Data Presentation, Analysis, & Statistics

Data Management

- Develop a hypothesis
- Design a study
- Collect data
- Graph data

Look for levels and trends

Describe the results (descriptive statistics)

Look at distributions and relationships

Describe the results (descriptive and inferential statistics)

Presentation of Data

- Graphic presentation of data
  - Frequency distribution
  - Graphs
  - Box Plots
- Key elements
  - Title, axis labels, legend
  - Major axis divisions, axis number formats
- Optional elements
  - Values, variation/error bars
  - Secondary axes

Why is it Important to Graph?

Graphing in Excel
Graphing in Excel

Questions to Ask

• Is the graph appropriate for the data?
  - Bar chart, pie chart, line graph, etc.
• Is the data scaled correctly?
  - Raw or normalized data
• Is the graph scaled properly?
  - Consistent if multiple graphs
  - Zero-baseline versus arbitrary baseline

Data Analysis

Types of Data

• Nominal
  - Categorical data without order or value
  - Example: classifications such as etiology, foot type, etc.
• Ordinal
  - Rank ordered (i.e., from a large value to a small value)
  - Difference among values unknown
  - Example: Survey responses such as "happy," and "no preference"
• Interval
  - Rank ordered
  - Difference among values the same
  - Zero does not correspond to minimum value
  - Example: Fahrenheit temperature scale
• Ratio
  - Rank ordered
  - Difference among values the same
  - Zero corresponds to minimum value
  - Example: time or percent correct

Visual Analysis

• Inter and intra-phase analysis
• Level changes
  - Mean
  - Median
• Slope changes
  - Slope of trend line
  - Hand-draw (i.e. split-middle technique)*
  - Compute (i.e. regression fit)
• Other issues
  - Immediate change
  - Latency - a delay in the desired effect

* Richards et al. "Single Subject Research" p. 272

* Richards et al. "Single Subject Research" p. 272
**Visual Analysis**

- Different methods of visual analysis
  - Mean, median, range, etc.

**Visual Analysis**

- Complexity of real data

**Visual Analysis**

- Split-middle technique

**Group Statistics**

**Key Elements**

- Statistical analysis tests hypothesis
- Stats do not prove results
  - Statistical analysis performed on NULL hypothesis
- Statistics usually make assumptions
  - Type of distribution (i.e. normal dist.)
  - Independence of variables
  - Variability of the dataset(s)

**Types of Statistics**

- Descriptive statistics - Characterization of distribution, central tendency, and variability of data
  - Shape of the distribution (i.e., skewness)
  - Frequency of results (i.e., histogram)
  - Central tendency (i.e., mean, median, or mode)
  - Variability (i.e., standard deviation or standard error)
- Inferential statistics - Estimation of population characteristics based upon sampled data
  - Probability (i.e., p-value)
  - Confidence (i.e., confidence interval)
  - Hypothesis testing (i.e., t-test, ANOVA, Chi-Square, etc.)
Descriptive Statistics

• Used to characterize a single dataset
  • Measures of central tendency
    - Mean
    - Median
    - Mode
  • Measures of variability
    - Range
    - Variance
    - Standard deviation

Measures of Central Tendency

• Mean
  - The average number or score in a dataset

• Median
  - The number or score in a dataset above and below which 50% of all other numbers or scores are located

• Mode
  - The number or score that occurs most often in a dataset

Measures of Variability

• Range (R)
  - Distance between the highest and lowest measure
    \[ R = x_{\text{max}} - x_{\text{min}} \]

• Percentiles
  - The location of a score within a dataset

• Quartiles
  - The location of the score based on 25% increments

• Variance (V)
  - Average of the squared deviations from the mean
    \[ V = \frac{\sum (x - \bar{x})^2}{N-1} \]

• Standard deviation (σ)
  - Square root of variance
  - Often used to represent variability in means
    \[ \sigma = \sqrt{V} \]
Measures of Variability

- Variance ($V$)
  - Average of the squared deviations from the mean
  \[ V = \frac{\sum (x_i - \bar{x})^2}{N-1} \]

- Standard deviation ($\sigma$)
  - Square root of variance
  - Often used to represent variability in means

\[ \sigma = \sqrt{V} \]

Correlation

- Defines the linear relationship between variables
- Correlation coefficients range from +1.0 to -1.0
  - +1.0 = perfect positive correlation
  - 0.0 = no correlation
  - -1.0 = perfect negative correlation

Correlation Coefficient

- Pearson’s Coefficient ($r$)
  - Most popular measure of correlation
  - Applicable to Ratio/interval data
  - Assumes normal distribution
  - Correlation measured with r-statistic
  \[ r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \]
  - Available in Microsoft Excel

- Spearman Rho ($r_S$)
  - Applicable to ordinal/nominal data

Correlation in Excel

Step 1: Create chart & input data
Step 2: Compute coefficient
Syntax: =pearson(a1, a2)

Step 3: Evaluate coefficient
Analyze the correlation coefficient

Inferential Statistics

- Also known as “predictive statistics”
- Used to compare n datasets
- Inferential stats allow:
  - Scientifically test a hypothesis
  - Make inferences onto a larger population
  - Influenced by a study design and analysis choices
    - Sample size
    - Data variability
    - Desired probability
    - Potential error
- Inferential statistics depend on collected data
  - Parametric
  - Non-parametric
**Statistical Tests**

(from Lunsford and Lunsford, 1996)

**Probability**

- Statistics assess probability
- Level of significance
  - $\alpha$ level, $\alpha$
  - The probability a measured difference is due to chance
  - $\alpha = 0.05$ (5%) is the typical level of significance
- Power
  - The likelihood of finding true differences
  - Influences on statistical power
    - Sample size
    - Effect size (i.e., "the magnitude of the difference")
    - Data variability
    - Level of significance
  - Power often used to estimate sample size
    - Need effect size, variability, level of significance

**Error**

- Type I Error
  - A measured significant difference is really due to chance
- Type II Error
  - When no measured difference is really present

**Type of data/test**

<table>
<thead>
<tr>
<th>Parametric</th>
<th>Non-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two samples (related)</td>
<td>Paired T-test</td>
</tr>
<tr>
<td>Two samples (independent)</td>
<td>Independent T-test</td>
</tr>
<tr>
<td>Three samples</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Correlation</td>
<td>Pearson correlation coefficient</td>
</tr>
<tr>
<td>Frequency analysis</td>
<td></td>
</tr>
</tbody>
</table>

**T-test**

- Most common statistical test and most misused
- Used for only two datasets
  - Related samples
  - Independent samples

**Basic Statistics in Excel**

Step 1: Input data & create chart
Step 2: Compute means (i.e., average)

Note: T-Test can be done in Microsoft Excel
Basic Statistics in Excel

Step 3: Compute median values
Syntax: \( \text{median}(n_1, n_2, n_3, \ldots, n_k) \)

Step 4: Compute T-test p-value
Syntax: \( \text{ttest}(a_1, a_2, \text{tails}, \text{type}) \)

Step 5: analyze data
5a) Compare the means
5b) Compare the medians
5c) Analyze the p-value

Analysis of Variance

- Three or more datasets
- Significance requires follow-up ("post-hoc") test
  - Tukey Honestly Significant Differences (HSD)
  - Neuman-Kuels Test
  - Scheffe Test
  - Bonferroni Correction

Note: ANOVA can be done in Excel, but post-hoc test requires additional software or knowledge of appropriate formulas

Significance

- Statistical significance
  - Proper study design
  - Proper statistical analysis
  - Requires interpretation
    - What does significance mean?
    - Is statistical significance clinically relevant?

- Clinical significance
  - A difference that promotes a clinical change
  - Moves a patient outside daily fluctuations
  - Unfortunately, we do not often know daily changes
  - 5 - 10% is likely a good estimation

Suggested Reading

- Portney & Watkins
  - Chapter 17 – Descriptive Statistics
  - Chapter 18 – Statistical Inference
  - Chapter 19 – the T-Test
  - Chapter 20 – Analysis of Variance
  - Chapter 21 – Multiple Comparison Tests
  - Chapter 22 – Nonparametric Tests for Group Comparisons
  - Chapter 23 – Correlation
  - Chapter 24 – Regression
  - Chapter 25 – Measures of Association for Categorical Variables (Chi-Square)
  - Chapter 26 – Statistical Measures of Reliability
  - Chapter 27 – Statistical Measures of Validity
  - Chapter 28 – Epidemiology: Measuring Risk
  - Chapter 29 – Multivariate Analysis
  - Chapter 30 – Data Management

- Richards et al.
  - Chapter 13 – Methods for Analyzing Data

For Next Week

- Class Discussions
  - Use rubric as template

- Discussion schedule (30 min each)
  - 5/07 – Megan, Emily
  - 5/14 – Frances, Jed, Chatelaine
  - 5/21 – Sarah, Stephen, Jennifer

- Final proposal
  - Due no later than Friday, June 1st, 5pm
  - Electronic copies