Best Practices for Statistics in Your Own Projects

William F. Sensakovic, PhD, DABR, MRSC
Medical Physicist – Florida Hospital
Assoc. Prof. – University of Central Florida
Clin. Asst. Prof. – Florida State University
Adj. Prof. – Adventist University

E-mail: wfsensak@gmail.com
Twitter: @wfsensak

July 31, 2017

William F. Sensakovic, PhD

Relevant Conflicts of Interest

No relevant conflicts of interest... but I would be happy to have someone change that

July 31, 2017

William F. Sensakovic, PhD

1. Use a Statistician . . . Stay Involved

July 31, 2017

William F. Sensakovic, PhD
Clinical Significance . . . Not Their Job

• 52% dose difference
  – Matters? . . . Depends on who you ask
• Could get better quality with same dose using different vendor’s product
  – Matters? . . . More interesting
2. Multiple Comparisons

- 28 Patients
- 13 Tissue Types
- 3 Scanners
- 2 Kernels
- 0 Hypothesis testing

3. How Good is My Model: R²

- What we measure
  - Response variable
  - Outcome variable
  - Dependent variable

- Simple Linear Regression = determine the best fit line to the data

- Concentration of Iodine [mg/mL]

- What we investigate
  - Predictor variable
  - Explanatory variable
  - Independent

- Importance: determines the best fit line to the data
Coefficient of Determination

- Total Sum of Squares
  - How much does the data differ from the mean?
  - Variation in the data from all sources

- Regression Sum of Squares
  - How much do model-predicted values differ from the mean?
  - Variation of the model

- How much variation does the model explain?
  - \( R^2 = \frac{\text{Model Variation}}{\text{Total Data Variation}} \)
  - 0: accounts for no variation
  - 1: perfect fit

\[ y = 0.3437x \]
\[ R^2 = 0.933 \]
How much variation does the model explain?
- $R^2 = 0.933$ or ~93%
- So not bad, but does that mean it is the correct model... Maybe

What about the other 7%?
- Random?
- Measurement error
- Other predictor?
- Wrong model?

Linear model
- $R^2 = 0.933$
- Power model
- $R^2 = 0.995$

Is power model the correct fit?... maybe
- $R^2$ just tells you how much variation the model accounts for, nothing more
4. Bland-Altman Analysis

- Pearson Correlation
  - Linear relationship
- Spearman Correlation
  - Non-linear relationship
- Best if dependent and independent variables are different categories
- Not an indication of accuracy!

4. Bland-Altman Analysis

- Best choice when measuring the same quantity with different methods!
  - Bias
  - Variance
  - Trends
5. Capturing “Clinical Practice”
Case Selection

- Comparisons should be on same cases
  - Sensitivity 25%-100% depending on case selection
- The normal case subtlety must be considered to ensure sufficient number of false-positive responses
- Study disease prevalence does not need to match disease population prevalence
  - ROC AUC stable between 2%-28% study prevalence, but small increases in observer ratings are seen with low prevalence
- Study disease prevalence does not need to match disease population prevalence
  - ROC AUC stable between 2%-28% study prevalence, but small increases in observer ratings are seen with low prevalence

Observer Selection

- Observer Experience
  - Sp 0.9:
    - Se - 0.76 (high volume mammographers)
    - Se - 0.65 (low volume mammographers)

Table 1. Patient and Scan Demographics

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Standard/Control</th>
<th>Control</th>
<th>Control</th>
<th>Observer (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>15.4±4.3</td>
<td>15.5±4.3</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>45.9±10.2</td>
<td>45.9±10.2</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Similarly (mm)</td>
<td>23.8±5.5</td>
<td>23.8±5.5</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Spine length (cm)</td>
<td>41.2±17.1</td>
<td>41.2±17.1</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Spine width (cm)</td>
<td>57.6±17.1</td>
<td>57.6±17.1</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Spine height (cm)</td>
<td>27.4±17.0</td>
<td>27.4±17.0</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

87% Lower

6. Choosing the Correct Test

<table>
<thead>
<tr>
<th>Continuous Dependent Variables</th>
<th>Independent Variables</th>
<th>Independent?</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 normal</td>
<td>1 categorical</td>
<td>Yes</td>
<td>t-test</td>
</tr>
<tr>
<td>1 normal</td>
<td>1 categorical</td>
<td>No</td>
<td>paired t-test</td>
</tr>
<tr>
<td>1 non-normal</td>
<td>1 categorical</td>
<td>Yes</td>
<td>rank sum</td>
</tr>
<tr>
<td>1 non-normal</td>
<td>1 categorical</td>
<td>No</td>
<td>signed rank</td>
</tr>
<tr>
<td>1 normal</td>
<td>1 normal continuous</td>
<td></td>
<td>Pearson</td>
</tr>
<tr>
<td>1 non-normal</td>
<td>1 non-normal continuous</td>
<td></td>
<td>Spearman</td>
</tr>
<tr>
<td>1 normal</td>
<td>&gt; 1 categorical</td>
<td>Yes</td>
<td>ANOVA</td>
</tr>
<tr>
<td>1 non-normal</td>
<td>&gt; 1 categorical</td>
<td>No</td>
<td>Kruskal-Wallis</td>
</tr>
</tbody>
</table>

- **Parametric**
  - Non-normal & 15-20 samples per category
  - Mean describes the data
- **Non-parametric**
  - Deals with outliers better
  - Median describes the data

Observers rate “image quality” at different doses
Cite This Talk/Handout

Sensakovic, WF. MO-F-201-3: Best Practices for Statistics in Your Own Projects. Med Phys. 44 (6), 3714 (2017); 3091-3092 DOI: 10.1002/mp.12304

Thank You!