



Clinical Chemistry Trainee Council

Pearls of Laboratory Medicine

www.traineecouncil.org

“QC Design: Things You Need to Know” Series

TITLE: Performance Statistics

PRESENTER: Lakshmi Kuchipudi

Slide 1:

Hello, my name is Lakshmi Kuchipudi. I am a senior scientist at Bio-Rad Laboratories. Welcome to this Pearl of Laboratory Medicine on “Performance Statistics.” This is the 3rd Pearl in the “QC design: Things you need to know” series. In order to evaluate the quality of a testing system, i.e. to make sure the process is functioning as intended, we need to compute the testing system’s stable performance statistics. In the laboratory, we use statistics to monitor and verify testing system performance, interpret the results of laboratory tests, and evaluate quality control procedures. In this pearl, I will give a brief introduction to statistical distributions and computing performance metrics.

Slide 2:

Statistical distributions can be used to predict the probability characteristics of certain applicable real populations. There are several different statistical distributions. The one shown here is a normal distribution. It has a nice bell shaped curve. The mean of the normal distribution is at the center of the curve. The normal curve is symmetric around this mean value. The variance of the distribution measures the spread of the data making up the distribution with respect to the mean value. The true mean (μ) and true variance (σ) of the distribution are parameters which can be estimated.

Slide 3:

The estimates of the true mean and variance based on a sample of observations are referred to as the sample mean (\bar{x}) and sample variance (s^2). Given a sample size N and sample data x_1, x_2, \dots, x_N , the sample mean is an average of all the data points x_1 to x_N which is computed by adding up all N data points and dividing by N . The sample variance is the average of the squared deviation of each data point from the sample mean divided by $N-1$. Variance is always positive. A low variance indicates the data points tend to be very close to the mean; a high variance indicates the data points are spread out over a large range of values.

Slide 4:

Standard deviation is defined as the square root of the variance. Just like variance, standard deviation measures the dispersion of values making up the distribution from the mean and is always positive. The sample standard deviation (s) is computed as the square root of the sample variance.

Slide 5:

The coefficient of variation is a unit-less quantity defined as the standard deviation of a distribution divided by the mean of the distribution. You can think of it as noise to signal ratio. We can measure it in % by multiplying the ratio by 100. It is a useful statistic for comparing the degree of variation from one sample to another, even if the means are drastically different from each other. It's frequently used in the laboratory instead of standard deviation or variance. In the laboratory, the lower the CV, the better your process precision.

Slide 6:

Here's an example on how to compute these estimates. We have a sample size of 5. The mean of the sample is the average of the all the points in the sample. In this example, the mean of this sample is 4. The sample variance is the sum of the squared deviations of each of the data point in the sample from its mean divided by the sample size minus 1. The sample variance is 2.5. The sample standard deviation is the square root of the variance and is equal to 1.58. The sample CV in % is the sample standard deviation divided by the sample mean times one hundred which is equal to 39.5%.

Slide 7:

Let's say we want to compute an instrument's Mean, SD, and CV values on a daily basis. We run one level of liquid quality controls 5 times a day. We have the control data for day 1 and day 2. Mean for Day 1 is 3.6 and Day 2 is 4.4. SD for Day 1 is 1.14 and Day 2 is 1.34. Which of these days has the correct Mean and SD values?

In reality, there is only one correct value which is the truth. In our case, we don't know the correct value. As a result, we try to estimate these correct values. We have data from Day 1 and Day 2; we estimated the mean, SD, and CV from these data. Now, which one is the correct value? Are any of these values correct? If not, how can we estimate the correct values?

Slide 8:

When estimating the distribution's parameters mean, SD, CV, the sample size is very important. The bigger the sample, the more precisely you will estimate the correct parameter value. To illustrate this, let's see how well we estimate the correct parameters when the true mean is 50, true standard deviation is 10, and true coefficient of variation is 20%. In reality, we wouldn't know these true values. Here I do because I am simulating data from these true values. We estimate the distribution parameters with sample sizes of 10, 20, 50, and 100. The x-axis reflects the different sample sizes. The first panel shows sample estimates of the mean, the second panel shows sample estimates of the standard deviation, and the third panel shows sample estimates of the coefficient of variation. The solid horizontal lines are drawn at the true parameter values.

Let's look at the first panel which is estimating the mean. The red solid horizontal line is representing the true mean 50. Each of the red circles here are the estimated means computed from a different sample. The 1st set of vertical red circles representing the estimated means are computed from six different samples of size 10. You can see the estimated means; the red circles are different for different samples and are spread out.

The 2nd set of vertical red circles representing the estimated means are computed from six different samples of size 20. You can see the 6 estimated means from 6 different samples are precise and closer to the true mean 50. The 3rd and 4th set of vertical red circles representing the estimated means are computed from six different samples of size 50 and 100. You can see the 6 estimated means in this case are more precise and closer to the true value.

If you look at the 2nd and 3rd panel which estimates the SD's and CV's, respectively, you can see the higher the sample size, the more precise the estimates. A small sample size has a lot of variation in the estimates. Estimates of sample size of 10 are more spread out compared to the other sample sizes. The estimates from sample size 100 have less spread and are closer to the true value. The bigger the sample sizes, the closer the estimates to the true values.

Slide 9:

The estimates we computed (sample mean, sample SD, sample CV) all characterize the state of a testing process. Another characteristic of interest for an instrument is bias. An instrument's bias is the difference between the instrument's mean value and the specimen's target value (the reference mean). If you want to compute the bias in %, it's the absolute difference between the instrument mean and reference mean divided by the reference mean multiplied by 100. Estimating the bias in % may give you a better appreciation of the magnitude of the bias. For example, if an instrument gives a mean value of 5 units and the reference mean is 4 units, then the absolute bias is 1 unit and the % bias is 25%, which tells you the instrument has a high bias.

Slide 10: References**Slide 11: References****Slide 12: Disclosures****Slide 13: Thank You from www.TraineeCouncil.org**

Thank you for joining me on this Pearl of Laboratory Medicine on "Performance Statistics" from the "QC Design: Things You Need to Know" series. I am Lakshmi Kuchipudi.