# Quantitative Stats - 2

Kyungsik Han

# What We Will Learn?

- Summarizing data
  - variables
  - simple statistics
  - effect statistics
  - statistical models
  - complex models
- Generalizing from sample to population
  - precision of estimate
  - confidence limits
  - statistical significance
  - p value
  - errors

# Variable Types

- Independent Variables (IV)
  - X
- Dependent Variables (DV)
  - Y
- Control Variables (= constant variables)
- Moderating Variables (affects the strength of the relationship)

IV → DV

MoV

- Mediating Variables

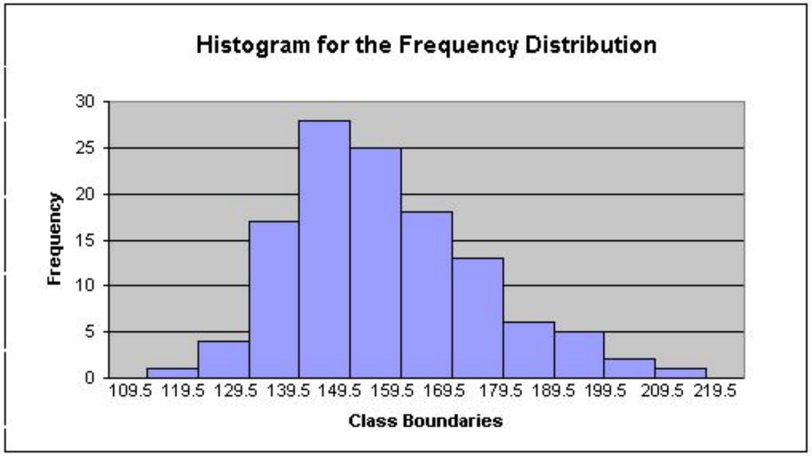
IV → MeV → DV

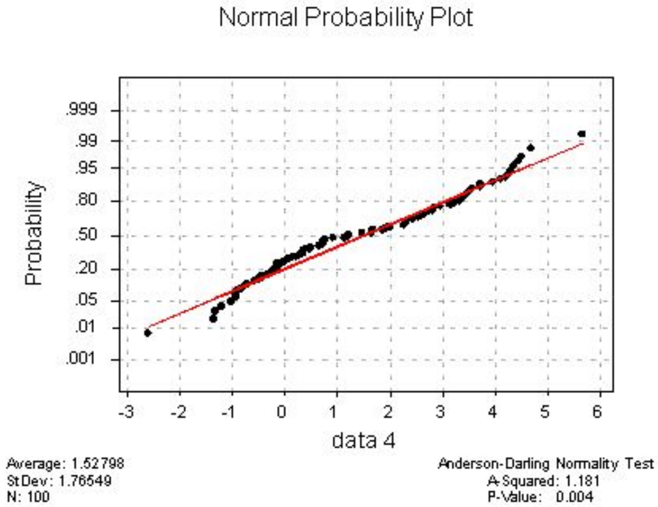
# Summarizing Data

- Data are a bunch of values of one or more variables
- A variable is something that has different values
  - Values can be numbers or names
  - **Numeric** (e.g., weight)
  - **Counting** (e.g., number of injuries)
  - **Ordinal** (e.g., competitive level): values are numbers or names
  - **Nominal** (e.g., sex): values are names

# Summarizing Data

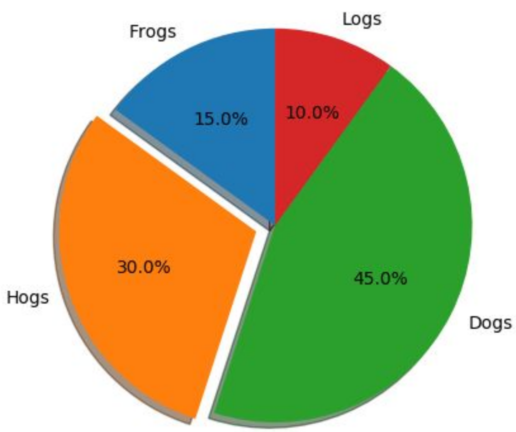
- When values are numbers, visualize the distribution of all values in a frequency histogram
- Can also use normal probability plots to visualize how well the values fit a normal distribution

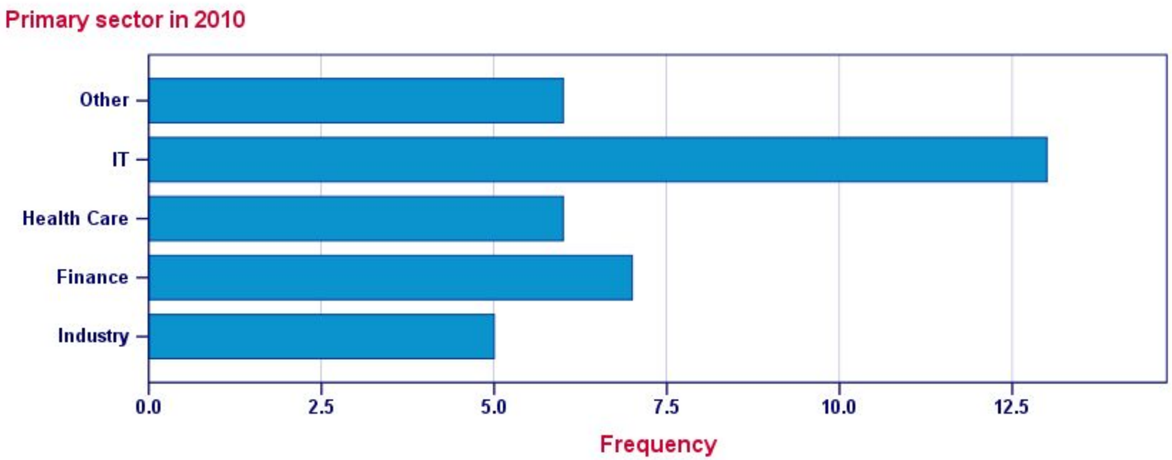


# Summarizing Data

- When values are names, visualize the frequency of each value with a pie chart or a list of values and frequencies

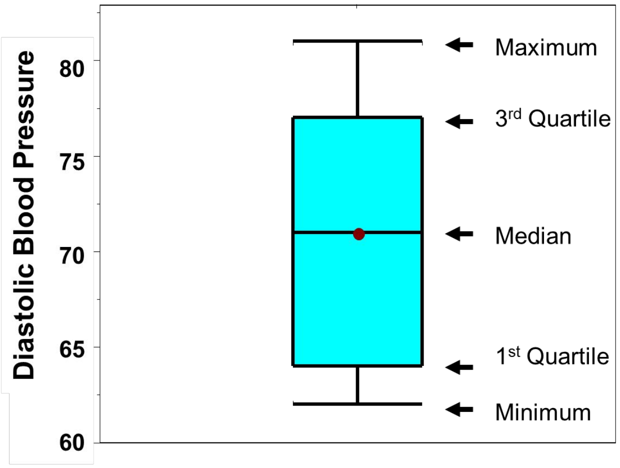


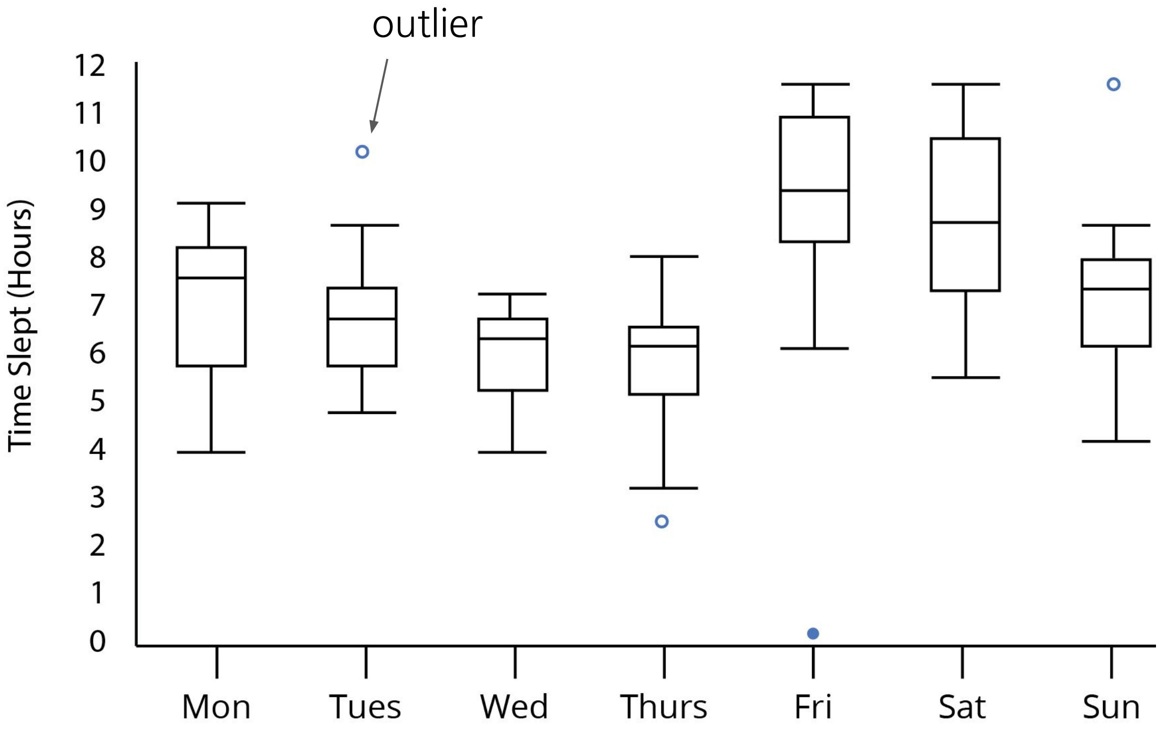



# Statistic

- A statistic is a number summarizing a bunch of values
  - Simple or univariate statistics summarize values of one variable
  - Effect or outcome statistics summarize the relationship between values of two or more variables
- Simple statistics for numeric variables
  - Mean: the average
  - Standard deviation: the typical variation
  - Standard error of the mean: the typical variation in the mean with repeated sampling
  - Median: middle value or 50th percentile
  - Quartiles: 25th and 75th percentiles
- Summarize these and other simple statistics visually with box and whisker plots

# Box and Whisker Plots



# More Statistics

- Simple statistics for nominal variables
  - Frequencies, proportions, or odds
  - Can also use these for ordinal variables

- Effect statistics
  - Derived from statistical model (equation) of the form Y (dependent) vs X (independent or predictor)
  - Depend on type of Y and X

# Significant Tests

Why do we need significant tests?

- When the values of the members of the comparison groups are all known, you can directly compare them and draw a conclusion. In this case, no significant test is needed since there is no uncertainty involved
- When the population is large, we can only sample a sub-group of people from the entire population
- Significant tests allow us to determine how confident we are that the results observed from the sampling population can be generalized to the entire population

# Significant Tests

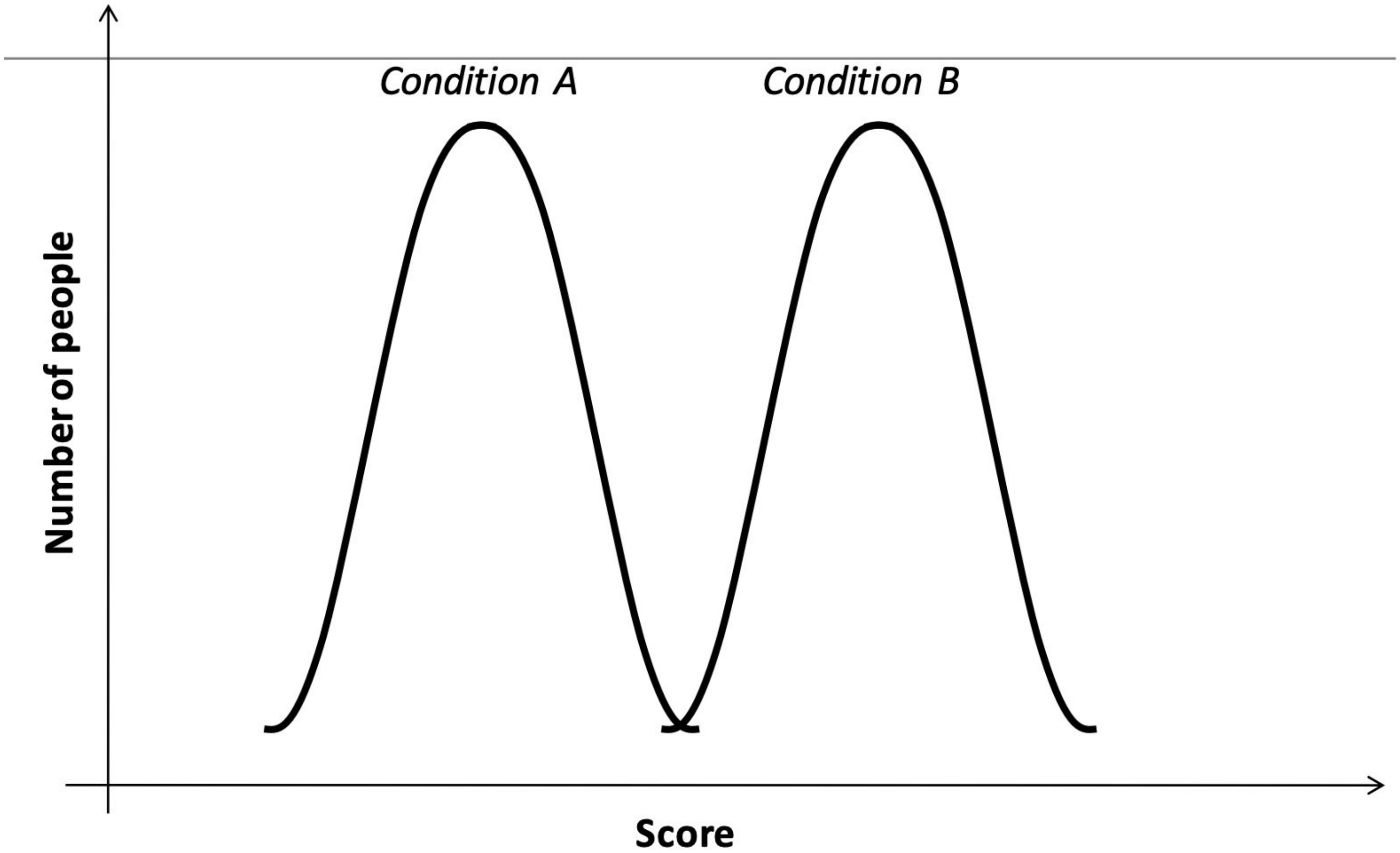
Example

- You recruit 30 people, 15 of which do a test using a touchpad, and 15 of which do the same test using a trackball. You end up with 30 measures of *throughput*. The mean for the touchpad is 4.30 clicks/s. The mean for the trackball is 5.08 clicks/s.

Can you conclude that trackball is better than the touchpad?

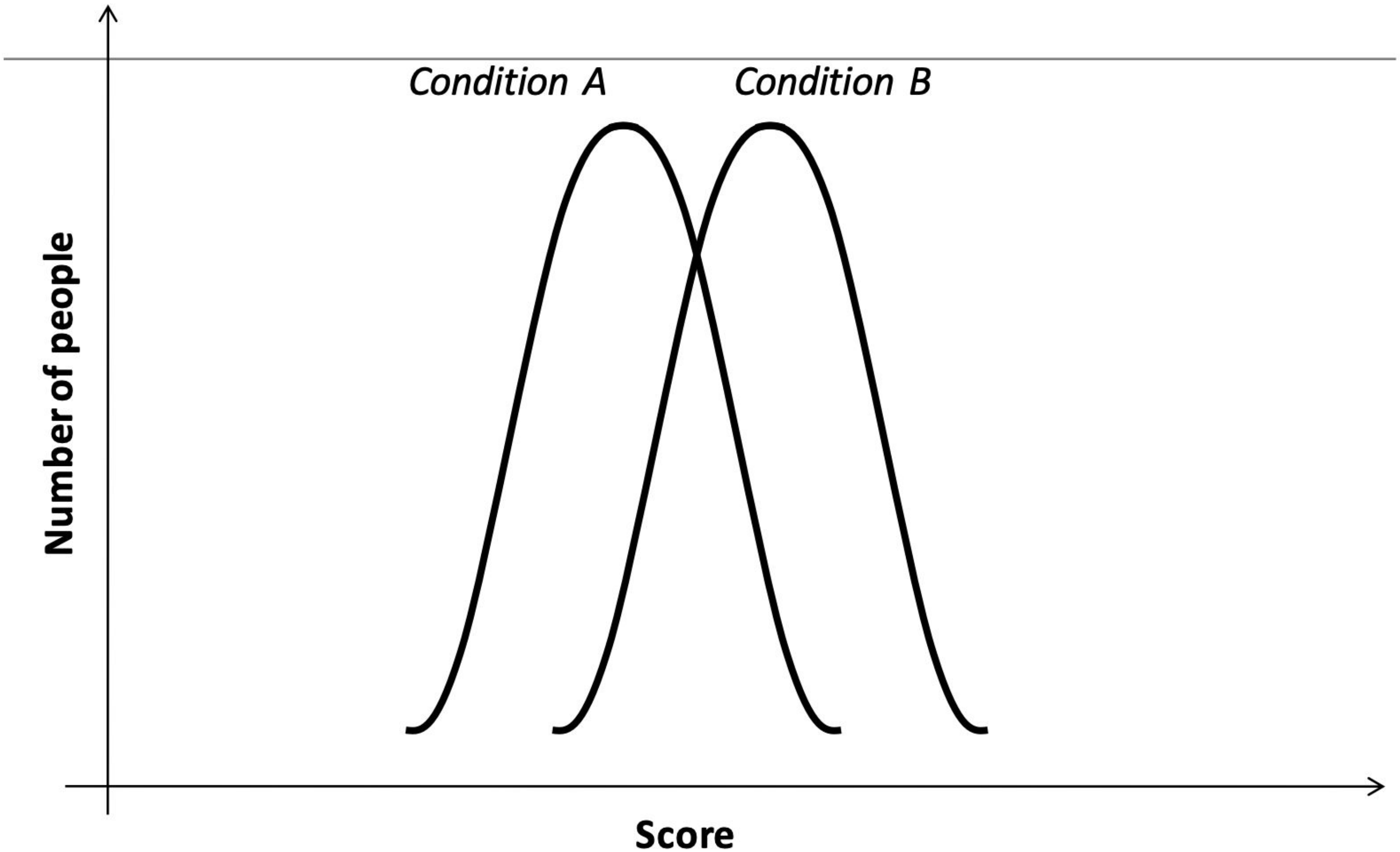
# Significant Tests

Are they different?



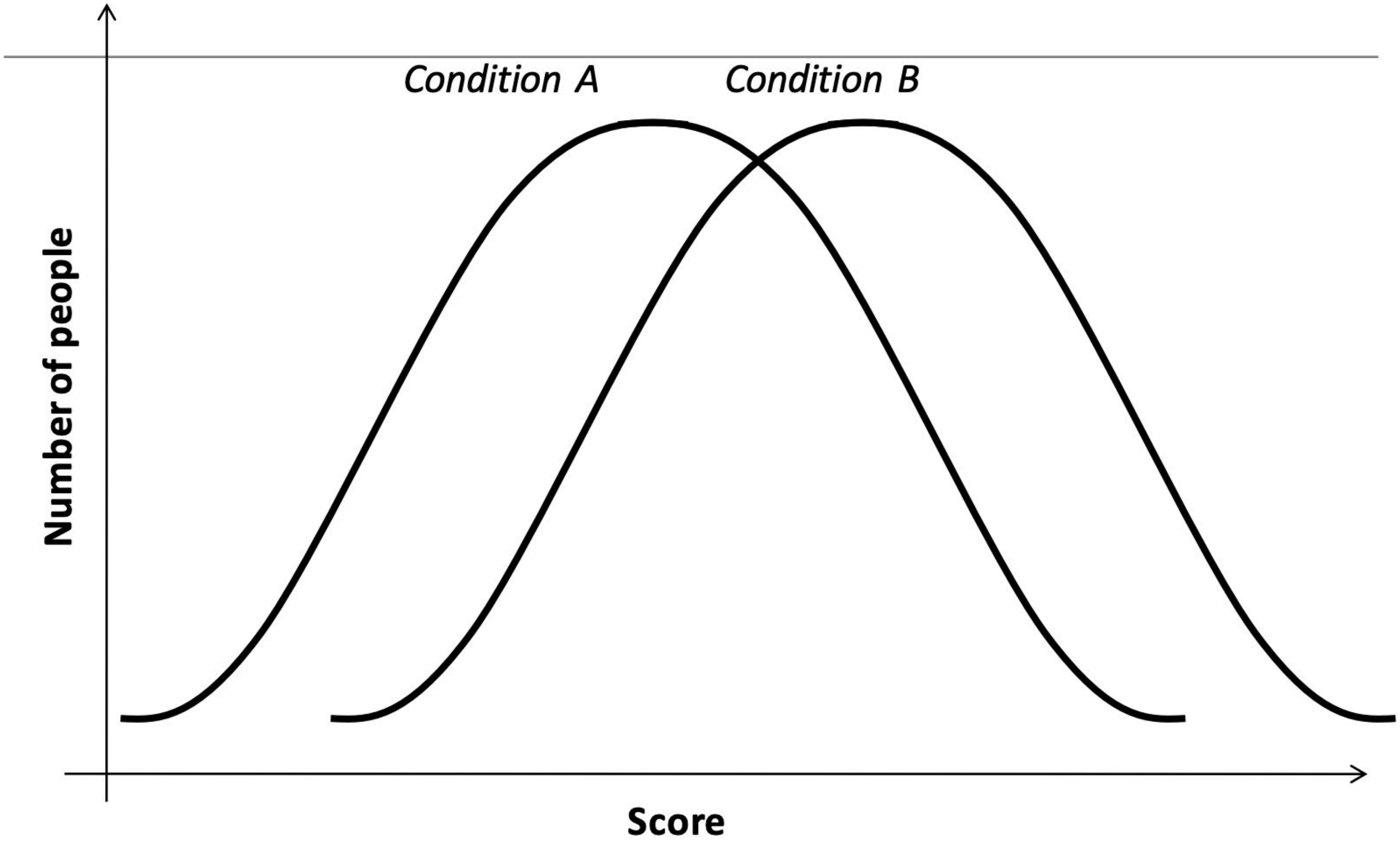
# Significant Tests

Are they different?



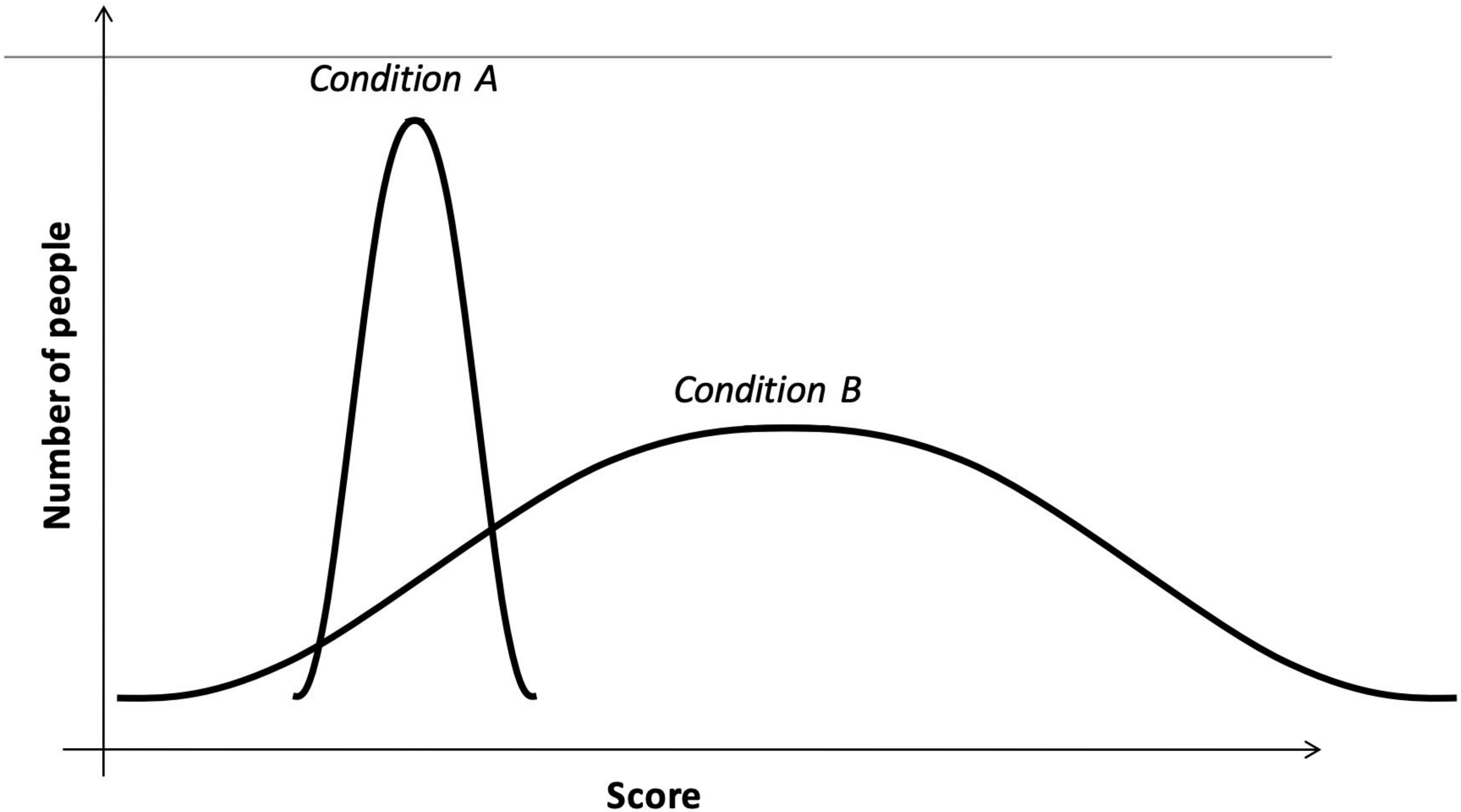
# Significant Tests

Are they different?



# Significant Tests

Are they different?



# Significant Tests

Bottom line

- You cannot just compare means
- You must take “spreads” into account
- Statistics can perform analyses of variance, or the “amount of spread” around means to tell us how reliable/probable a real difference is
    - A real difference is a “statistically significant difference”
    - An unreliable difference is a “statistically non-significant difference”

# Statistically Significant?

- Statistical significance is an old-fashioned way of generalizing, based on testing whether the true value could be zero or null
  - Assume the null hypothesis: that the true value is zero (null)
  - **Null hypothesis** ($H_0$): no difference between certain characteristics of a population
  - **Alternative hypothesis** ($H_1$): the opposite hypothesis if $H_0$ is rejected (i.e., there is a difference between two groups)
  - If you observed value falls in a region of extreme values that would occur only **5% (0.05)** of the time, you reject the null hypothesis
- The ***p-value*** helps you decide whether your result falls in the unlikely region
  - If $p < 0.05$, your result is in the unlikely region

# Significant Tests

p-values

- We perform an analysis of variance and get a p-value
- The p-value comes from the sampling distribution of the sample mean
- The p-value is the probability of randomly getting a test statistics as (or more) extreme than what you observed if the null hypothesis was true
    - i.e., the probability that your results occurred by chance
    - p = 0.45 means there is a 45% chance the data occurred by chance
    - p = 0.05 means there is a 5% chance the data occurred by chance

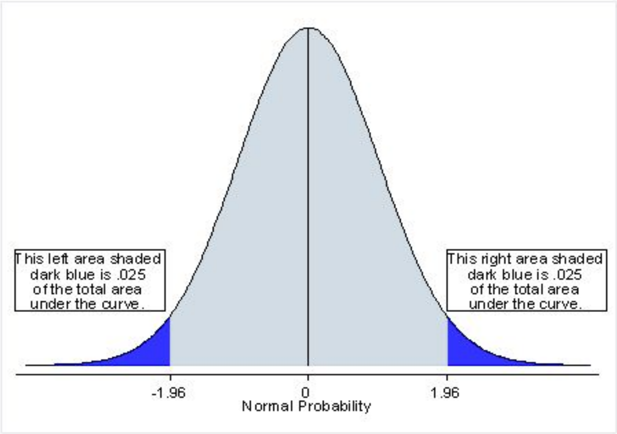
# Significant Tests

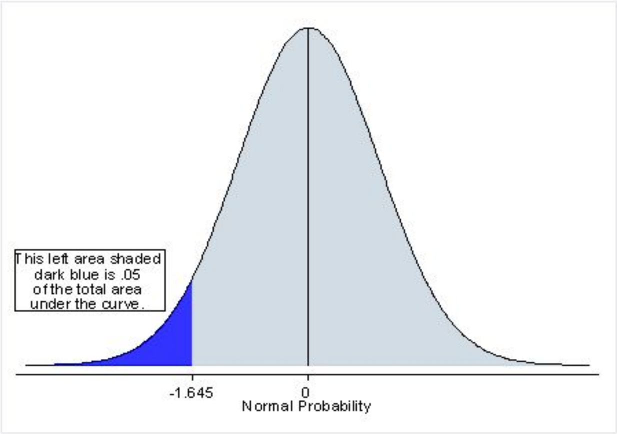
- We now need to use the p-value to choose a course of action
- Either reject the null hypothesis, or fail to reject the null hypothesis
- We need to decide if our sample result is unlikely enough to have occurred by chance
- Standard cutoff of $p < 0.05$ (we are at least 95% confident that our results did not occur by chance)

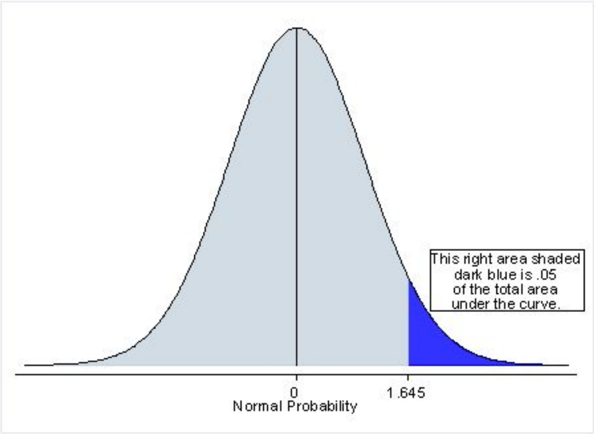
# Better Meaning of p-value

- If you observe a positive effect, **1 - p/2** is the chance that the true value is **positive**, and **p/2** is the chance that the true value is **negative**
  - Example: you observe a 1.5% enhancement of performance ($p$=0.08). Therefore there is a 96% chance that the true effect is any "enhancement" and a 4% chance that the true effect is any "impairment"
- If you must use p-values, **showing exact values is more preferred** than just $p<0.05$ or $p>0.05$
- Generally (and empirically), p-values with 0.01, 0.05, 0.10 are used
  - But be careful using 0.10 (meaning marginally significant)

# One-tailed or two-tailed t-test




two-tailed t-test

one-tailed t-test

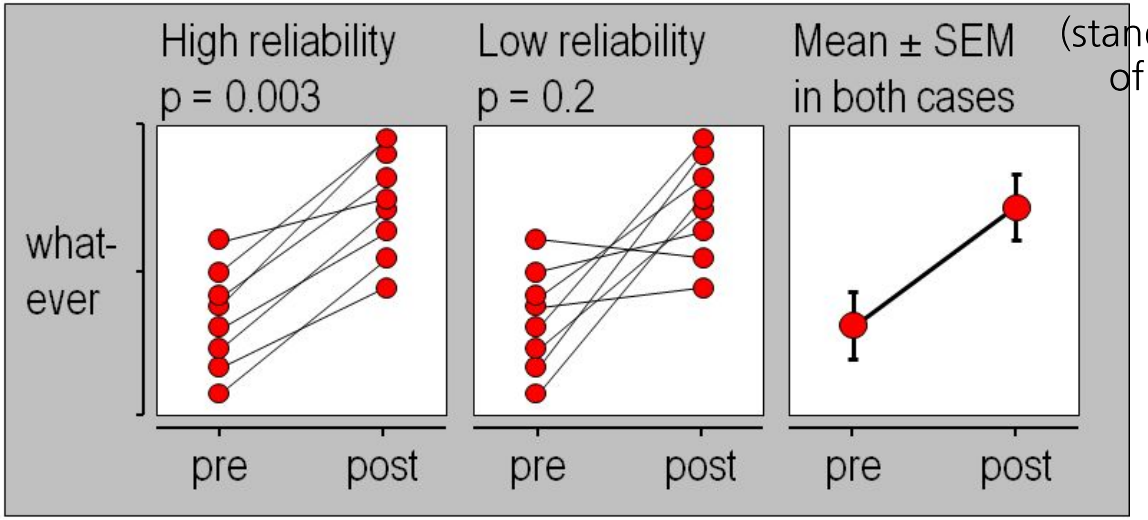
# More on p-value

- If the true value is zero, there's a 5% chance of getting statistical significance: the Type I error rate, or rate of false positives or false alarms
- There's also a chance that the smallest worthwhile true value will produce an observed value that is not statistically significant: the Type II error rate, or rate of false negatives or failed alarms

| | | Reality: True | Reality: False |
|---|---|---|---|
| Measured or Perceived | True | Correct ☺ | **Type 1 error** False Positive |
| | False | **Type 2 error** False Negative | Correct ☺ |

# Standard Error of the Mean

- Typical variation in the mean from sample to sample
- It can convey statistical significance
  - Non-overlap of the error bars of two groups implies a statistically significant difference, but only for groups of equal size (e.g., males vs females)
  - In particular, non-overlap does NOT convey statistical significance in experiments



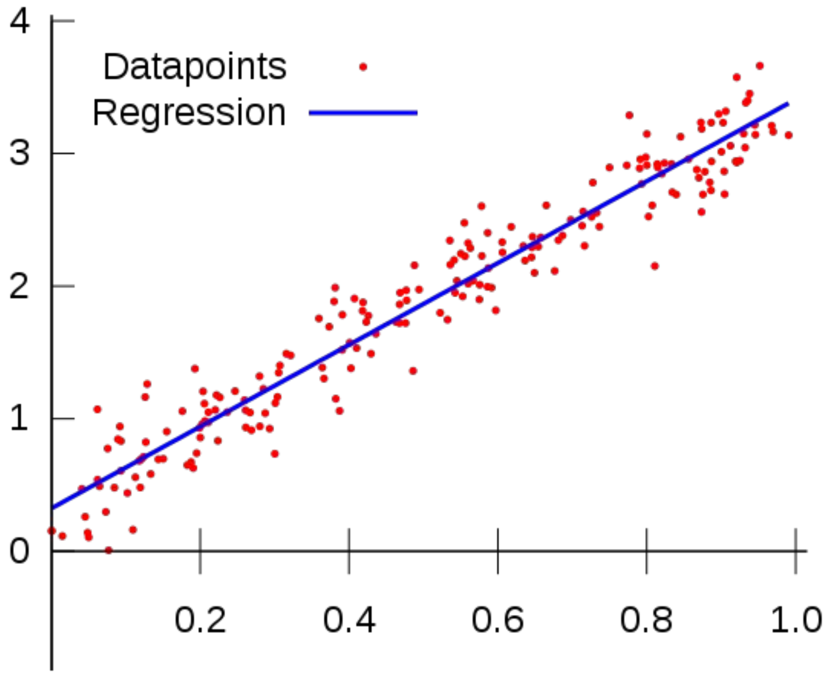

(standard error of mean)

# Statistical Test Summary

| X | Y | Model/Test | Effect Statistics |
|---|---|---|---|
| numeric | numeric | regression | slope, intercept, correlation |
| numeric | nominal | categorical | frequency ratio |
| nominal | numeric | t-test, ANOVA | mean difference |
| nominal | nominal | chi-square | frequency difference or ratio |

# What is Goodness-of-Fit?

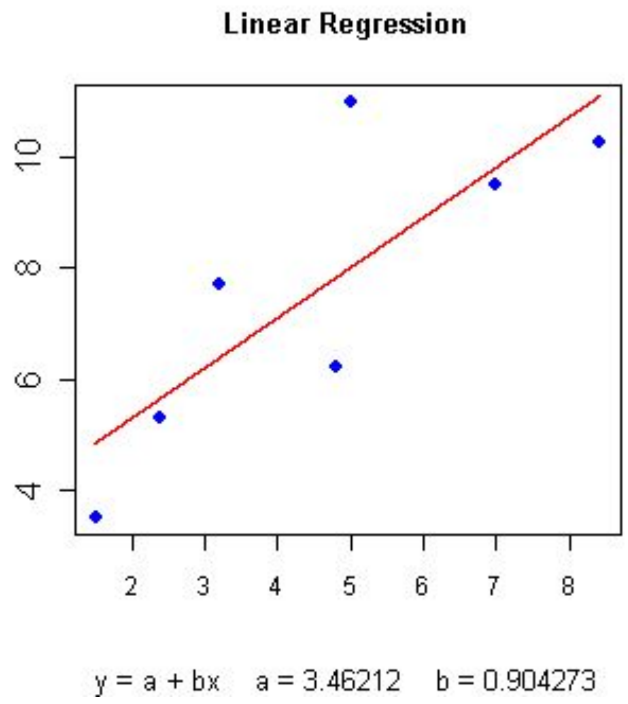
- Description of how well a statistical model fits a set of observations
- Summarize the discrepancy between observed values and the values expected under the model in question



# Numeric (X) vs Numeric (Y)

- Example
  - Body fat vs sum of skinfolds
- Model or test
  - Linear regression
- Effect statistics
  - Slope and intercept = parameters
  - Correlation coefficient OR variance explained = measures of goodness of fit

# Numeric (X) vs Numeric (Y) - Cont'd




univariate regression

**Coefficients[a]**

| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | (Constant) | -11.619 | 82.306 | | -.141 | .897 |
| | AGE | -.786 | .752 | -.558 | -1.045 | .373 |
| | CHILDC | -5.626 | 4.035 | -.621 | -1.394 | .258 |
| | FINISH | -3.017 | 5.535 | -.167 | -.545 | .624 |
| | GENDER | 4.960 | 10.515 | .311 | .472 | .669 |
| | HEALTHC | .485 | 1.424 | .416 | .341 | .756 |
| | INCOMEC | .566 | .989 | .709 | .572 | .607 |
| | LIFESATC | .809 | .329 | 1.148 | 2.460 | .091 |
| | MARRIED | -20.542 | 6.736 | -1.216 | -3.050 | .055 |
| | SES | 9.132E-02 | .375 | .127 | .244 | .823 |
| | SMOKE | 3.173 | 14.296 | .188 | .222 | .839 |
| | SPIRITC | .427 | .182 | .645 | 2.345 | .101 |

a Dependent Variable: LIFESAT7

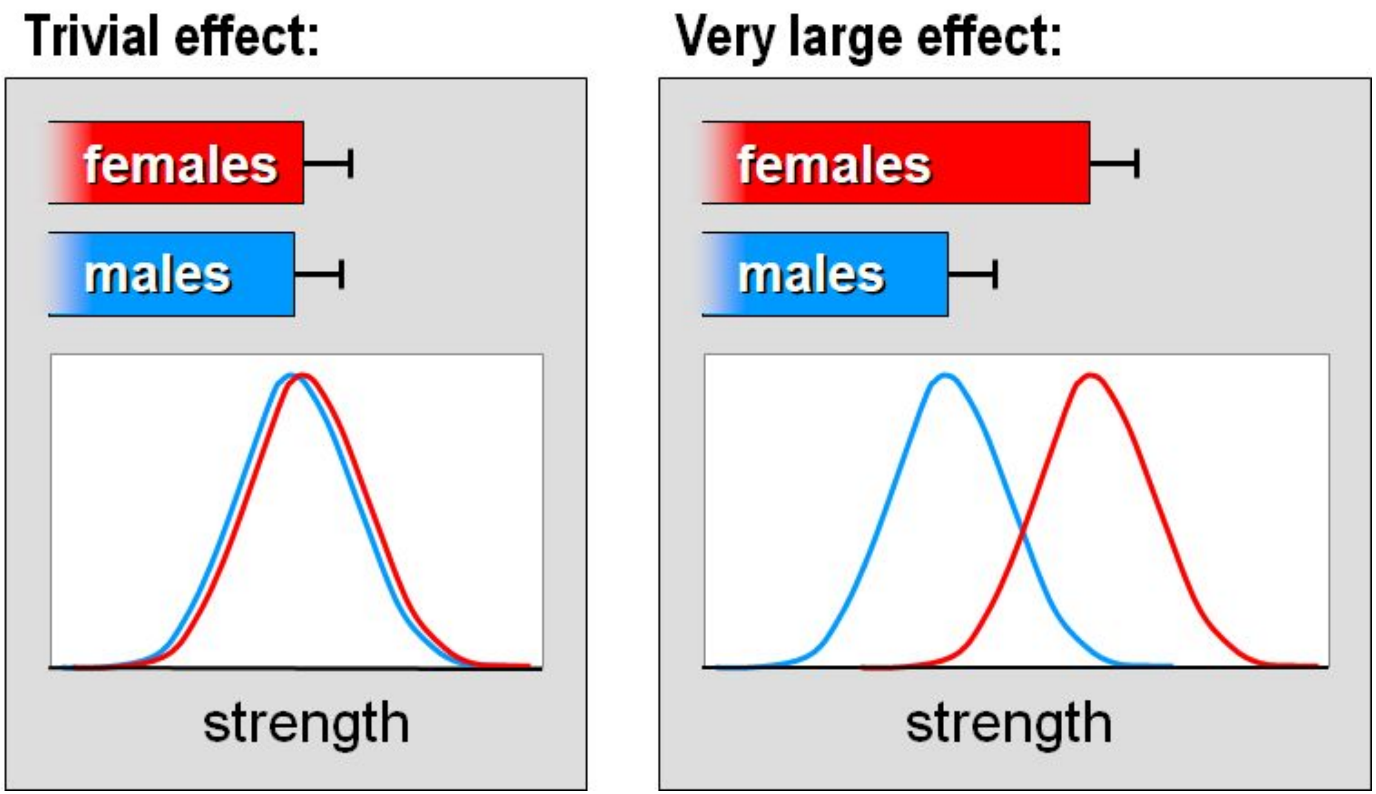
multivariate regression

# Nominal (X) vs Numeric (Y)

- Example
  - Sex vs Strength
- Model or test
  - t-test (2 groups)
  - 1-way ANOVA ( > 2 groups)
- Effect statistics
  - Difference between means expressed as raw difference, percent difference, or fraction of the root mean square error (Cohen's effect-size statistic)
  - variance explained or better = measures of goodness of fit

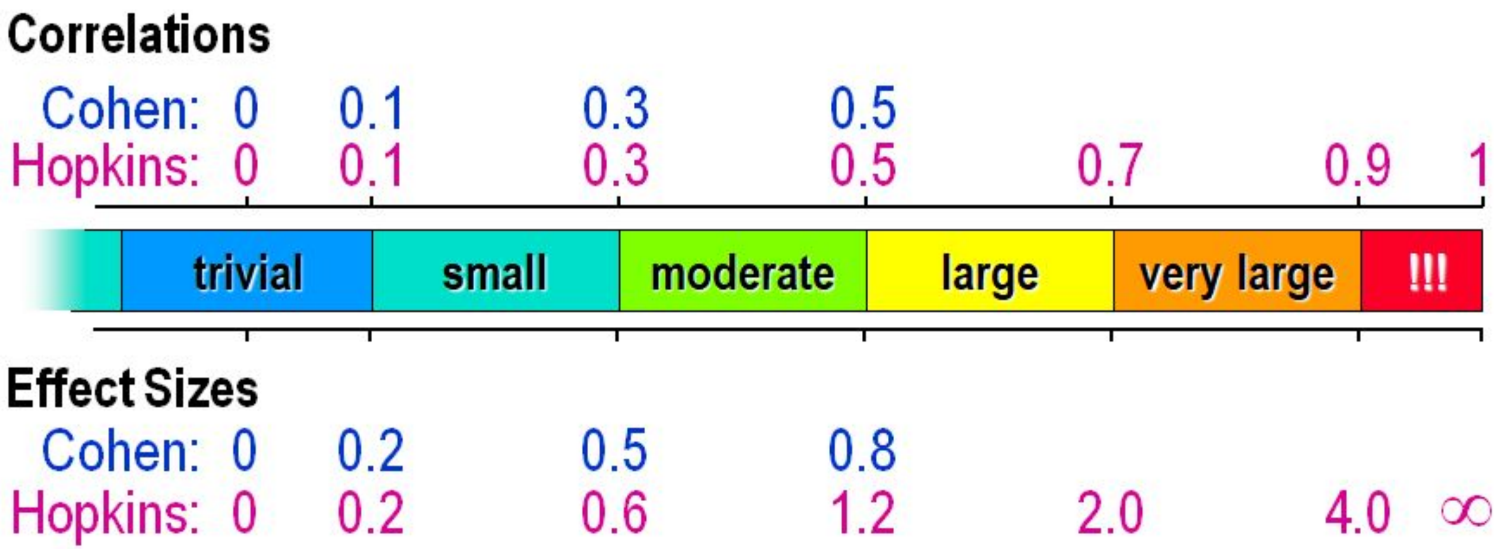
# Nominal (X) vs Numeric (Y) - Cont'd

- More on expressing the magnitude of the effect
  - What often matters is the difference between means relative to the standard deviation



# Correlations & Effect size

- Fraction or multiple of a standard deviation is known as the effect-size statistic (or Cohen's "d")
- Cohen suggested thresholds for correlations and effect sizes
- Hopkins agrees with the thresholds for correlations but suggests others for the effect size:



# More on Effect Size

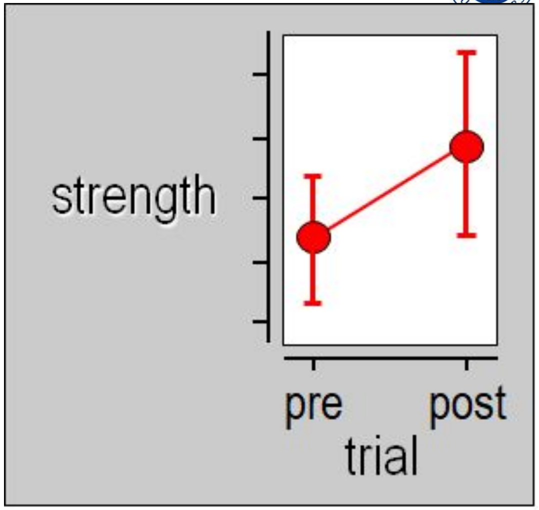
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174/

TABLE 1 **Common Effect Size Indices**[a]

| Index | Description[b] | Effect Size | Comments |
|---|---|---|---|
| Between groups | | | |
| Cohen's $d$[a] | $d = M_1 - M_2 / s$<br>$M_1 - M_2$ is the difference between the group means ($M$);<br>$s$ is the standard deviation of either group | Small 0.2<br>Medium 0.5<br>Large 0.8<br>Very large 1.3 | Can be used at planning stage to find the sample size required for sufficient power for your study |
| Odds ratio (OR) | $\frac{\text{Group 1 odds of outcome}}{\text{Group 2 odds of outcome}}$<br>If OR = 1, the odds of outcome are equally likely in both groups | Small 1.5<br>Medium 2<br>Large 3 | For binary outcome variables<br>Compares odds of outcome occurring from one intervention vs another |
| Relative risk or risk ratio (RR) | Ratio of probability of outcome in group 1 vs group 2;<br>If RR = 1, the outcome is equally probable in both groups | Small 2<br>Medium 3<br>Large 4 | Compares probabilities of outcome occurring from one intervention to another |
| Measures of association | | | |
| Pearson's $r$ correlation | Range, −1 to 1 | Small ±0.2<br>Medium ±0.5<br>Large ± 0.8 | Measures the degree of linear relationship between two quantitative variables |
| $r^2$ coefficient of determination | Range, 0 to 1;<br>Usually expressed as percent | Small 0.04<br>Medium 0.25<br>Large 0.64 | Proportion of variance in one variable explained by the other |

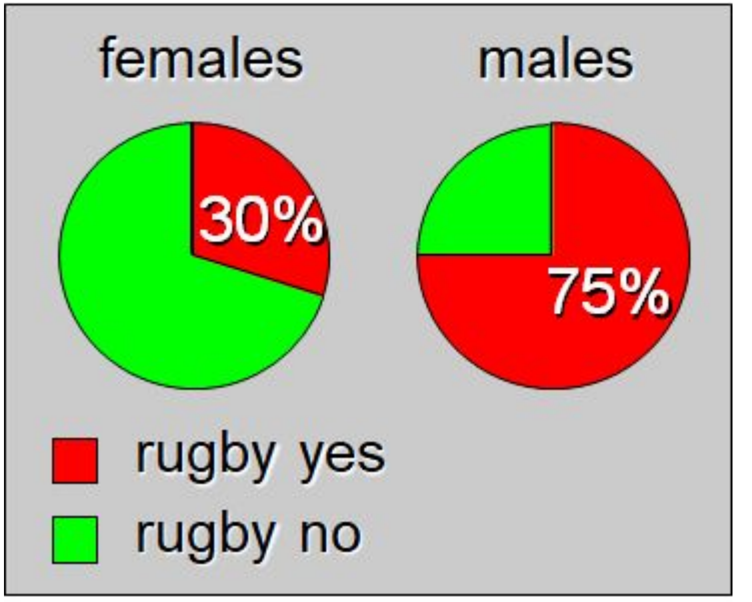
[a] Adapted from Ferguson et al.[9]
[b] Based on Soper.[7]

# Nominal (X) vs Numeric (Y) - Cont'd



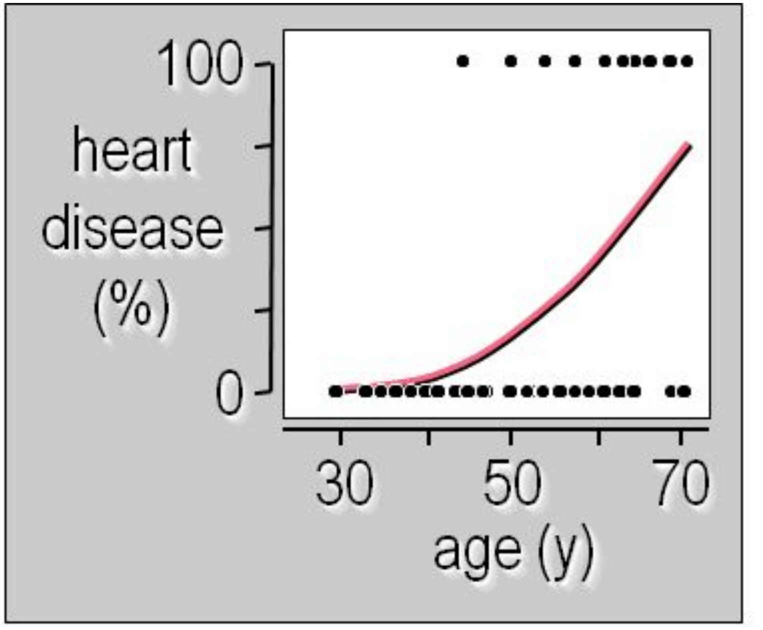

- Example
  - Trial vs Strength
- Model or test
  - paired t-test (2 trials)
  - repeated-measures ANOVA with one within-subject factor ( > 2 trials)
- Effect statistics
  - Change in mean expressed as raw change, percent change, or fraction of the pre-standard deviation

# Nominal (X) vs Nominal (Y)




- Example
  - Sex vs Sports
- Model or test
  - Chi-squared test or contingency table
- Effect statistics
  - Relative frequencies, expressed as a difference in frequencies, ratio of frequencies, or ratio of odds (odds ratio)
  - Relative risk is appropriate for cross-sectional designs (e.g., across cases)
    - risk of having rugby disease for males relative to females is $(75/100)/(30/100) = 2.5$
  - Odds ratio is appropriate for case-control designs (e.g., within male case and female case)
    - calculated as $(75/25)/(30/70) = 7.0$

# Numeric (X) vs Nominal (Y)




- Example
  - Age vs heart disease
- Model or test
  - Categorical modeling
- Effect statistics
  - Relative risk or odds ratio per unit of the numeric variable
- Model
  - logistic regression or generalized linear modeling