choosing Covariates

- Variables that affect or have the potential to affect the dependent variable
  - Demographic information
  - Inherent characteristics
  - Differences in group characteristics due to sampling
- Number of covariates depends on:
  - Known relationship or previous research
  - No. of independent variables or groups
  - Total no. of subjects

## Example

Suppose we want to compare the effect of drugs on the weights of a particular group of patients(homogeneous among themselves). We can analyze the data by performing the ANCOVA by regarding:

$y$: the final weight of the patients taking drugs, after a specified period as the response variable.

$x$: the initial weight of the animals at the time of starting the experiment as the covariate.

# Model

So, our model becomes:

$$y_{ij} = \mu + \alpha_i + \beta\left(x_{ij} - \overline{x_{00}}\right) + \varepsilon_{ij}$$ Where,

- $\mu$ is the general mean effect
- $\alpha_i$ is the (fixed) additional effect due to the $i$th treatment (i=1,2,….,p)
- $\varepsilon_{ij}$ is the random error effect (j=1,2,…, $n_i$)
- $\beta$ is the coefficient of regression of y on x
- $x_{ij}$ is the value of covariate variable corresponding to the response variable $y_{ij}$