

LINEAR REGRESSION

- Continuous Outcomes
- Main Effect
- Covariates
- Correlation
- Regression
- Adjusted Model
- Assumptions
- Linearity
- Independence
- Homogeneity

OBJECTIVE

To identify key concepts associated with linear regression

PURPOSE

To review assumptions and appropriate use of linear regression

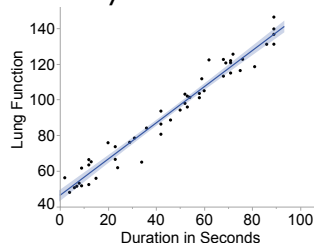
COMMON VARIABLES

- **Continuous outcomes:** BMI value, weight (in lbs or kg), systolic blood pressure (mmHG), disease specific scales like the MHI (0-100), Timed 25 Foot Walk (in seconds)
- **Main Effect:** In clinical trials, the main predictor (effect) is the focus of the study, e.g. treatment arms, intervention arms, etc.
- **Any type of covariate variables:** continuous or categorical variables for adjustment, e.g. sex, race, age, disease stage, etc.

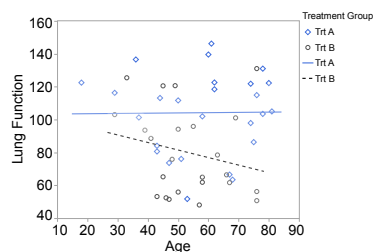
LINEAR REGRESSION

Regression is the most commonly used statistical method when predicting continuous outcomes as a function of treatment groups or interventions, with or without adjusting for relevant covariates.

- **Simple Linear Regression:** This is the most basic form of linear regression, when there is a single continuous outcome and a single continuous predictor. Also used to determine **Correlation, also called Pearson's correlation or Pearson's r**. Consider lung function output predicted by duration of inhaled medication:



- **Multiple (Covariate Adjusted) Regression:** The main effect of interest is usually treatment group, with adjustment for covariates (variables also related to the outcome but not of main interest):



INTERPRETATION & CONSIDERATIONS

Goal: The goal with linear regression is to explain as much of the variability in the outcome with the least amount of predictors, this is called parsimony.

Interpretation: The primary focus is to answer the question "Is there a difference in the outcome by Treatment Group?"

- Report the mean, or how the mean changes, by treatment group
- Explain how the covariates are related to the outcome, if applicable

Assumptions: The following need to be shown to be true or assumed to be true to use linear regression:

- **Linearity:** using a line to predict the outcome must be reasonable, there should not be any curvature or obvious pattern in the plots
- **Independence:** The observations must come from independent participants, that is, the observations cannot be related people
- **Homogeneity:** This is another word for variance, the variance in the outcome needs to be roughly the same for all levels of the predictors

Why are these important? If these assumptions are not valid, then neither are the linear regression results.

Bottom line: Use a statistician when you need linear regression.

