

PathMD™: Board Review Letter

Author: Philip Ferguson, M.D.

Statistics - Part 1

Volume 1, Number 19

1. Of the series of numbers on the website under question #1, calculate the mode, median, and mean, respectively.

C. 3,4,4.33

Answer: C. Mode = most frequent number in a series (3), Median = the midpoint of a series (4), and Mean = average of a series ($39/9 = 4.33$). *Test taking strategy.* On questions where they may be asking you to calculate, look at the answers first. Notice in this case you don't even have to average the numbers. Really this question is only asking to differentiate mode and median of a series. (Henry, JB & Dufour, DR)

2. A series of data for a new clinical lab test has a Gaussian distribution. What percent of the data will be more than 3 standard deviations above the mean?

D. 0.15%

Answer: D. 3 standard deviations of data in a Gaussian distribution encompasses 99.7% of the test data. The remaining 0.3% is distributed at the two ends of the distribution, and only half (0.15%) will be above the mean. *Test taking strategy.* Remember that 1 standard deviation contains 68% of the data; 2 standard deviations \rightarrow 95.5% of the data; and 3 standard deviations \rightarrow 99.7% of the data. (Henry, JB & Dufour, DR)

3. A new instrument is installed for CBCs, and a correlation test is performed. The reference method is plotted on the X-axis and the new method on the Y-axis for multiple samples tested. Linear regression analysis is performed ($y=mx + b$). Which of the following represents the correlation coefficient, r.

C. Calculation of how well the points fit the line

Answer: C. The correlation coefficient represents how well data points fit a line. The Y-intercept represents constant bias, and the slope of the line represents proportional bias (e.g. the test always runs 5% higher than the previous test). (Henry, JB & Dufour, DR)

4. Which of the following correlation coefficients (r) represents the minimal number you would like to see when validating a new test?

C. 0.97

Answer: C. While an r-value of >0.99 is usually achieved when setting up a new test, 0.97 is the lower limit of an acceptable coefficient. An r-value of greater than 0.3 indicates the two methods are affected by the same variables, but this would be horrible for two tests that are supposed to be the same. (Henry, JB & Dufour, DR)

5. Of the following answers, which is the best tool to detect bias between two different methods?

E. Bland-Altman plot

Answer: E. The Bland-Altman plot is the most reliable. The difference between the new and old method results are plotted on the y-axis (usually as a percentage), and the reference result on the x-axis. *Test taking strategy.* The linear regression analysis can also show bias between two methods. The key to answer is realizing the author is testing whether the test taker recognizes what a Bland-Altman plot is (a.k.a. difference or bias plot). The F-test assesses the reproducibility of a new method, which has a Gaussian distribution of data. (Dufour, DR)

6. For a test with a Gaussian distribution, what is the minimum number of representative samples needed to determine a reference range?

B. 120

Answer: B. 20 is the minimum number of “healthy person” samples needed to validate a previously established reference range (if no more than 2 values fall outside the reference limits). 120 is the minimum number of representative samples needed to create a reference range (if it represents a Gaussian distribution). 200 is the minimum number of representative samples needed to create a reference range, which does not have a Gaussian distribution. (Dufour, DR)

7. For a test without a Gaussian distribution, what is the minimal number of representative samples needed to establish a reference range?

D. 200

Answer: D. 20 is the minimum number of “healthy person” samples needed to validate a previously established reference range (if no more than 2 values fall outside the reference limits). 120 is the minimum number of representative samples needed to create a reference range (if it represents a Gaussian distribution). 200 is the minimum number of representative samples needed to create a reference range, which does not have a Gaussian distribution. (Dufour, DR)

8. To validate a previously established reference range, what is the minimal number of healthy person samples needed?

A. 20

Answer: A. 20 is the minimum number of “healthy person” samples needed to validate a previously established reference range (if no more than 2 values fall outside the reference limits). 120 is the minimum number of representative samples needed to create a reference range (if it represents a Gaussian distribution). 200 is the minimum number of representative samples needed to create a reference range, which does not have a Gaussian distribution. (Dufour, DR)

PathMD™: Board Review Letter

Author: Philip Ferguson, M.D.

Statistics - Part 1

Volume 1, Number 19

9. If 20 tests were performed on a “healthy” individual, what is the likelihood all of the results will be within the reference range?

D. 36%

Answer: D. This question points to the fact that reference ranges by convention exclude 5% of the results (95% confidence interval), which is also approximately 2 standard deviations (95.5%). Considering this, the likelihood of n tests falling within the reference range is $= 0.95^n$. *Test taking strategy.* This is an easy question if you know the equation and have a scientific calculator. Most will not have either. Let’s think of this question more simply. If you consider that reference ranges represent the 95% confidence interval of a distribution, then it will be likely that 1 in 20 tests (i.e. 5%) will be “abnormal” in a “healthy” person. Answers A through C are either close to or greater than 50%; leaving answer D (36%) as the most likely answer. It is recommended that “none of the above” not be used unless all of the other answers can be confidently excluded, or the correct answer is known and not given. (Henry, JB & Dufour, DR)

10. The coefficient of variation best represents which of the following:

B. Precision

Answer: B. The coefficient of variation = standard deviation divided by the mean (x 100) to be expressed as a percentage. The smaller the standard deviation means the more precise the test. Accuracy is usually expressed as bias, which can be an absolute of percentage (think correlation plots). Efficiency is term used in relative operating characteristic curves. (Henry, JB & Dufour, DR)

References:

Clinical Diagnosis and Management by Laboratory Methods. Henry, JB, et al. 19th edition. 1996. pp. 125-135.

The Osler Institute: Statistics and Quality Control. Dufour, DR. 2004.

Notes for question set:¹

¹ PathMD strives for the highest quality and accuracy. However, the *PathMD: Board Review Letter* is for review purposes and not meant for clinical decision making. It should not be used in place of review of primary reference texts and the current medical literature. If inaccuracies are identified, please notify us so that a correction may be published. (info@PathMD.com)