Statistics Bests

marty@triolastats.com
Best Trends in Statistics

**Obsolete:**

- Printed tables
- $\sigma$ known case
- Normal approximation to binomial
Best Trends in Statistics

Analysis of Big Data

• Every 2 min now: As much data as from beginning of time to 2000
Best Trends in Statistics

From Technology:

• Simulations
• Bootstrapping
• Normal quantile plots (check requirements)
Wald CI for $p$

Wald CI: $\hat{p} - E < p < \hat{p} + E$ where $E = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Disadvantage of Wald:

“Coverage probability” is typically less than chosen confidence level.
Simulation Activity

• Generate 10,000 results from binomial:

\[ n = 100, \ p = 0.1 \]

• Find and compare:

95% coverage probability with Wald, plus 4, Score
Some Results

10,000 times: $n = 100, \ p = 0.1$

Wald: 0.9305

Plus 4: 0.9549

Score: 0.9397
Teaching CI for p

1. Wald (normal approximation)
2. “Plus four”: Do a few
3. Mention “there are others”:  
   • Score, Clopper-Pearson,  
   • Score with continuity correction  
   • Arcsine transformation.
And now . . .

a microcosm of the modern statistics course...
What is the single most important topic in statistics?
Variation
Best Intro to Variation

Bank lines changed from multiple lines to single line:

• Change was made to ... reduce variation!

• But $\bar{x}$ doesn’t change.
Mean = 7.15  Med = 7.2  Mode = 7.7  M.R. = 7.1

**Single** bank waiting line (*wait times* in minutes):
6.5  6.6  6.7  6.8  7.1  7.3  7.4  7.7  7.7  7.7

$s = 0.5 \text{ min}$

**Multiple** lines (*wait times* in minutes):
4.2  5.4  5.8  6.2  6.7  7.7  7.7  8.5  9.3  10.0

$s = 1.8 \text{ min}$
“See” Variation

Same mean, median, mode, midrange:

Wait Times (min)
Go beyond the formula!
Beyond the Formula

\[ s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} \]

Range Rule of Thumb

- Significantly low values: \( \mu - 2\sigma \)
- Values not significant: \( \mu \)
- Significantly high values: \( \mu + 2\sigma \)
Beyond the Formula

\[
\begin{align*}
99.9 & \quad 98.3 & \quad 99.3 & \quad 98.1 & \quad 98.5 & \quad 98.3 & \quad 99.0 & \quad 98.7 & \quad 98.3 & \quad 98.5 \\
97.9 & \quad 97.8 & \quad 98.5 & \quad 99.1 & \quad 97.4 & \quad 98.8 & \quad 99.0 & \quad 98.2 & \quad 99.5 & \quad 99.2
\end{align*}
\]

**Bad:** Give formula for \( s \), calculate \( s \), show all work.

**Good:** Find mean and s.d. In the context of these data, is 98.2 “significantly low”?
“Smoking is one of the leading causes of statistics.”

- Fletcher Knebel
Best Probability Problem
South Platte River
Glass of Water
River Seine
Unbelievable!

Probability that a molecule from a glass of water dumped in the South Platt River will show up in a glass scooped from the Seine in front of the Eiffel Tower:

\[ p = 0.9999+ \]
Best Probability Activity

• 3 closed doors: car, goat, goat
• You choose a door.
• Monte Hall opens another door without the car.
• Now he asks you to *stick* or *switch*.

What should you do?
Best Way to Teach Probability

• Stress *significance*: role in stats

• Practical: Redundancy

• Use *intuitive* rules, such as: “add without double counting”

• Keep it simple
Duke: James F. Bonk
Classic: Flat Tire

• 4 students miss test: “flat tire”
• Makeup: “Which tire?” (for 95 pts)
• Class activity: Which tire?
• Probability:
  \[ P(4 \text{ students choose } \text{specific} \text{ tire}) = \frac{1}{256}; \]
  \[ P(4 \text{ students choose } \text{same} \text{ tire}) = \frac{1}{64} \]
• Goodness-of-fit
Best Answer

Which tire went flat?

the spare!
## Best Data Set

<table>
<thead>
<tr>
<th>98.6</th>
<th>98.6</th>
<th>98.0</th>
<th>98.0</th>
<th>99.0</th>
<th>98.4</th>
<th>98.4</th>
<th>98.4</th>
<th>98.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.6</td>
<td>98.8</td>
<td>98.6</td>
<td>97.0</td>
<td>97.0</td>
<td>98.8</td>
<td>97.6</td>
<td>97.7</td>
<td>98.8</td>
</tr>
<tr>
<td>98.0</td>
<td>98.3</td>
<td>98.5</td>
<td>97.3</td>
<td>98.7</td>
<td>97.4</td>
<td>98.9</td>
<td>98.6</td>
<td>99.5</td>
</tr>
<tr>
<td>97.3</td>
<td>97.6</td>
<td>98.2</td>
<td>99.6</td>
<td>98.7</td>
<td>99.4</td>
<td>98.2</td>
<td>98.0</td>
<td>98.6</td>
</tr>
<tr>
<td>97.2</td>
<td>98.4</td>
<td>98.6</td>
<td>98.2</td>
<td>98.0</td>
<td>97.8</td>
<td>98.0</td>
<td>98.4</td>
<td>98.6</td>
</tr>
<tr>
<td>97.8</td>
<td>99.0</td>
<td>96.5</td>
<td>97.6</td>
<td>98.0</td>
<td>96.9</td>
<td>97.6</td>
<td>97.1</td>
<td>97.9</td>
</tr>
<tr>
<td>97.3</td>
<td>98.0</td>
<td>97.5</td>
<td>97.6</td>
<td>98.2</td>
<td>98.5</td>
<td>98.8</td>
<td>98.7</td>
<td>97.8</td>
</tr>
<tr>
<td>97.1</td>
<td>97.4</td>
<td>99.4</td>
<td>98.4</td>
<td>98.6</td>
<td>98.4</td>
<td>98.5</td>
<td>98.6</td>
<td>98.3</td>
</tr>
<tr>
<td>98.8</td>
<td>99.1</td>
<td>98.6</td>
<td>97.9</td>
<td>98.8</td>
<td>98.0</td>
<td>98.7</td>
<td>98.5</td>
<td>98.9</td>
</tr>
<tr>
<td>98.6</td>
<td>97.1</td>
<td>97.9</td>
<td>98.8</td>
<td>98.7</td>
<td>97.6</td>
<td>98.2</td>
<td>99.2</td>
<td>97.8</td>
</tr>
<tr>
<td>98.4</td>
<td>97.8</td>
<td>98.4</td>
<td>97.4</td>
<td>98.0</td>
<td>97.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Best Terms

Homoscedasticity

Bootstrapping
Best Technology

- TI-83/84 Plus calculator
- StatCrunch: online
- Excel: need an add-in
- Minitab: great, underused
- Statdisk: free, easy
- SPSS, SAS, JMP, ...

Use 2 technologies!!!
Best Graph

Losses of Soldiers in Napoleon's Army During the Russian Campaign (1812–1813)
(Width of band shows size of army.)

Best “Impactful” Graph

Deaths:
British Military Hospitals During Crimean War

Legend:
Outer: Deaths from preventable diseases
Middle: Deaths other than wounds or preventable diseases
Inner: Deaths from wounds in battle
Best Misleading Graph
Best “Worst” Graph

% College Enrollment Under 25

<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>72.0</td>
</tr>
<tr>
<td>1973</td>
<td>70.8</td>
</tr>
<tr>
<td>1974</td>
<td>67.2</td>
</tr>
<tr>
<td>1975</td>
<td>66.4</td>
</tr>
<tr>
<td>1976</td>
<td>67.0</td>
</tr>
</tbody>
</table>
Best New Controversy

**NHST:**

“Null Hypothesis Significance Test”
P-Value and NHST Controversy

Journal of Basic and Applied Social Psychology:

“Accept no articles that include $P$-values”

Because...

• $P$-values: excuse for lower quality research
• $P$-value bar is too easy to pass
Supplement NHST with ...

• Sample size (small/big)
• Power
• Practical significance
• Effect size
• C.I., OR, RR, replication, ...
Example: Carpal Tunnel

Clinical Trial Success Rates

Surgery: $\frac{67}{73} = 92\%$

Splints: $\frac{60}{83} = 72\%$

$P$-Value $= 0.000896$

Surgery is better!
The rest of the story ...

95% CI: 8% < $p$ < 31%

Splint/surgery for failure:
• Odds Ratio
  4.28 (1.6 < OR < 11.2)
• Power = 0.80 (for 20% difference)
If analysis is important...

• Use several approaches:
  Parametric, nonparametric, bootstrap, simulation

• Several methods:
  H.T., C.I., Power, OR, RR

• Replicate

• Practical Significance
Educators
go beyond teaching...
Best Goals of Teaching Stats

• Critical thinking
• Technology
• Collaboration
• Speaking
Best way to improve course

Projects/Activities
Best Activity

First Class

Assign 200 coin tosses
• Half of class: actual tosses
• Half of class: fabricate tosses

Randomness!
Projects

Capstone Project:
• Groups of 3 or 4
• Oral presentation by all
• Computer printout
• Brief written report
Best Project

• Hip replacement costs / Whoppers
• Coke/Pepsi
• The story of Chris