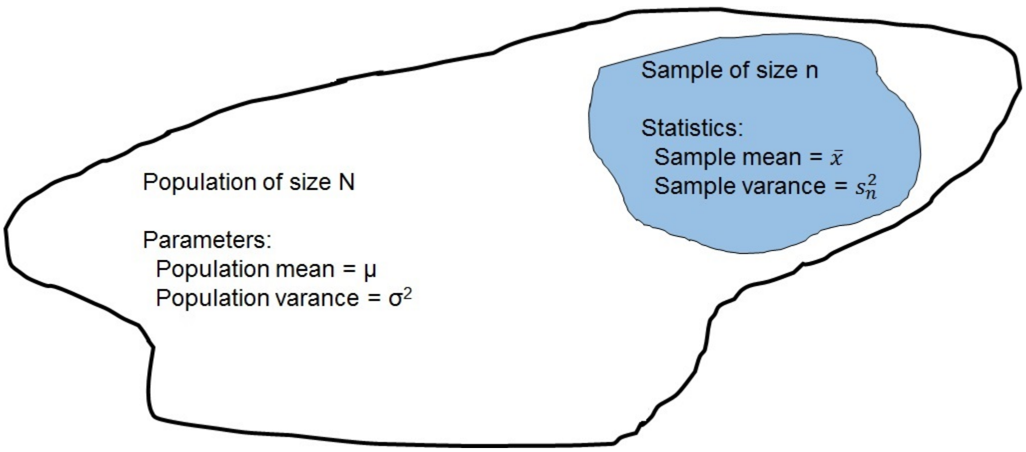
Point of this document is to give a compact summary of statistics usage and nomenclature as it relates to image sensors.

Descriptive statistics



Population mean and variance

Population is a group of phenomena with something in common. Parameter is characteristic of a population.

Population mean $= \mu = \frac{\sum_{i=1}^{N} x_i}{N}$

Population variance $= \sigma^2 = \frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}$

Sample mean and variance

Sample is randomly-drawn subset of manageable size used to make inferences about population. Statistic is characteristic of a sample.

Sample mean $= \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Sample variance biased $= s_n^2 = \frac{\sum_{i=1}^{n} (x_i-\overline{x})^2}{n}$

Sample variance unbiased (usually use this one)$= s_{n-1}^2 = \frac{\sum_{i=1}^{n} (x_i-\overline{x})^2}{n-1}$

Mean and variance notes

For intuition, variance is the average distance of values from mean. Use square to treat positive and negative distances the same.

If mean has units of DN, variance has units of DN^2, so often people will use standard deviation (square root of variance) which has units of DN.

Alternative formulations of variance for faster calculation

Population variance $= \sigma^2 = \frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}$

$= \frac{\sum_{i=1}^{N} (x_i^2-2x_i\mu+\mu^2)}{N}$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} + \frac{\sum_{i=1}^{N}(-2x_i\mu)}{N} + \frac{\sum_{i=1}^{N}(\mu^2)}{N}$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - 2\mu\frac{\sum_{i=1}^{N} x_i}{N} + \mu^2\frac{\sum_{i=1}^{N}(1)}{N}$$

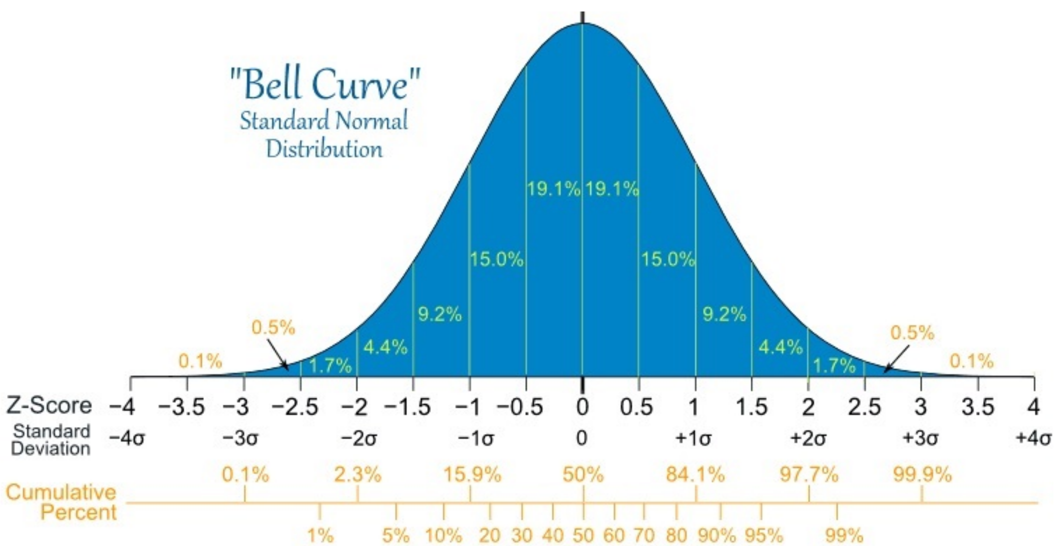
$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - 2\mu\frac{\sum_{i=1}^{N} x_i}{N} + \mu^2$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - 2\mu^2 + \mu^2$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2$$

Normal distribution and empirical rules



Gaussian normal distribution $= P(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$

One sigma captures 68.2% of values
Two sigma captures 95.4% of values
Three sigma captures 99.7% of values

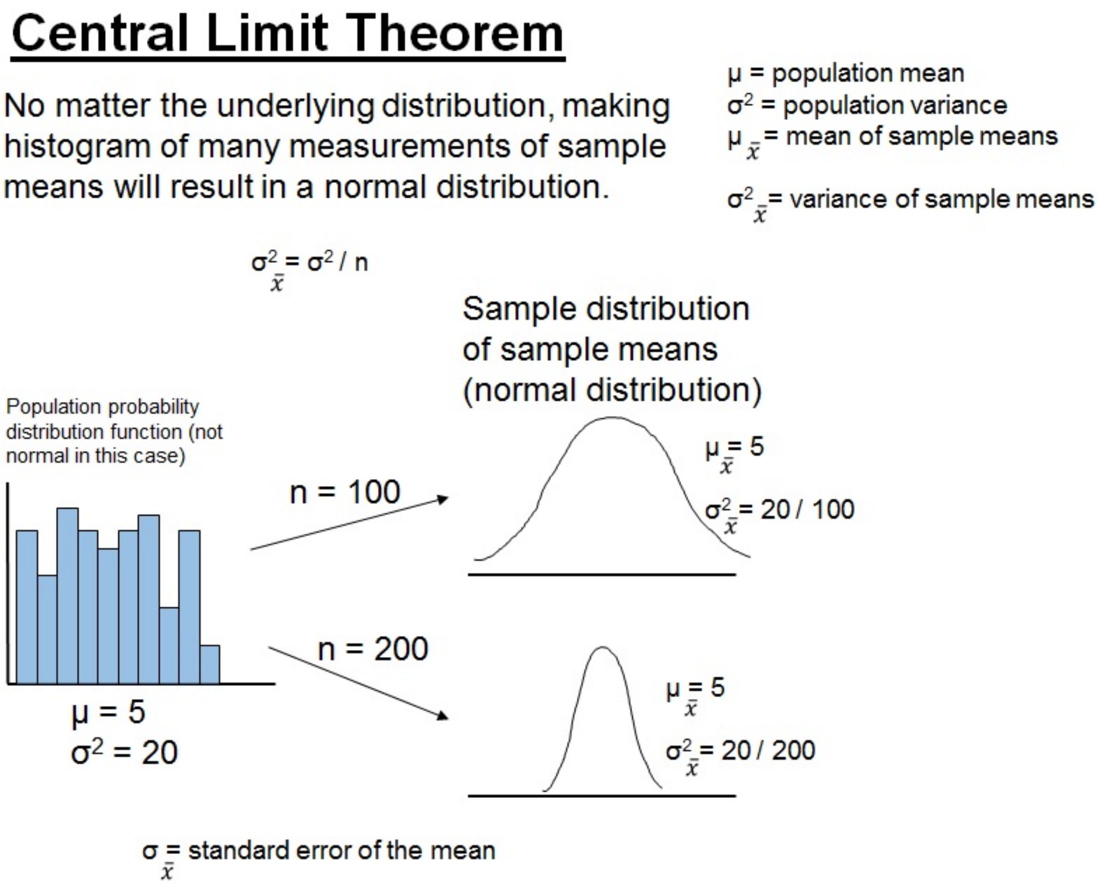
Law of large numbers

Average of results obtained from large number of trials should be close to expected value. More to do with probability than statistics.

Central limit theorem

Histogram of large number of independent, identically distributed values will be approximately normal regardless of underlying distribution. Underlying distribution can be continuous or discrete.

The histogram is the sample distribution of sample mean. Each observation of sample mean is one count for the sample distribution.



Standard error of mean and confidence interval

Larger the sample size (n), the smaller the variance of sample distribution of sample mean. Intuitively this makes sense. The larger your sample, the more accurate and tightly grouped will be the average values of the sample.

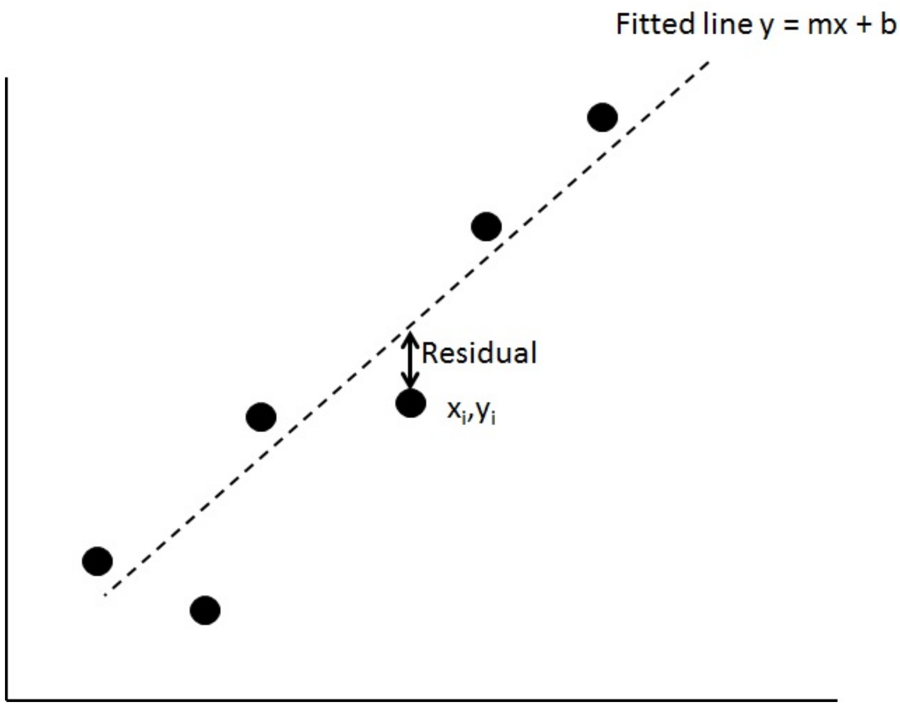
Variance of sampling distribution of sample mean equals variance of original distribution divided by sample size $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

Sometimes data are plotted with error bars (aka confidence intervals). These are often equal to twice the standard error. Meaning of confidence interval should be specified when used.

Regression

Fitting line to data

Best line fit (regression line) has form $y = mx + b$. The "best" fit is the line that gives least-squares. It minimizes sum of squared residuals.



More formally, assume n points of (x_i, y_i) and fit has form $\hat{y}_i = m\hat{x}_i + b$

$$\text{Sum of squares} = Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - m\hat{x}_i + b)^2$$

RSS, Tolerance Analysis

RSS = root sum squares
Assuming independent random variables with normal distributions, then total variance of all variables working in concert is: $\sigma_{total}^2 = \sum_{i=1}^{n} \sigma_i^2$

Intuitively, consider that you may consider independent processes to be orthogonal to each other, similar to two sides of a right triangle with size a and size b. The size of the hypotenuse $= \sqrt{a^2 + b^2}$, which is of similar form to standard deviation σ_{total}.

This applies to independent noise sources for image sensors such as read noise, shot noise, and fixed pattern noise.

It also applies to tolerance analysis (tolerance stackup) in manufacturing. For a manufacturing process, you may be given mean and standard deviation of the final product dimensioning. For something like 6-sigma analysis, you would calculate RSS values considering $+/-3\sigma$ from mean.

Poisson Distribution

For Poisson process, all events are independent. In such cases, mean equals the variance for the distribution. This is true of photon arrival.

When looking at data from Poisson process, be careful. Even though there may be clumping, grouping, filaments, etc there is no pattern. Your brain is hard-wired to look for patterns. For instance, to first approximation, there is no pattern in the starts. Yet people see constellations and attribute meaning to them.