Part 5
Lecture 1 More on Modeling

1

Who I am...

Pascal Tyrrell, PhD

Associate Professor

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine

Department of Statistical Sciences, Faculty of Arts and Science

Overview of Statistical Models

	Type of Predictors Categorical	Continuous	Continuous and Categorical
Type of Response **Continuous**	**Analysis of Variance (ANOVA)**	Ordinary Least Squares (OLS) Regression	Analysis of Covariance (ANCOVA)
Categorical	Contingency Table Analysis or Logistic Regression	Logistic Regression	Logistic Regression

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Overview

- Are there any differences among the population means?
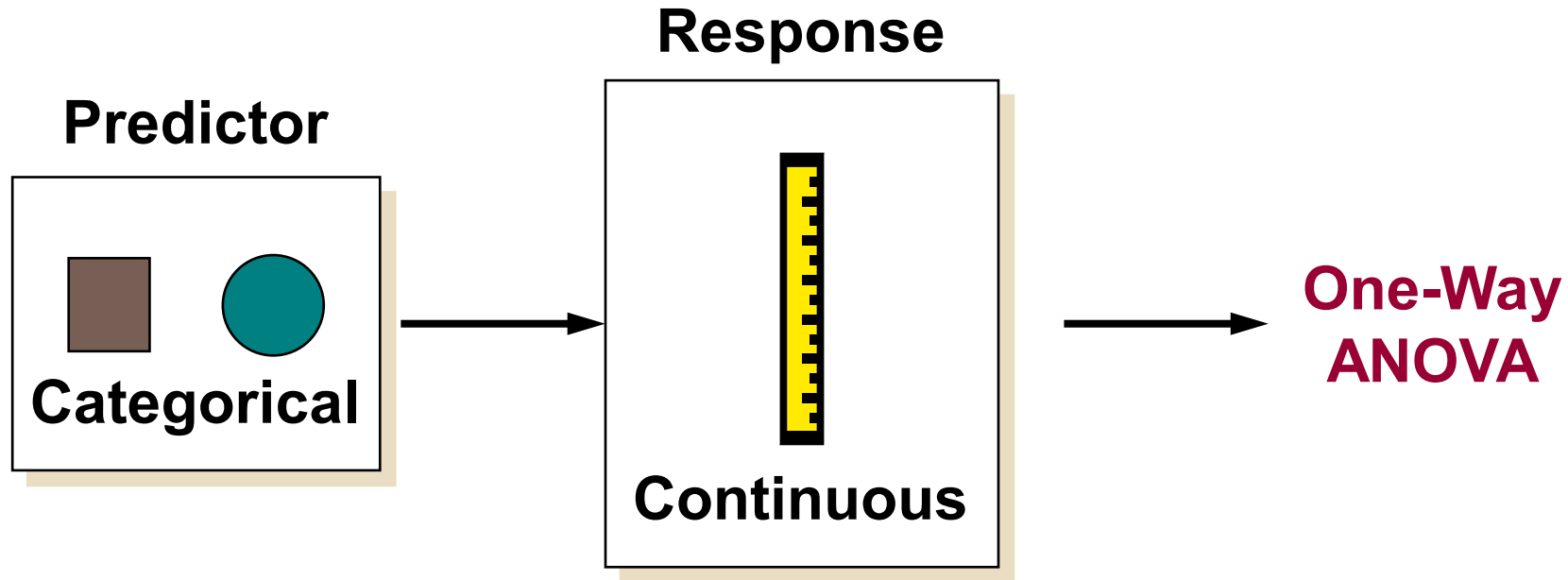
Another way of asking: Does information about group membership help predict the level of a numeric response?

Research Questions for One-Way ANOVA

- Do accountants, on average, earn more than teachers? *

- * Is this a case for a *t* test?

Research Questions for One-Way ANOVA

- Do people treated with one of two new drugs have higher average T-cell counts than people in the control group?

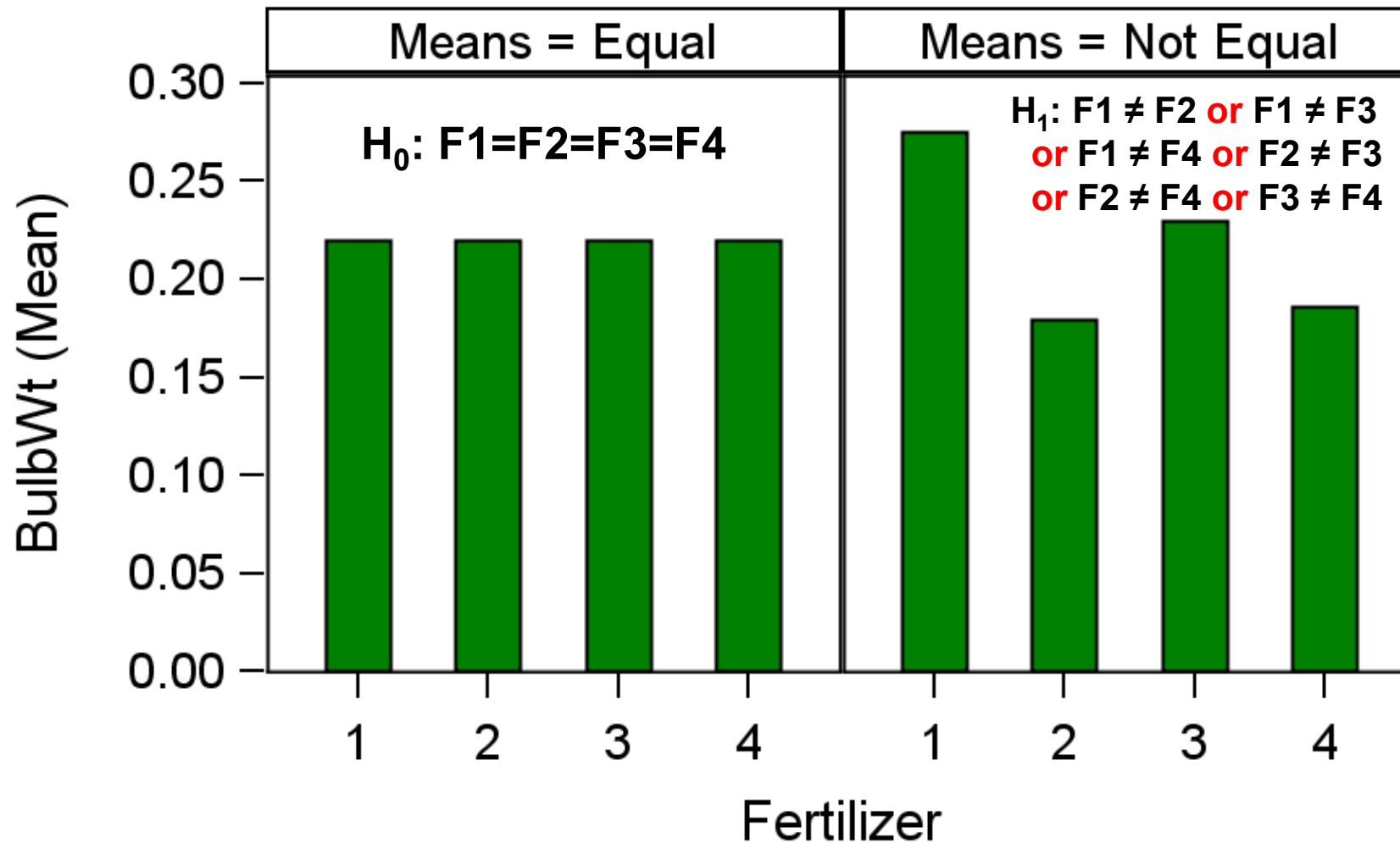Placebo

Treatment 1

Treatment 2

Research Questions for One-Way ANOVA

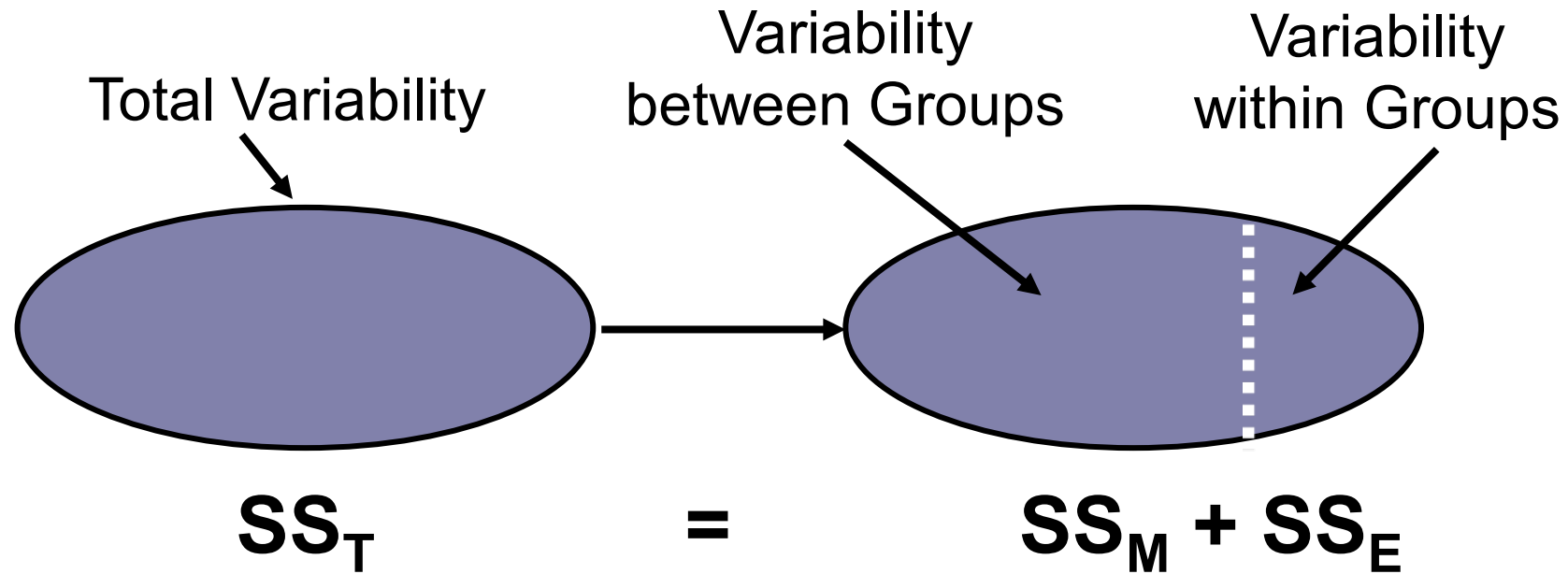- Do people spend different amounts depending on which type of credit card they have?
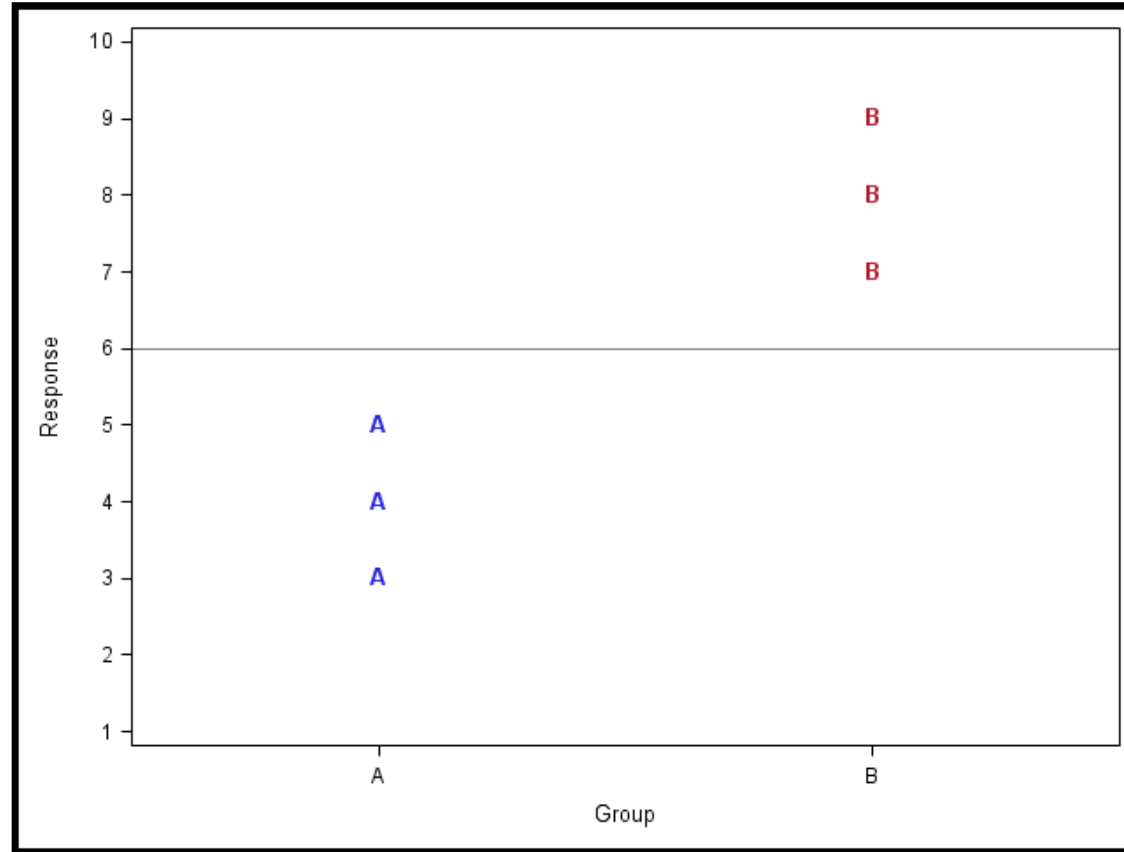
The ANOVA Hypothesis

Partitioning Variability in ANOVA

Total Variability

Variability between Groups

Variability within Groups
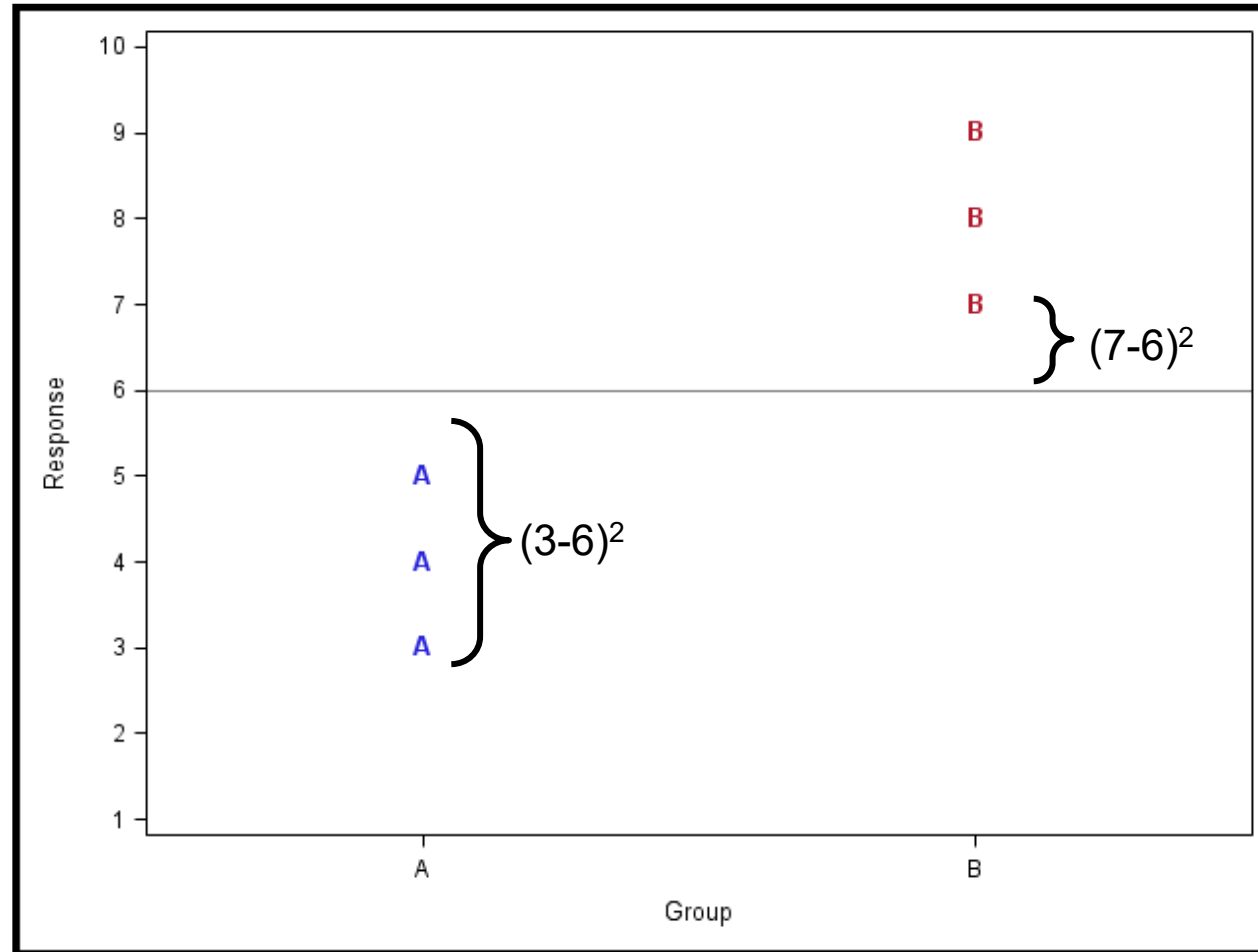
SS_T = $SS_M + SS_E$

Sums of Squares

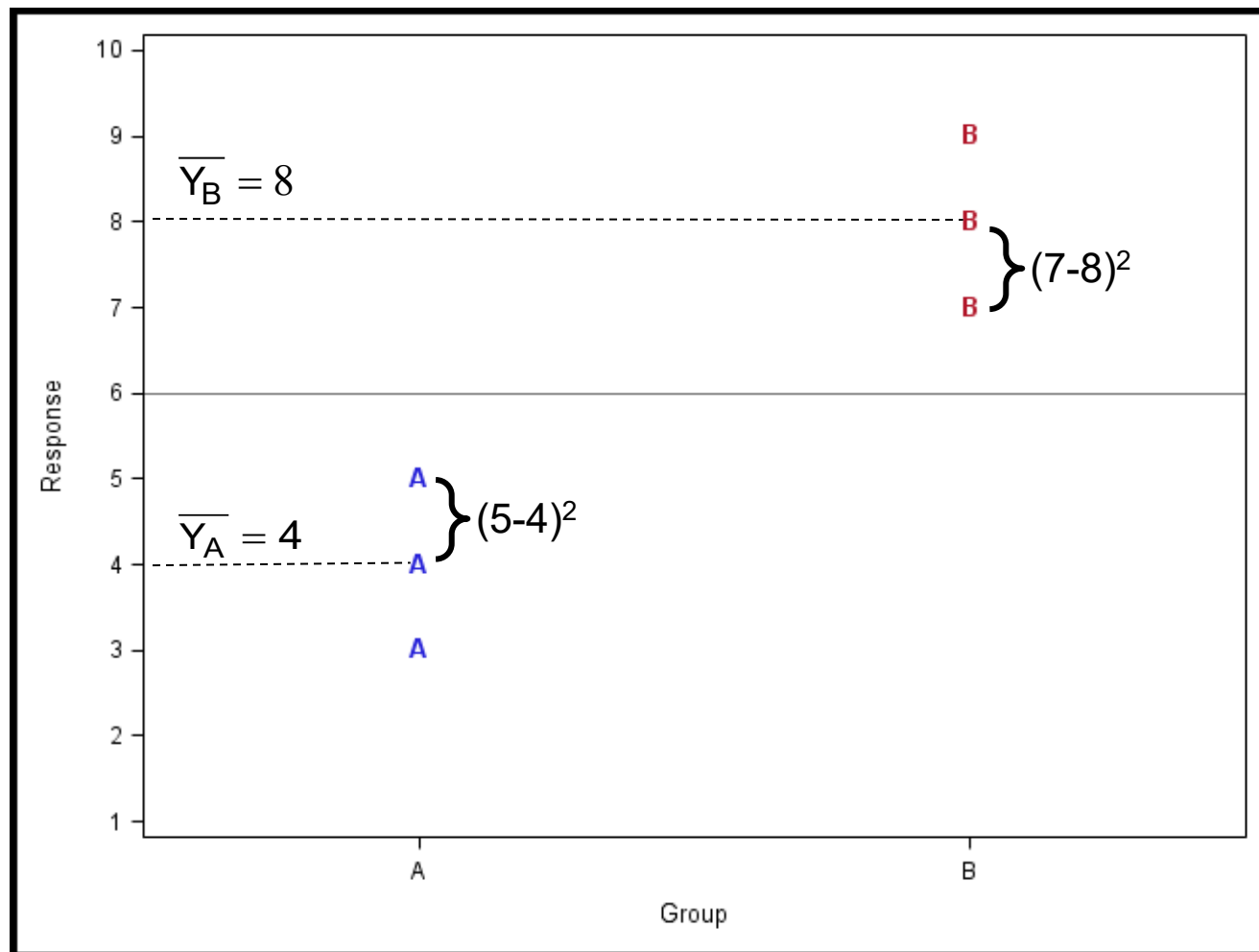$$\text{Overall mean} = \bar{\bar{y}} = \frac{3+4+5+7+8+9}{6} = \mathbf{6}$$

Total Sum of Squares

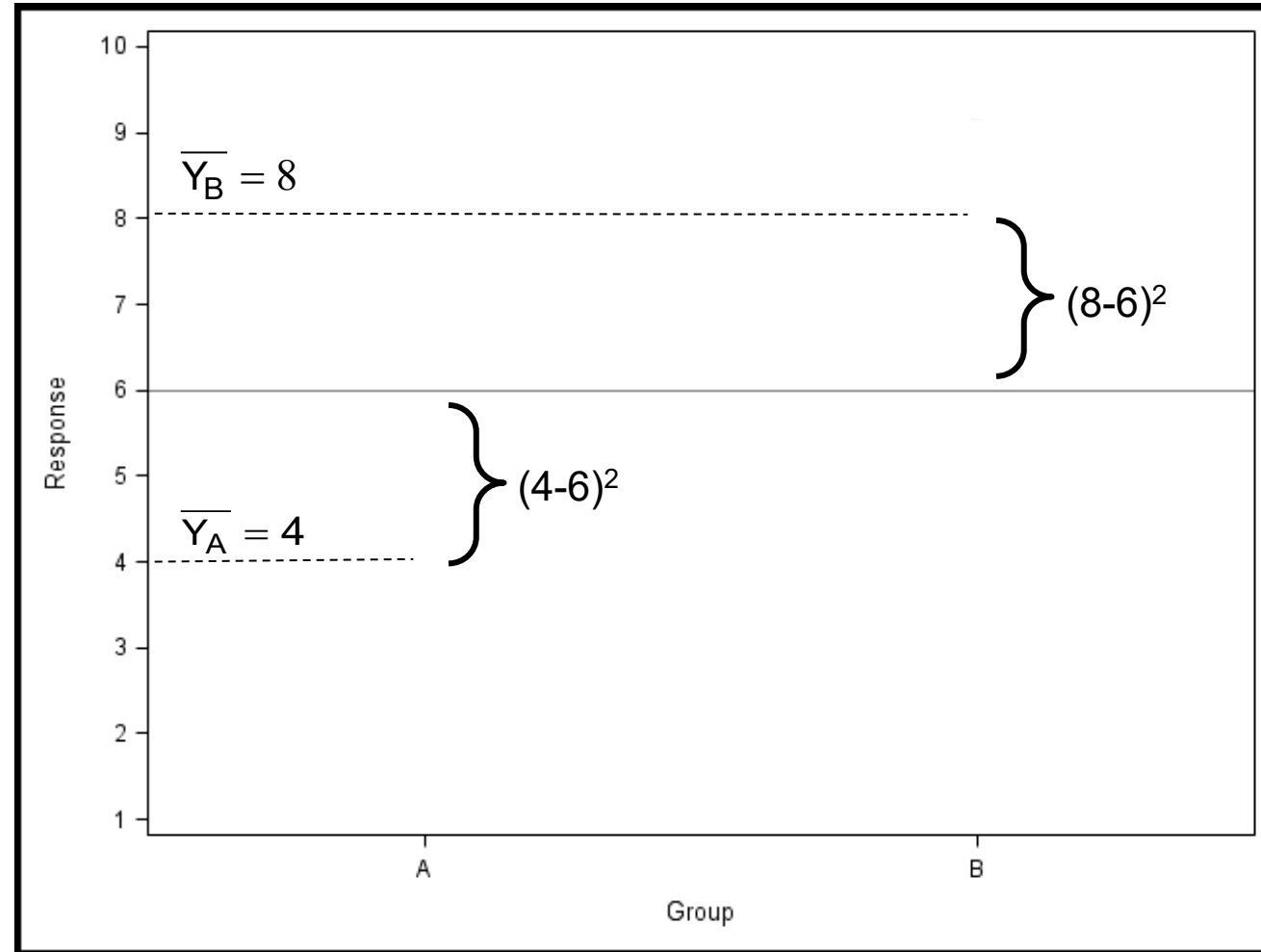$$SS_T = (3-6)^2 + (4-6)^2 + (5-6)^2 + (7-6)^2 + (8-6)^2 + (9-6)^2 = 28$$

Error Sum of Squares

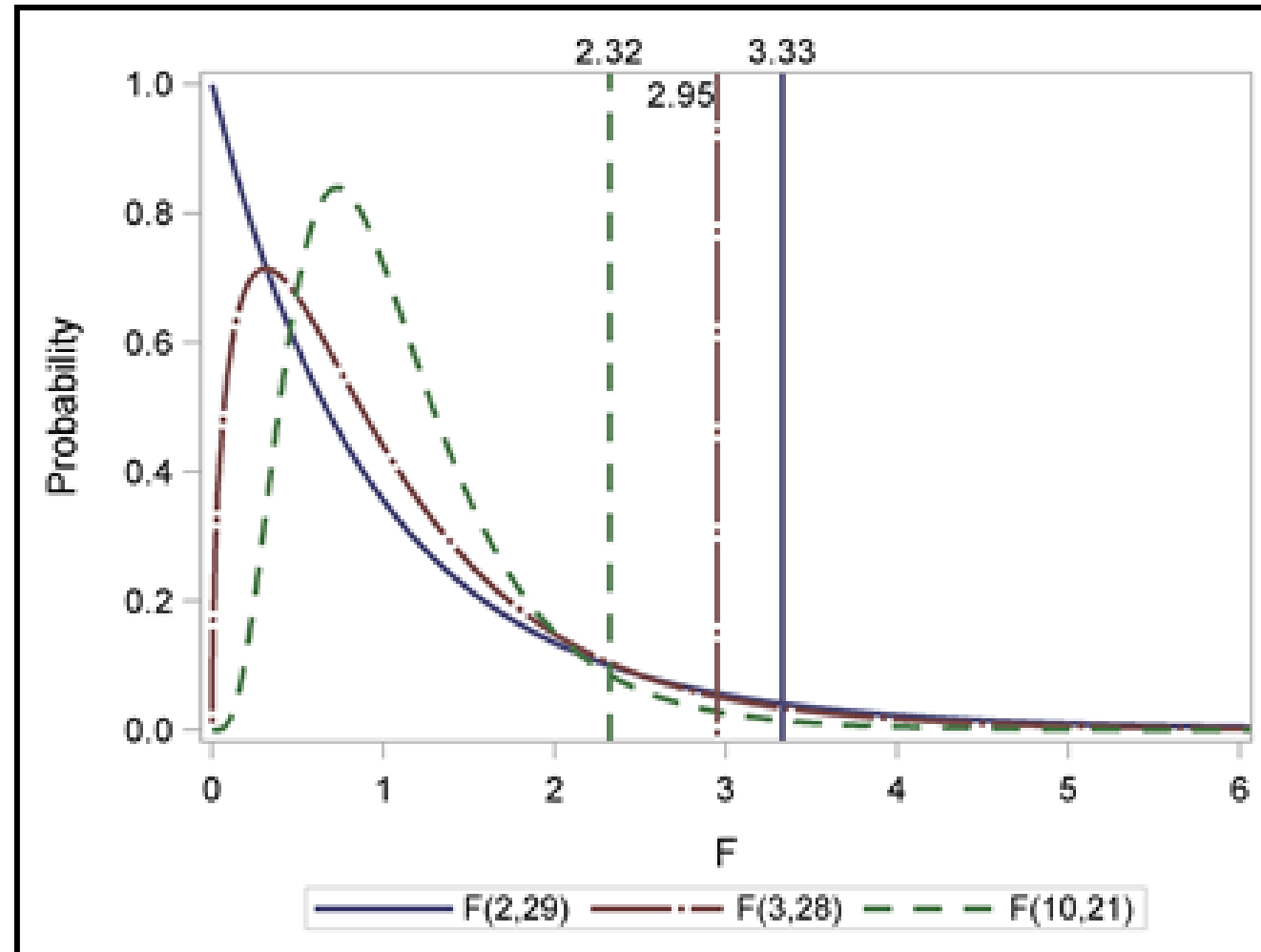$$SS_E = (3-4)^2 + (4-4)^2 + (5-4)^2 + (7-8)^2 + (8-8)^2 + (9-8)^2 = \mathbf{4}$$

Model Sum of Squares

$$SS_M = 3*(4-6)^2 + 3*(8-6)^2 = 24$$

F Statistic and Critical Values at α=0.05

$$F(\text{Model df, Error df}) = MS_M / MS_E$$

Coefficient of Determination

$$R^2 = SS_M / SS_T$$

"Proportion of variance accounted for by the model"

The ANOVA Model

$$\text{BulbWt} = \text{Base Level} + \text{Fertilizer} + \text{Unaccounted for Variation}$$

$$Y_{ik} = \mu + \tau_i + \varepsilon_{ik}$$

The GLM Procedure

- General form of the GLM procedure:

PROC GLM DATA=*SAS-data-set* PLOTS=*options*;
    CLASS *variables*;
    MODEL *dependents=independents </ options>*;
    MEANS *effects </ options>*;
    LSMEANS *effects </ options>*;
    OUTPUT OUT=*SAS-data-set keyword=variable…*;
RUN;
QUIT;

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Assumptions for ANOVA

- Observations are independent.
- Errors are normally distributed.
- All groups have equal error variances.

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Overview of Statistical Models

Type of Response ╲ Type of Predictors	Categorical	Continuous	Continuous and Categorical
Continuous	Analysis of Variance (ANOVA)	Ordinary Least Squares (OLS) Regression	Analysis of Covariance (ANCOVA)
Categorical	Contingency Table Analysis or Logistic Regression	Logistic Regression	Logistic Regression

Two Continuous Variables
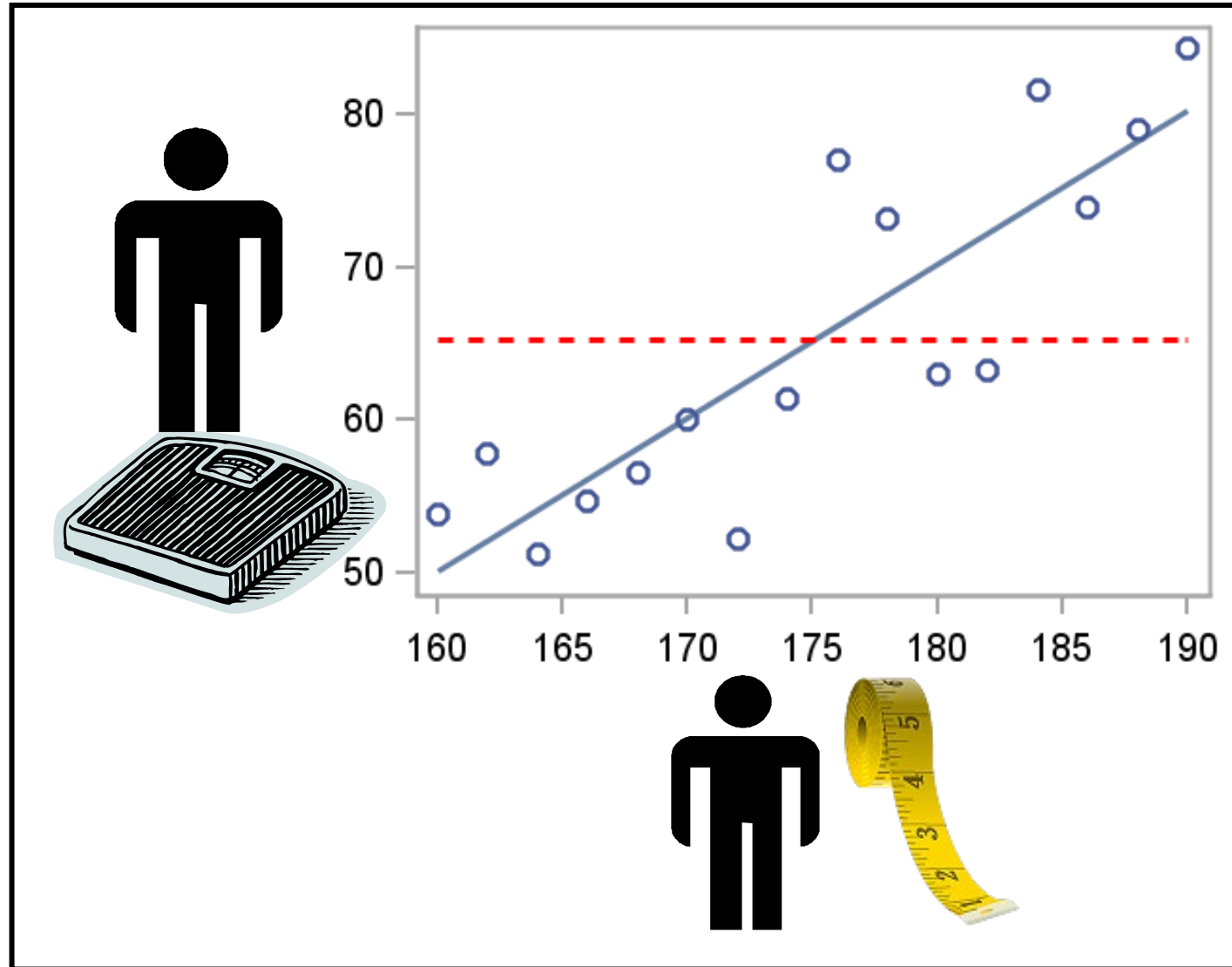
Scatter Plots
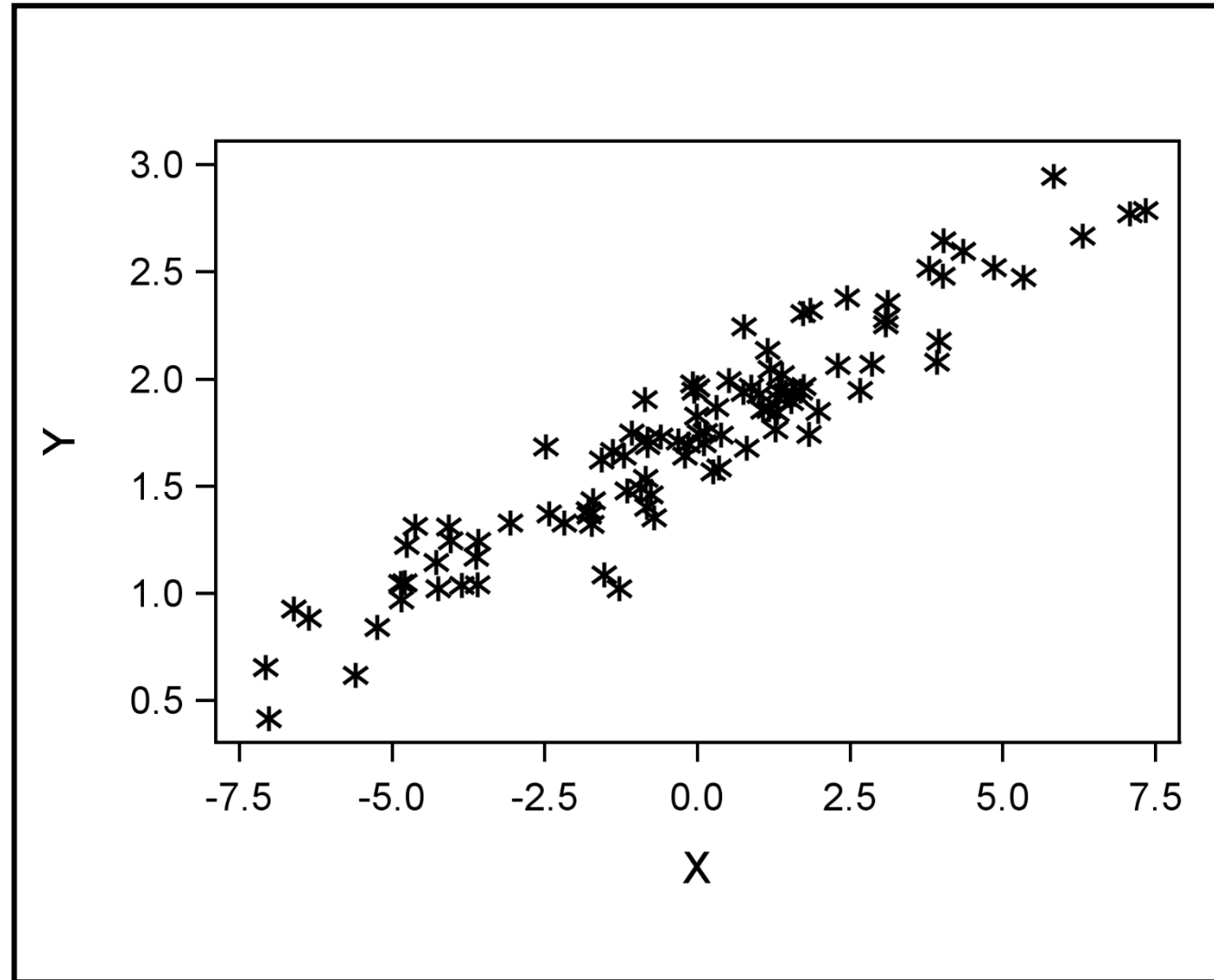
Relationships between Continuous Variables