Statistics 101 for POCT

What do the numbers mean?

Marcia L. Zucker, Ph.D., FAACC
ZIVD LLC
Objectives

- Interpret statistical analyses as reported by commercial programs
- Identify the statistical analyses relevant to the question being asked
- Critically evaluate data presented in package inserts for mis-used statistics
Statistics

- Definition of Statistics: The science of producing unreliable facts from reliable figures.
  - Evan Esar

- Be able to analyze statistics, which can be used to support or undercut almost any argument.
  - Marilyn vos Savant

- Statistic: a function of a set of observations from a random variable.
  - CLSI Harmonized Database
Method Validation

- A new POCT is to be implemented
  - Multiple replicates of controls run
  - Run side by side patient samples with current method
  - Data is:
    - Entered into EP Evaluator OR
    - Entered into Excel spreadsheet and analyzed using AnalysisToolPak or Analyse-It OR
    - Sent to manufacturer
  - Report returned with lots of statistics
    - Report may indicate pass/ fail to unknown specifications
    - Manufacturer rep explains it is all good

- How do I know it is OK?
Resources

- www.qimacros.com
- YouTube videos on performing analyses in Excel
- CLSI EP documents
  - The lab may have copies
- https://www.wikihow.com/Calculate-Precision
Some Basics

- Quantitative Methods
  - Statistics we use assume a normal distribution
Precision

- Measure of the variability of the system
  - How close are multiple replicates?
- Higher number of replicates allows better estimate of precision
- Outliers affect small numbers much more significantly
- Calculations assume a Normal Distribution
  - Frequently untrue assumption, but used anyway.
Precision
Precision – N affects result

N=10

N=20

N=100
Precision Statistics

- **Mean** – central tendency of the data
  - Peak of the bell curve (Average used in practice)

- **Median**
  - Value where 50% of samples are lower & 50% higher

- **Standard deviation (SD)** – measure of variability
  - Width of the bell curve
  - Relates to difference between individual results and the mean

- **Standard error (SE)** – measure of SD of the mean
  - Calculated from variance \((SD^2)\) & N

- **95% Confidence interval**
  - Estimate of “truth” from data collected
  - 95% probability that the “true” value is within the interval defined
## Statistics Calculated

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N=10</th>
<th>N=20</th>
<th>N=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.90</td>
<td>4.17</td>
<td>4.22</td>
</tr>
<tr>
<td>95% CI mean</td>
<td>3.65 – 4.14</td>
<td>4.00 – 4.35</td>
<td>4.14 – 4.27</td>
</tr>
<tr>
<td>SE</td>
<td>0.11</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>SD</td>
<td>0.34</td>
<td>0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>CV = ((\frac{\text{Mean}}{\text{SD}})) * 100</td>
<td>8.7%</td>
<td>9.1%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Median</td>
<td>3.99</td>
<td>4.21</td>
<td>4.25</td>
</tr>
<tr>
<td>95% CI median</td>
<td>3.45 – 4.20</td>
<td>4.01 – 4.44</td>
<td>4.19 – 4.29</td>
</tr>
</tbody>
</table>
Outlier Removal

N=8

N=98
# Outliers

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N=10</th>
<th>N=8</th>
<th>N=100</th>
<th>N=98</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.90</td>
<td>4.04</td>
<td>4.22</td>
<td>4.24</td>
</tr>
<tr>
<td>95% CI mean</td>
<td>3.65 – 4.14</td>
<td>3.92 – 4.16</td>
<td>4.14 – 4.27</td>
<td>4.20 – 4.28</td>
</tr>
<tr>
<td>SE</td>
<td>0.11</td>
<td>0.05</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>SD</td>
<td>0.34</td>
<td>0.14</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>CV = (\frac{\text{Mean}}{SD}) \ast 100</td>
<td>8.7%</td>
<td>3.5%</td>
<td>5.7%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Median</td>
<td>3.99</td>
<td>4.05</td>
<td>4.25</td>
<td>4.25</td>
</tr>
<tr>
<td>95% CI median</td>
<td>3.45 – 4.20</td>
<td>3.86 – 4.23</td>
<td>4.19 – 4.29</td>
<td>4.20 – 4.30</td>
</tr>
</tbody>
</table>
Statistics often look better at higher mean values
- If mean is 0.1 an SD of 0.05 is 50% CV
- If mean is 100 an SD of 5.0 is 5% CV

Evaluate values reported in inserts
- Should be near clinical decision points
- Required for newer products
- For older products expect to see more variability in end-user results
Accuracy

- Comparison to “truth”
  - Truth usually defined as current system
  - Truth a myth for many analytes
    - Notably coagulation, troponin I, other non-standardized analytes

- How close does POCT come to lab result
  - Correlation using patient samples
Accuracy
Correlation Graph

- **Data points**
  - Each split sample generates one point
  - Horizontal (X) axis is Lab (current system)
  - Vertical (Y) axis is point of care (new) device

- **Regression line**
  - Mathematical prediction of relationship between two devices
Results - Correlation Graph

Regression line
Data points
Regression equation

\[ y = 1.03x + 3.6 \]
\[ R = 0.965 \]
Correlation Graph

Regression equation

- 3 parts: \( Y = mX + b \) (\( y = 1.03x + 3.6 \))
  - \( Y \) = POC (new) result; \( X \) = lab (current) result
  - \( m \) = slope - perfect correlation \( m = 1.0 \)
  - \( b \) = intercept - perfect correlation \( b = 0.0 \)

- \( r \) value - correlation coefficient
  - NOT \( r^2 \)
  - Describes how much of the change in \( Y \) value is due to the change in the \( X \) value
  - \( r = 0.91 \) mean 91% correlation
Correlation - Is this good?

- Cannot judge
  - All values close to normal range
  - Nothing above 150
- Evaluate the axes when looking at correlation graphs

Glucose

\[ y = 1.08x + 5.53 \]

\[ R = 0.906 \]
Correlation – What to look for

- Assay range to 500, so spread seems OK
  - Isolated value drives correlation

- Original data set showed out of range values
  - These **must** be excluded before regression run

- Revised data has same issues as prior glucose results
Data need to span the clinically important range
  › Single extreme values should be omitted
  › Out of range values must be omitted

Correlates does **NOT** mean Match
Bias evaluation

- Difference plot
  - Bland Altman analysis
  - Plot either reference result or average of two methods as X
    - Reference result used when considered “truth”
      - e.g., POC electrolytes versus lab
    - Average used when “truth” is uncertain
      - e.g., ACT comparisons
  - Plot difference between two results as Y
Bland-Altman Plot

- Look for bias
  - Constant or variable?
  - Clinically significant?

![Graph showing Bland-Altman plot with data points and axes labeled: Mean INR, Current INR - New INR, Mean INRs]
Look for clinical differences

- Change of clinical decision limit can maintain current practice standards

$y = 1.09x - 7.53$

$R = 0.915$

Target Time change from 480 to 520 seconds
## Evaluate clinical differences

<table>
<thead>
<tr>
<th>LAB</th>
<th>POC A</th>
<th>&gt;0.1</th>
<th>&lt;0.1</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;0.1</td>
<td>28</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;0.1</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td>93%</td>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>93%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LAB</th>
<th>POC B</th>
<th>&gt;0.1</th>
<th>&lt;0.1</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;0.1</td>
<td>18</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;0.1</td>
<td>12</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td>60%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70%</td>
</tr>
</tbody>
</table>
Sensitivity & Specificity

- **Sensitivity**
  - ability of an assay to identify patients with a specific condition \((true\ positives)\)

- **Specificity**
  - ability of an assay to identify patients without a specific condition \((true\ negatives)\)

- **Positive predictive value**
  - likelihood that a patient with a positive result (or above the cut-off) truly has the condition

- **Negative predictive value**
  - likelihood that a patient with a negative result (or below the cut-off) is truly normal
## 2 x 2 Table

<table>
<thead>
<tr>
<th>New System Result</th>
<th>&quot;True&quot; Result</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Concordance</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>True positive (TP)</td>
<td>False positive (FP)</td>
<td>Positive predictive value (PPV)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>False negative (FN)</td>
<td>True negative (TN)</td>
<td>Negative predictive value (NPV)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Concordance</th>
</tr>
</thead>
</table>
| \( Sensitivity = \frac{TP}{TP + FN} \) | \( Specificity = \frac{TN}{TN + FP} \) | \( Concordance = \frac{TP + TN}{Total \ Sample \ Number} \)
Qualitative tests always include sensitivity and specificity claims

- Older products have limited clinical data
  - Only spiked samples evaluated
  - Only frozen clinical samples evaluated
  - Too few samples evaluated

- Newer products will include confidence intervals
  - Do not want test where CI spans 50% (coin toss)
**Probability (p-value)**

- **Paired t-test**
  - Compare the difference between paired samples
  - Null hypothesis is tested
    - mean difference is zero
  - Means of populations compared
  - Assume normal distribution; equal variance

- **ANOVA (Analysis of Variance)**
  - Compare means of groups of measurement
  - Null hypothesis is tested
    - means of the measured variables are the same
  - Variances of populations compared
  - Assume normal distribution; equal variance
### p-value

<table>
<thead>
<tr>
<th>P-VALUE</th>
<th>INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>HIGHLY SIGNIFICANT</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>0.049</td>
<td>OH CRAP. REDO CALCULATIONS.</td>
</tr>
<tr>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>0.051</td>
<td>ON THE EDGE OF SIGNIFICANCE</td>
</tr>
<tr>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P&lt;0.10 LEVEL</td>
</tr>
<tr>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>0.099</td>
<td>HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS</td>
</tr>
<tr>
<td>≥0.1</td>
<td></td>
</tr>
</tbody>
</table>
p-value

- Statistical significance can be defined at multiple levels
- For diagnostics, generally defined as $p \leq 0.05$
  - 95% confidence
  - $\sim \pm 2$ SD from mean
Interpreting $p$

- If viewing results of analysis:
  - $p \leq 0.05$: two samples are different
  - $0.05 < p < 0.1$: trend towards difference
  - $p > 0.1$: two samples are the same
What else?

- There are as many ways to crunch data as there are people to do it.
- Keep in mind what you are looking for:
  - **Clinical utility**
    - statistical difference may not matter
- Understand what you want BEFORE you collect the data:
  - Define studies by the information you want
There are three kinds of lies: lies, damned lies and statistics.

- Benjamin Disraeli

Torture numbers, and they’ll confess to anything.

- Gregg Easterbrook
QUESTIONS?

Marcia L. Zucker, Ph.D., FAACC
ZIVD, LLC
Mlzucker.zivd@gmail.com