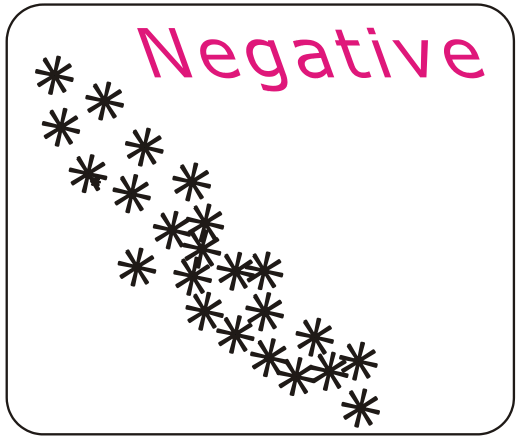
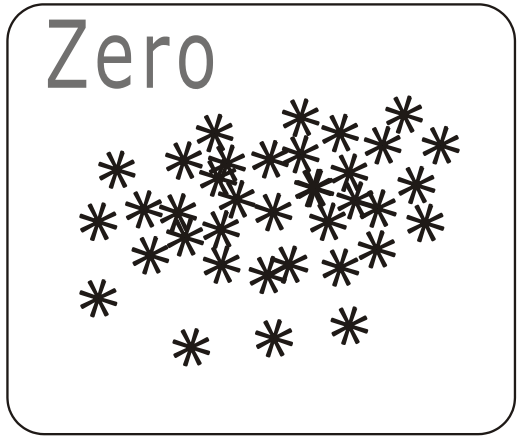
Part 5
Lecture 1 More on Modeling

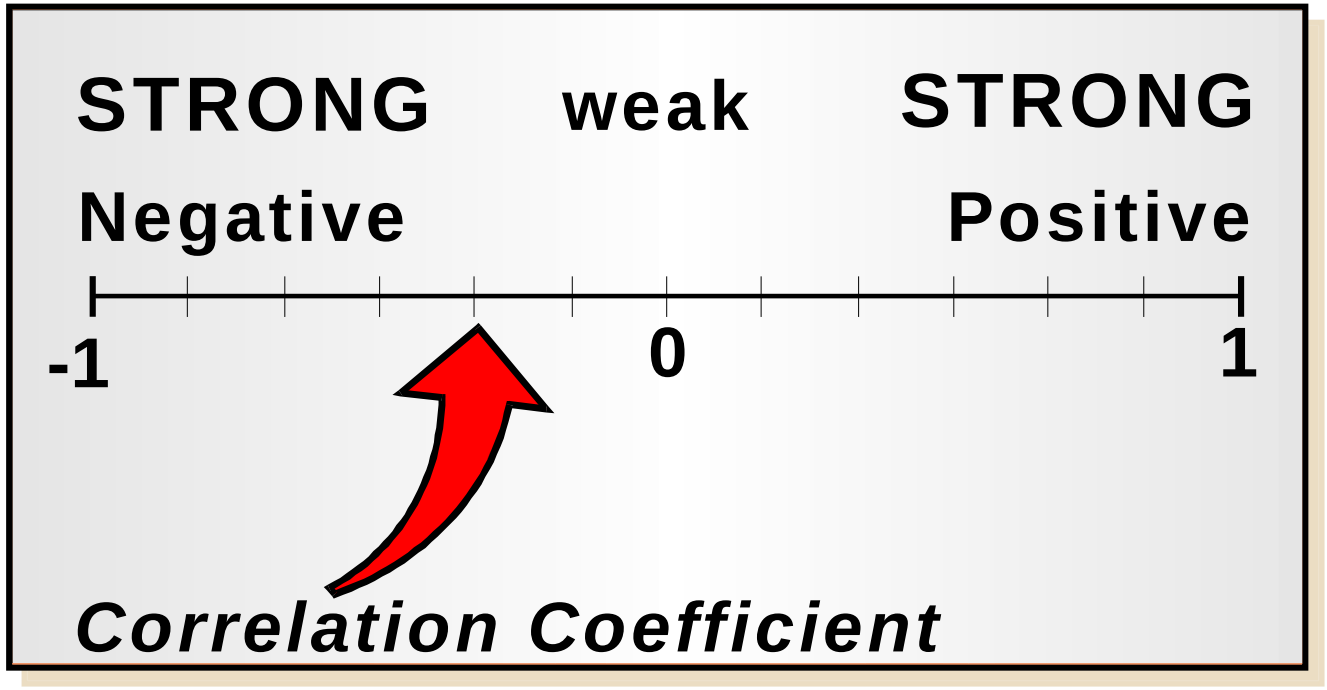
1

Correlation





Pearson Correlation Coefficient



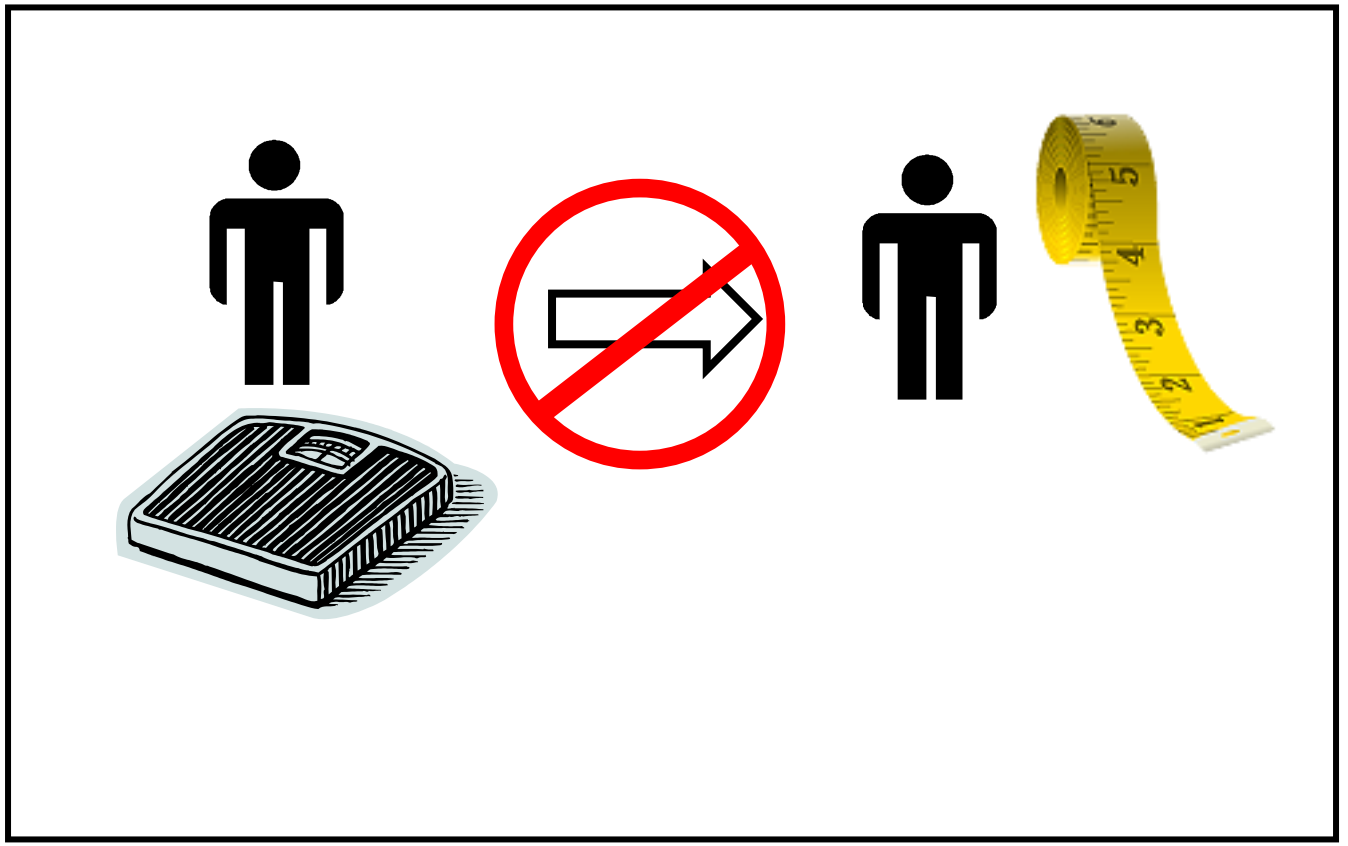
Hypothesis Test for a Correlation

- The parameter representing correlation is ρ.
  - ρ is estimated by the sample statistic *r*.

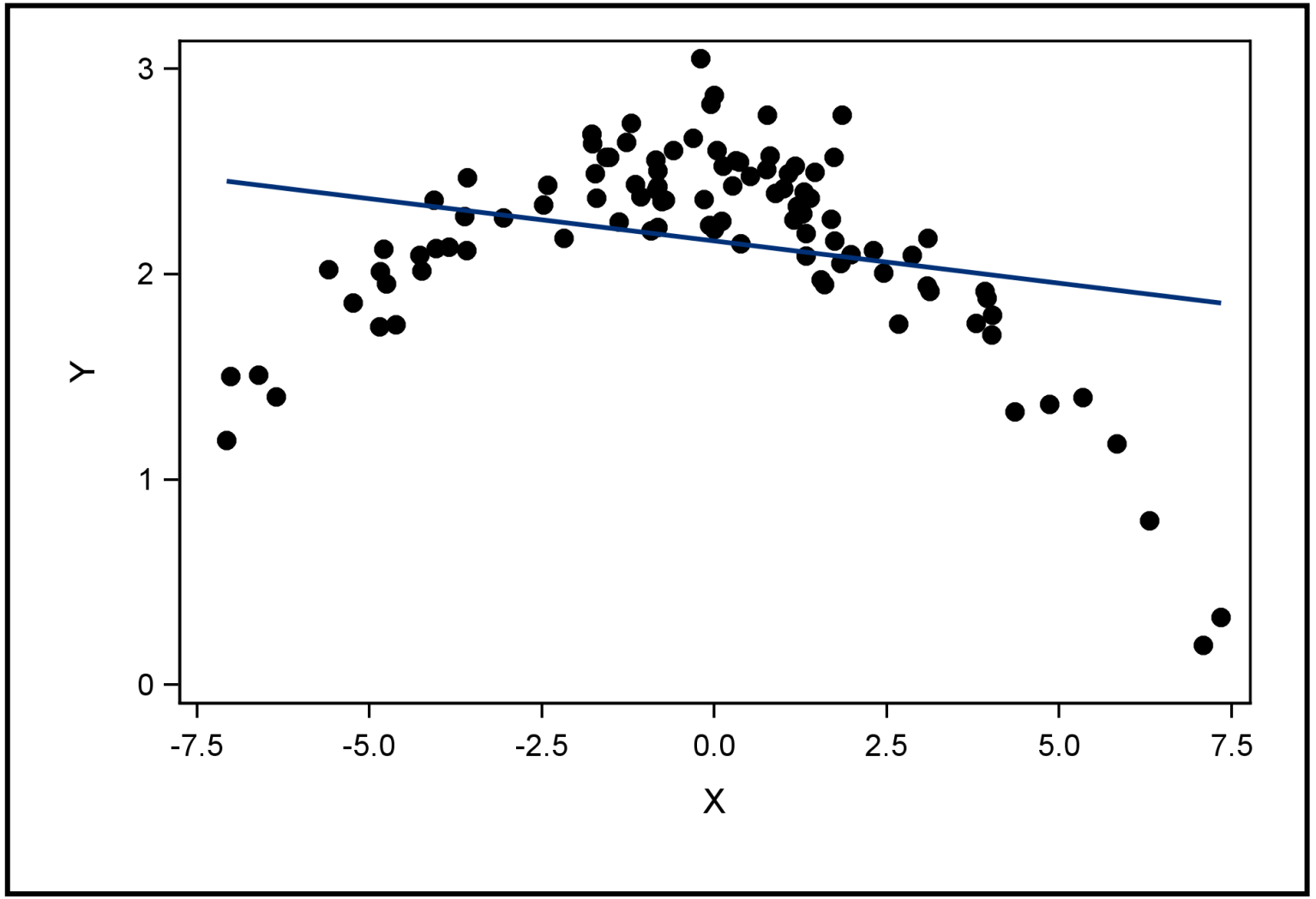
- H_0: $\rho=0$
  - Rejecting H_0 indicates only great confidence that ρ is not exactly zero.

- A *p*-value does not measure the magnitude of the association but is affected by sample size.

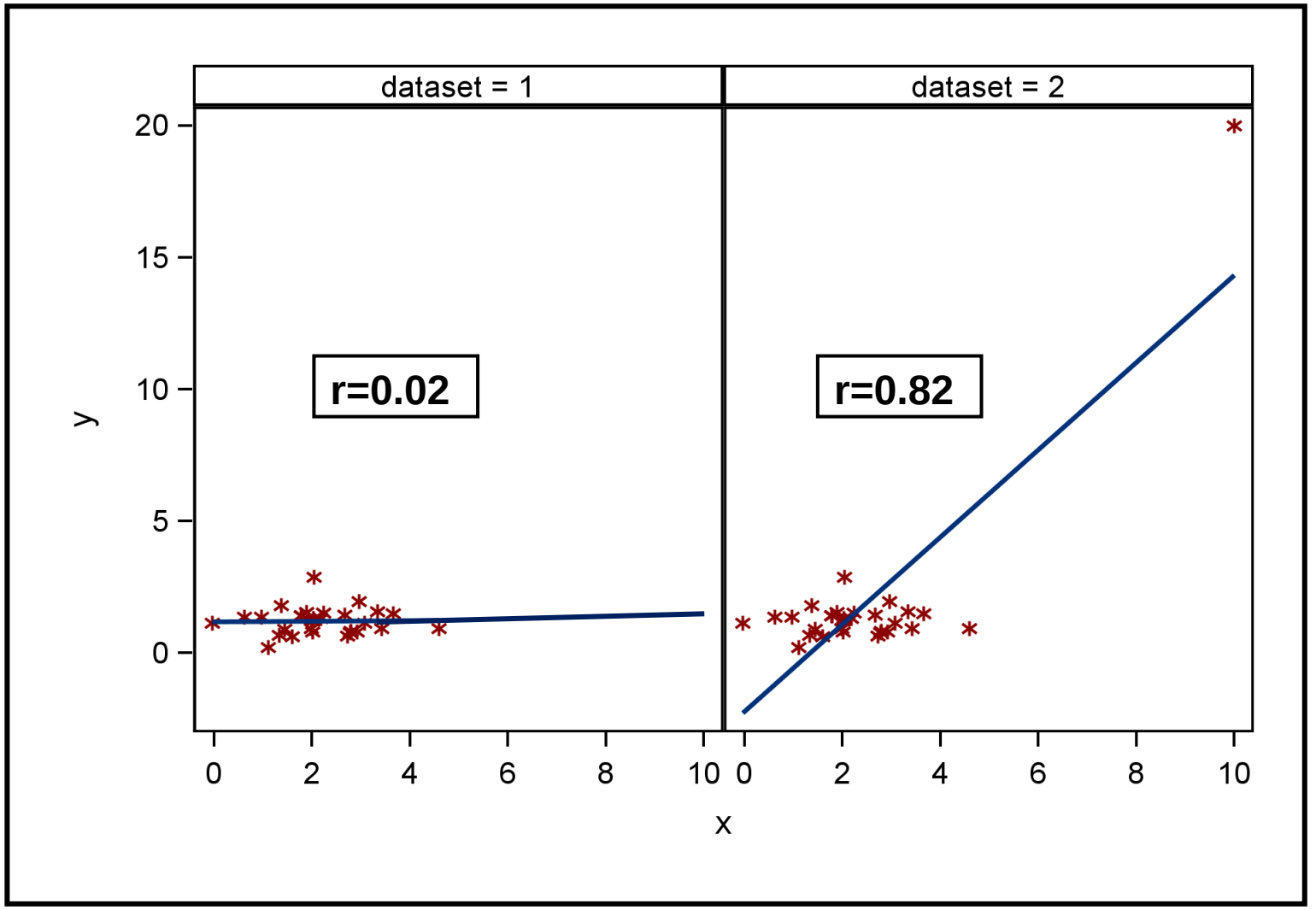
Correlation versus Causation



Missing Another Type of Relationship



Extreme Data Values



The CORR Procedure

- General form of the CORR procedure:

```
PROC CORR DATA=SAS-data-set <options>;
      VAR variables;
      WITH variables;
      ID variables;
RUN;
```

Simple Linear Regression Analysis

- The objectives of simple linear regression are as follows:
  - assess the significance of the predictor variable in explaining the variability or behavior of the response variable
  - predict the values of the response variable given the values of the predictor variable

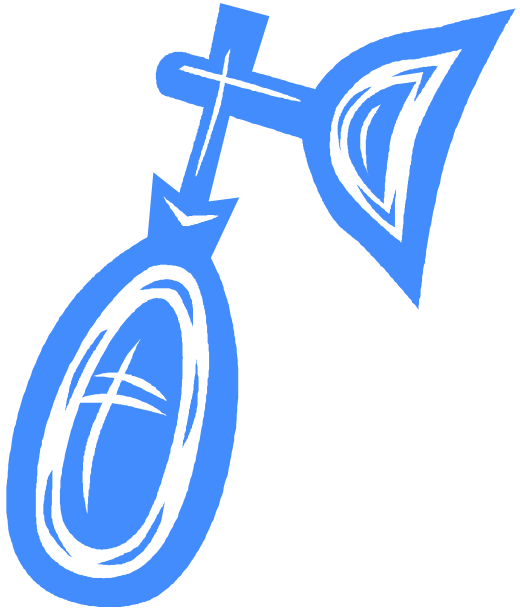
Fitness Example

PREDICTOR

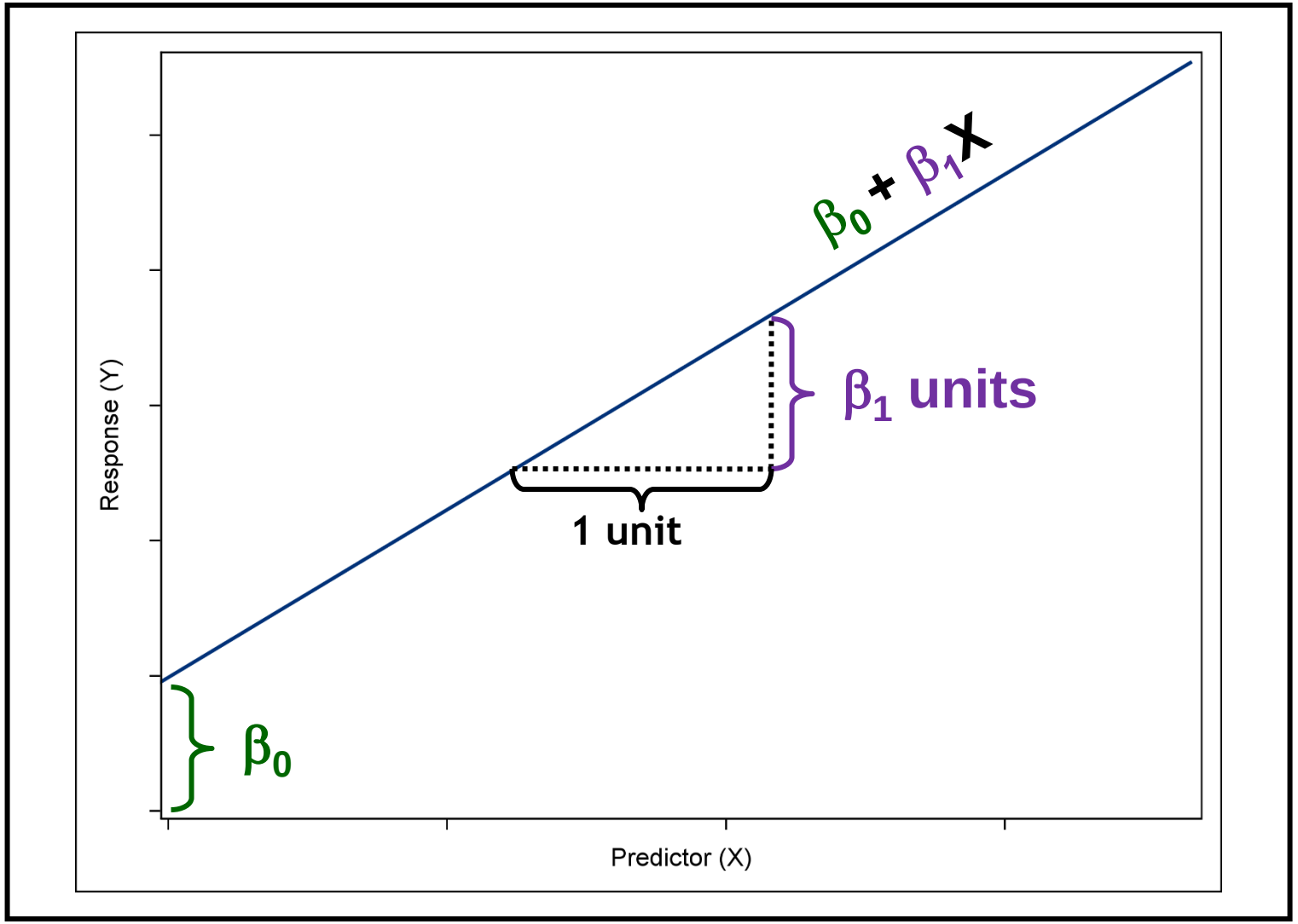
RunTime

RESPONSE

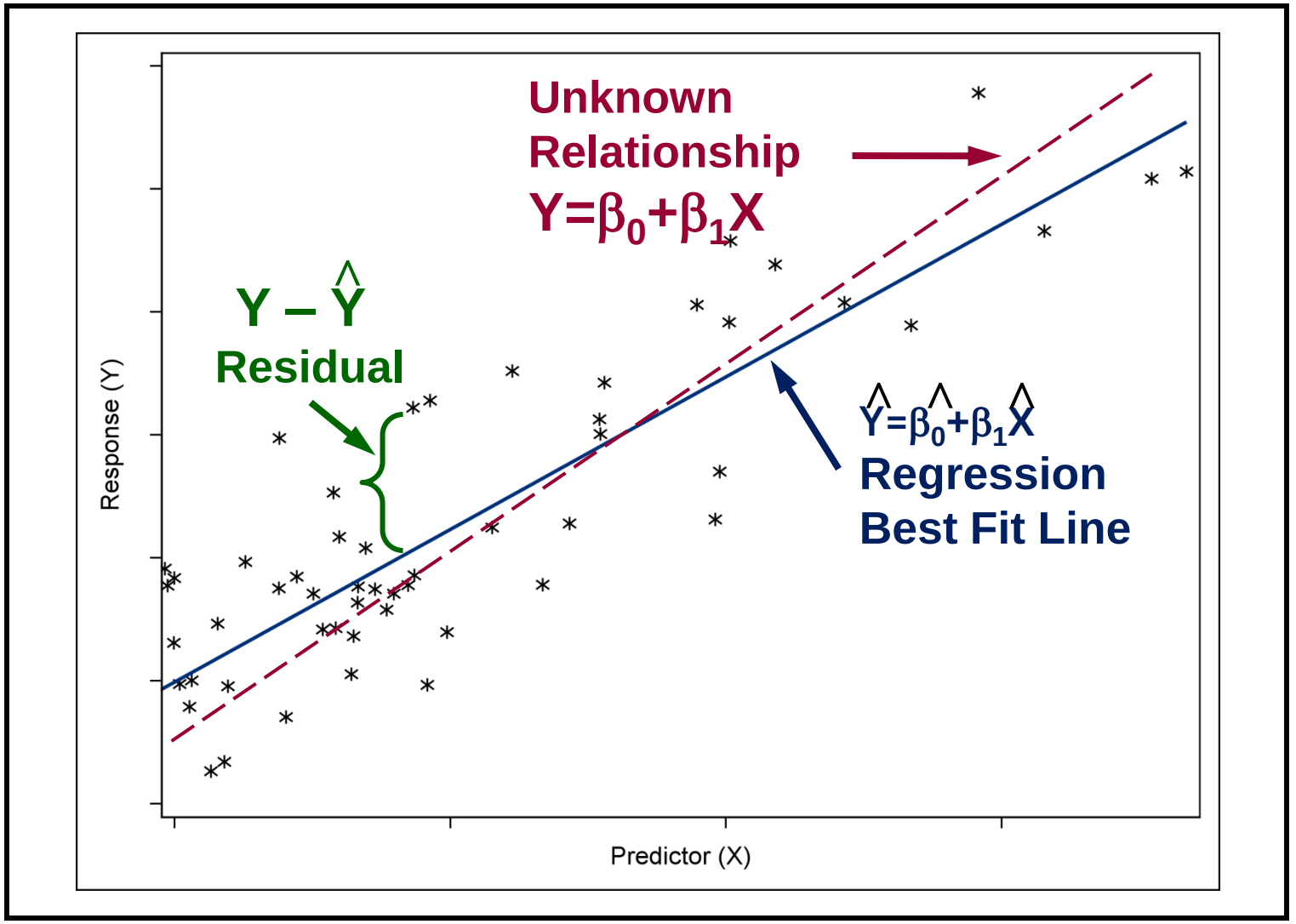
Oxygen_Consumption



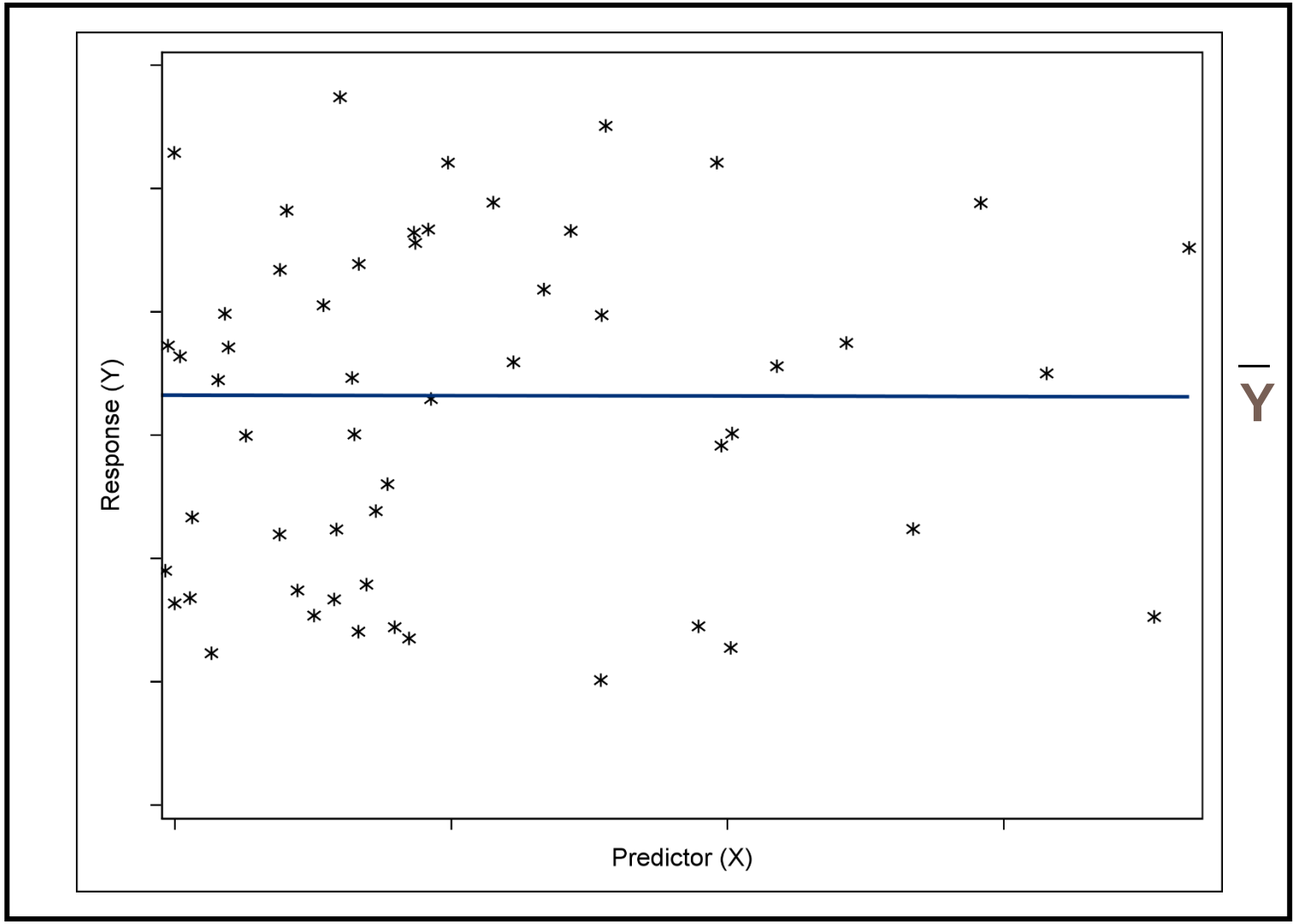
Simple Linear Regression Model



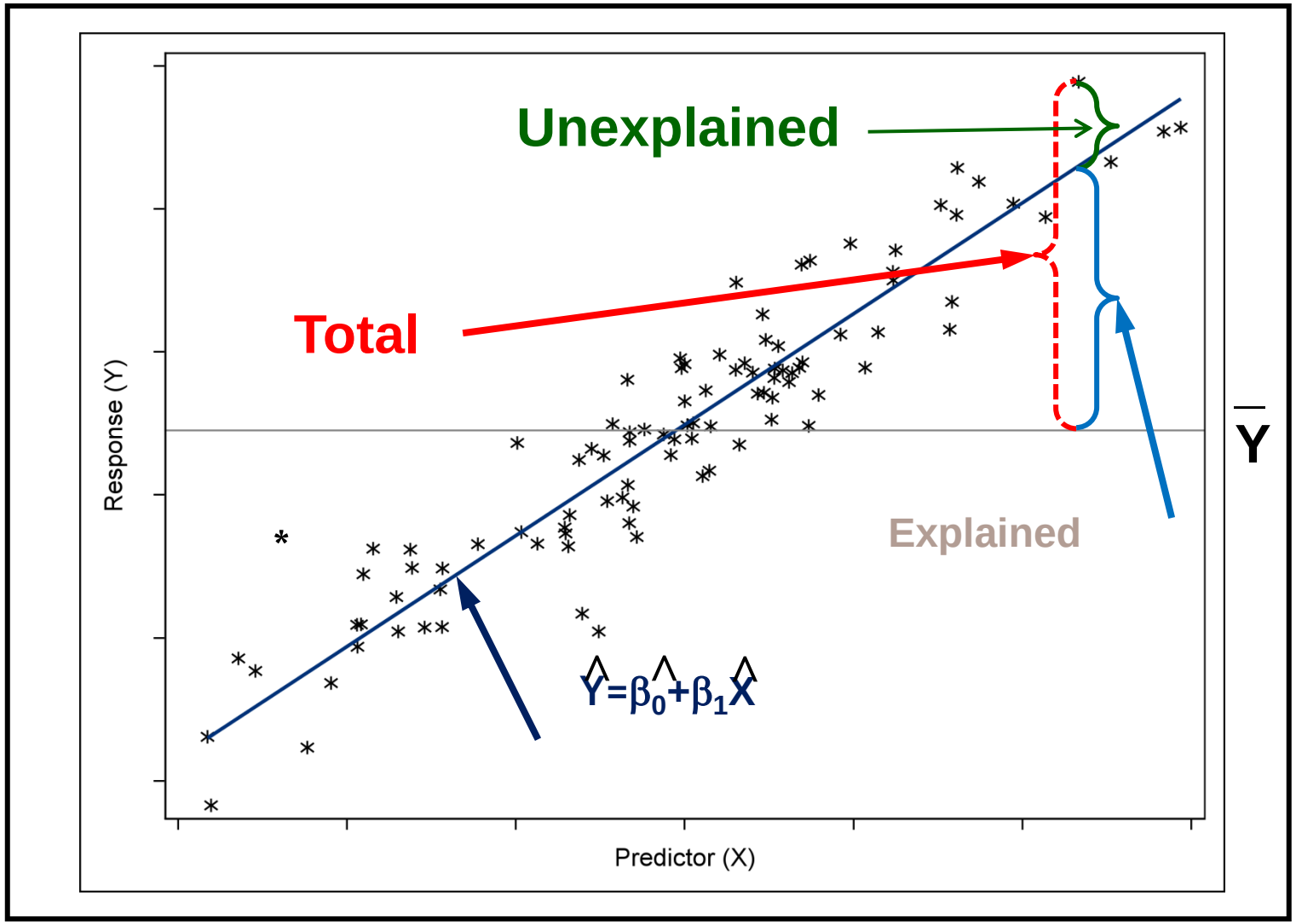
Simple Linear Regression Model



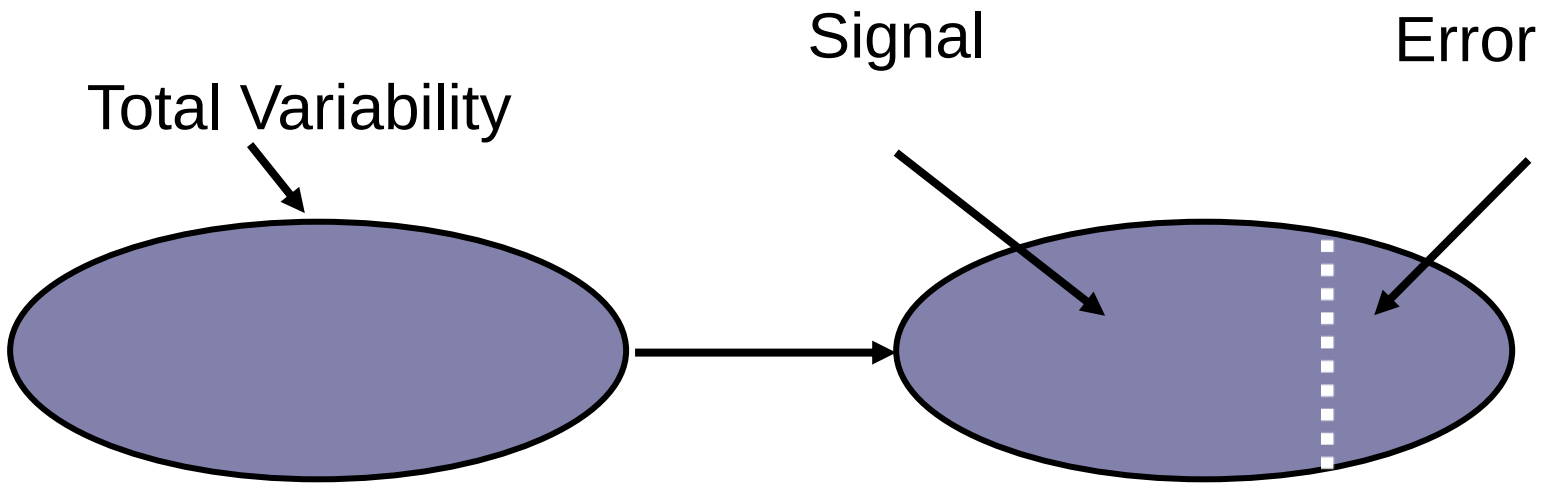
The Baseline Model



Explained versus Unexplained Variability



Partitioning Variability in Regression



$SS_T = SS_{reg} + SS_{res}$

Coefficient of Determination

$$R^2 = SS_{reg} / SS_T$$

"Proportion of variance accounted for by the model"

Correlation coefficient

- In this special case of simple regression:

$$r = \sqrt{R^2}$$

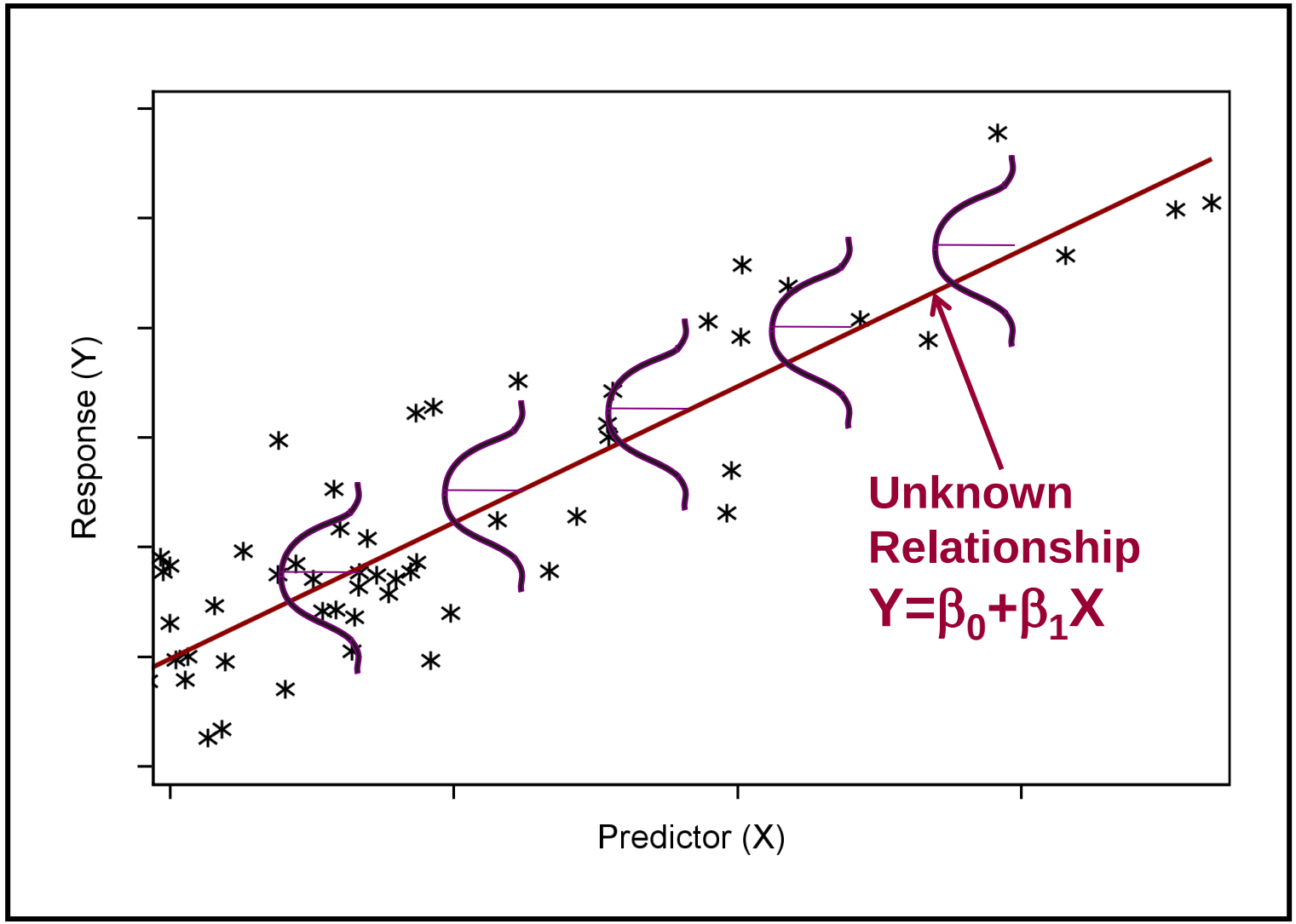
Pearson Correlation Coefficient

- Of historical note:

$$r = \frac{cov(X,Y)}{\sqrt{var(X)x\ var(Y)}}$$

*Covariance of X and Y is the product of the deviation of X and Y from their respective means

Assumptions of Simple Linear Regression



End of Lecture 1

Next up in Part 5 Lecture 2: Reliability