Types of Variables

- Quantitative
  - Nominal
  - Ordinal
- Qualitative
  - Interval
  - Ratio

Differentiate between discrete + continuous variables
Discrete variables are countable and have a true 0 (# of pregnancies) while continuous may fall btwn integers + don't have a true 0 (temperature)

Measures of Central Tendency

Mean	Median	Mode
• Calculated avg • Continuous, normally distributed data	• 50th percentile • Ordinal or continuous data • Great for skewed data	• Most common value • Nominal, ordinal or continuous • Bimodal, trimodal, etc. • Does not describe meaningful distribution of large range of infrequent values

Measures of Data Spread / Variability

- Standard Deviation: Measure of variability around the mean
  - The larger the SD, the (more or less) spread out the results
  - Square root of the variance
  - Continuous data w/ normal to near normal distribution
  - 68 % of data w/in 1 SD
  - 95 % of data w/in 2 SD
  - 99.7 % of data w/in 3 SD

Measures of Data Spread / Variability (cont'd)

- Range: Difference btwn largest + smallest data points
  - Usually reported as lowest + highest data points
  - Very sensitive to outliers
- Percentiles: The value e which point the values larger than XX% of the data
  - Does not assume normal distribution
  - Interquartile range: 25 % to 75 %

Normal (Gaussian) Distribution

- $\bar{x}$ = sample mean s = standard deviation

* ALSO KNOWN AS PARAMETRIC DATA BC IT MEETS *
* DEFINING CRITERIA W/ PARAMETERS, MEAN, + SD *

Hypothesis Testing + Error

H_0: No difference exists between the groups being compared
H_A: A difference exists between the groups being ared
P-value: Whether a statistically significant difference exists btwn groups
Confidence Interval: 95% confident true value falls btwn these 2 values.

- Interpret the difference btwn wide + narrow CIs: Wide CI = true value can fall in a wide range of #s, Narrow CI = true value can fall in a smaller range of values
- 95% CI for risk ratios containing 1 indicate equal risk, thus no statistical significance.
- 95% CI for continuous variables containing 0 indicate no difference in variables, thus no statistical significance.
- Type I Error = (α or β)
- Type II Error = (α or β)
- Power: the ability to detect a difference if one exists
  - Equation: $1-\beta$

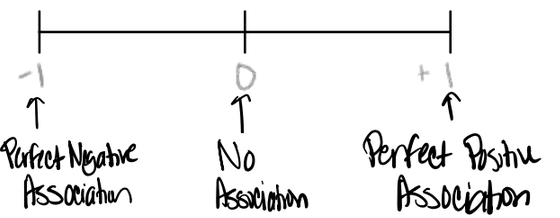
Test Result	Underlying Truth / Reality: H_0 is True	H_0 is False
H_0 is Accepted		Type II Error False (+ or −)
H_0 is Rejected	Type I Error False (+ or −)	

Outcome Variable		Exposure Variable	Normal Distribution	Skewed Distribution
Continuous		1 Group	1 sample t test	Sign test / Signed Rank Test
		2 Group	2 sample t test	Mann-Whitney U test
		Paired	Paired t test	Wilcoxon signed rank test
		> 2 Group	One-Way ANOVA test	Kruskal-Wallis test
Categorical		1 Group	Chi-square test / Exact test	
		2 Group	Chi-square / Fishers exact test	
		Paired	McNemar's Test	
		> 2 Group	Chi-square test / Fisher's exact Test / Logistic Regression	
Survival		2 Groups	Kaplan-Meier Plot w/ Log-Rank Test	
		> 2 Groups	Kaplan-Meier Plot w/ Log-Rank Test	
		Continuous	Cox Regression	

Differentiate between correlation + regression

Correlation = Association
Regression = Causation

Correlation Scale



Why does regression matter?
It helps us develop a prediction model
How can we determine accuracy?
Coefficient of determination (R^2)
$R^2 = 0$ → Model explains no variability Outcome is random w/ respect to predictors
$R^2 = 1$ → Perfect Prediction. Model explains all variability