Midterm Study Guide

- Know each of the definitions and be able to explain their significance
- Be able to explain any important relationships between concepts or statistics that we discussed in class (e.g. how random treatment assignment relates to causal inference and how standard deviation is related to variance)
- Be able to explain the ideas a formula expresses in plain English
- Be familiar with all of the notation in a formula and what it means

Definitions

Causal Inference

- Random assignment
- Difference in means estimator
- Treatment effect
- Average treatment effect
- Fundamental problem of causal inference
- Counterfactual
- Random treatment assignment
- Randomized controlled experiment/trial
- Representative Sample
- Sampling frame
- DAG
- Treatment
- Control
- Pre-treatment characteristics
- Observational and experimental data

Populations and Samples

- Population
- Sample
- Sampling frame
- Representative sample
- Random sample
- Quota sample

Descriptive Statistics

- Mean
- Median
- Robust statistic
- Standard deviation
- Variance
- Z score
- Correlation

Regression

- Predictors, features, independent variables
- Outcome variable, dependent variable
- Predicted values, fitted values
- Residuals, errors
- Sum of squared residuals
- Total sum of squares
- Linear regression
- Ordinary least squares
- Root mean squared error
- R^2
- Statistical Model
- Data generating process
- Model fitting

Formulas

Formulas

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Standard Deviation

$$Z_i^X = \frac{(X_i - \bar{X})}{sd(X)}$$

Z score

Formulas

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Linear regression

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

Linear regression line

$$cor(X, Y) = \frac{\sum_{i=1}^{n} Z_i^X \times Z_i^Y}{n}$$

Correlation coefficient