Design of Experiments and Statistical Process Control

Basics

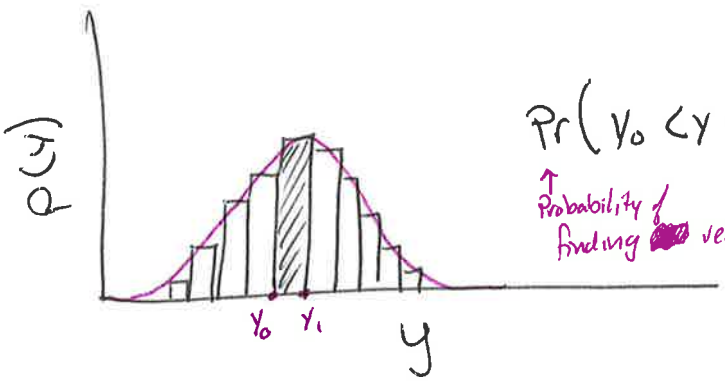
1. Error → variation in a measurement or process that is often unavoidable

   → must plan for in D.O.E. / data Analysis

2. Distributions → collection of data points representing "population"

   represented by a line diagram

   → probability distribution



$$Pr(y_0 < y < y_1) = \int_{y_0}^{y_1} p(y)\,dy$$

Probability of finding [redacted] result

3. Statistics -

   distinguish between sample and population quantities.

   sample ↑ what has been measured, (n) a subset of the population

   population ↑ everything that exists (N)

$$\eta = \sum y/N \qquad \bar{y} = \sum \frac{y}{n}$$

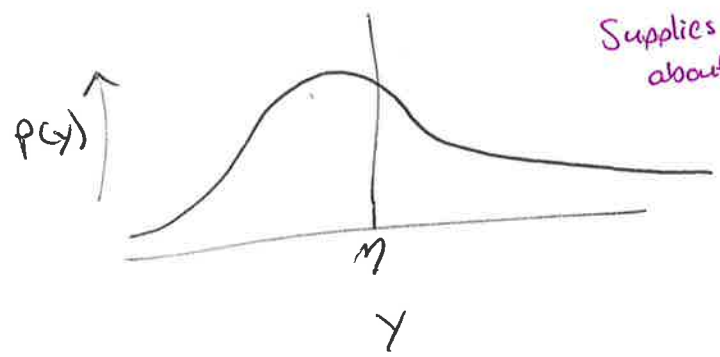
population mean — sample average

if hypothesis of random sampling holds, then $\bar{y}$ represents η

(must watch out for autocorrelation of measurements i.e. consider daily weather!)

Variance

We defined population mean η which defines the first moment of the distribution y



Supplies information about the location of the distribution

Variance describes spread of the population, measures how far, on average, the data lie from the mean.

population variance

$$\sigma^2 = \frac{\sum (y-\eta)^2}{N}$$

population std. dev.

$$\sigma = \sqrt{\frac{\sum (y-\eta)^2}{N}}$$

for a sample:

Variance

$$s^2 = \frac{\sum (y-\bar{y})^2}{n-1}$$

Std. dev

$$s = \sqrt{s^2}$$

$n-1$ here is used to account for loss in degrees of freedom associated with using sample $\bar{y}$ to estimate η, instead of knowing the value.

"Bessel's correction"

$n-1$ degrees of freedom.

$$\frac{\sigma}{\eta}, \frac{s}{\bar{y}} \equiv \text{Coefficient of Variation}$$ (or percent error)

$$\frac{\bar{y}}{s} \equiv \text{signal to noise ratio}$$

Standard Error of Average (or of Effect)

$$S.E. = \frac{s}{\sqrt{n}}$$

error of the mean, not error of an individual measurement.

used to calculate confidence intervals of the mean.

point of possible confusion...

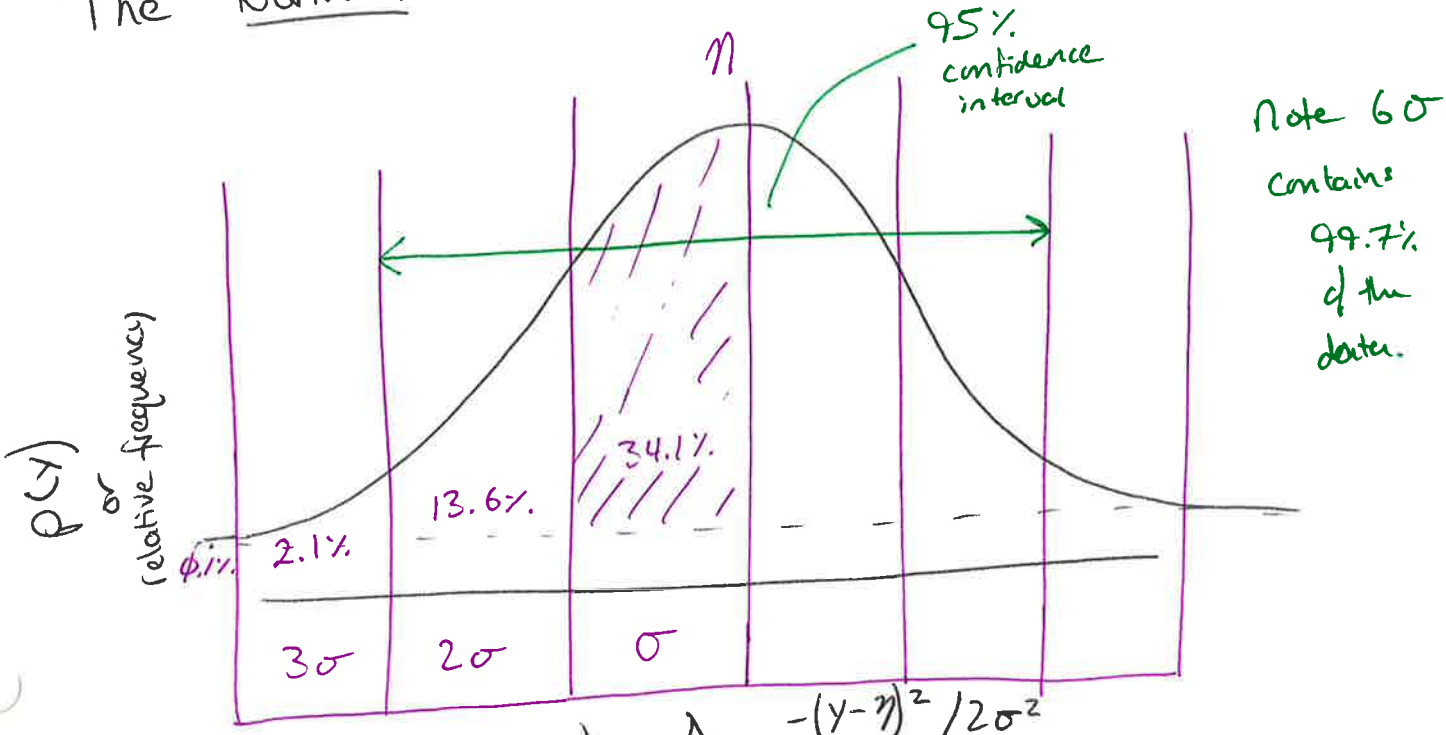
S.E. describes how well sample mean represents population mean

S represents the actual <u>spread</u> (or error) in the measurements

Relative Standard Error

$$RSE \equiv SE/\bar{y}$$

The <u>Normal</u> Distribution : Gaussian



$$p(y) = \frac{A}{\sigma} e^{-(y-\eta)^2/2\sigma^2}$$

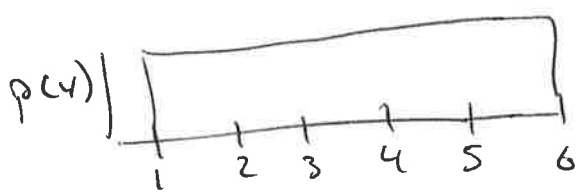
Why do many error distributions tend to be normal (gaussian)?

Central Limit Effect

* random error is the result of the sum of errors from non-correlated variables.

Consider throwing dice

1 die → non gaussian



← equal probability

↓ average value of multiple die tend toward normal distribution

Note: the tending toward normality happens NO MATTER WHAT the initial p(y) looks like!

3a.

η = mean = 3.5

e_n = deviation from average = error

$e = \sum \frac{e_i}{n}$

total error is linear function of uncorrelated errors

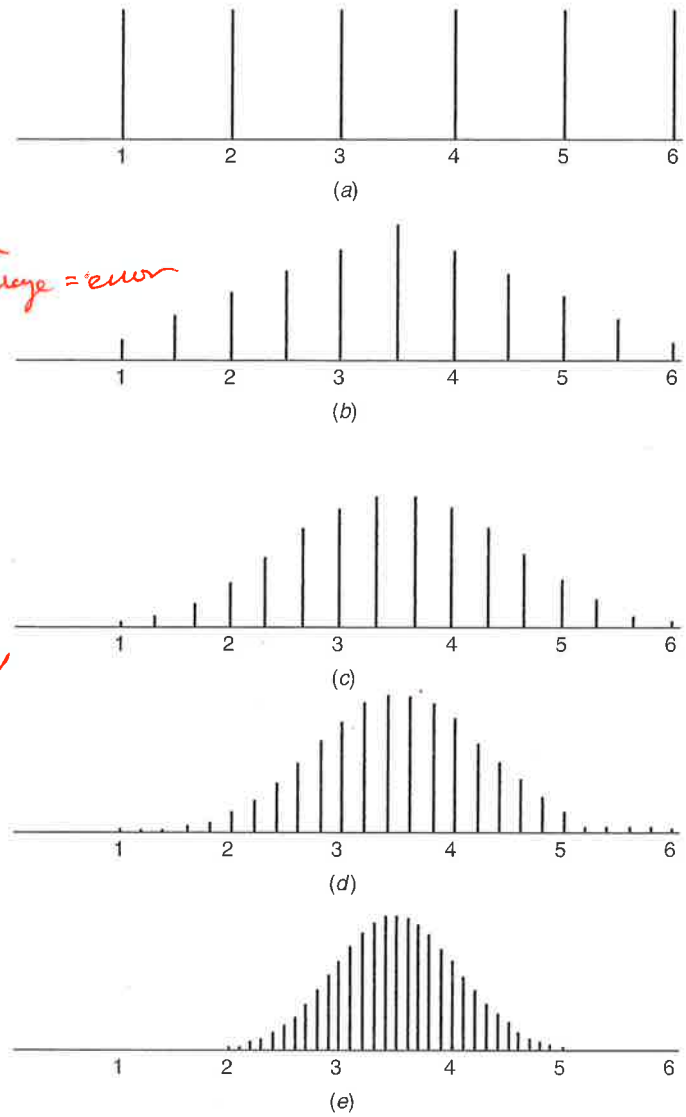


Figure 2.6. Distribution of averages scores from throwing various numbers of dice: (*a*) 1, (*b*) 2, (*c*) 3, (*d*) 5, and (*e*) 10 dice.

normality. You should of course be continually on the look-out and check for *gross* violations of this and all other assumptions.

Characterizing the Normal Distribution

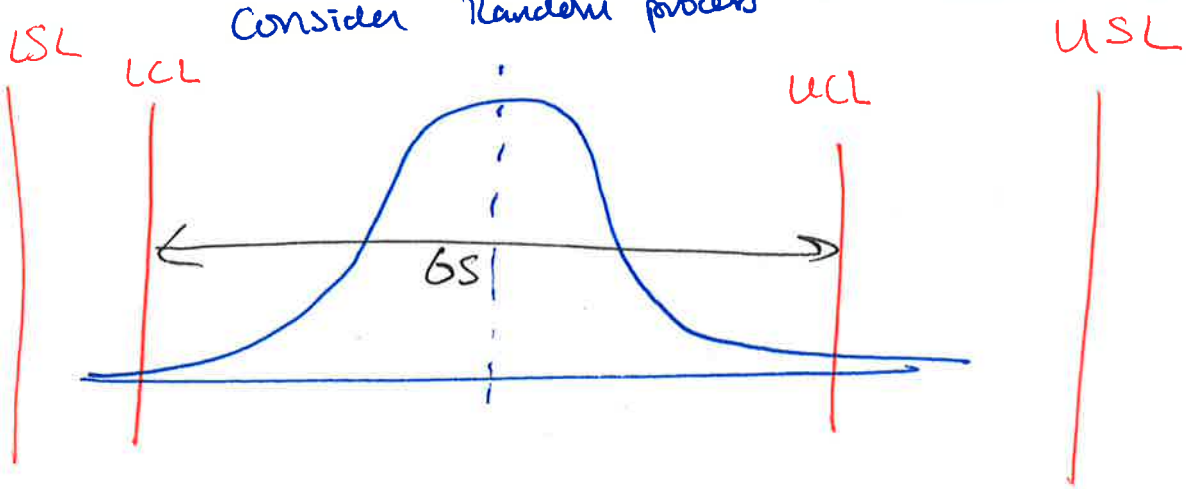
Once the mean η and the variance σ^2 of a normal distribution are given, the entire distribution is characterized. The notation $N(\eta, \sigma^2)$ is often used to indicate

4b

Statistical Process Control - SPC

Statistical process control and the Schwehart or western Electric Rules.

Consider Random process (we know from Central limit theorem...)



Spec limits → determine whether or not you have to get rid of product.

Control limits → based on statistics typically ±3σ or all the data

---

Western electric rules indicate whether or not a process is in statistical control (Normal) or whether there is a problem.

(Review Western Electric Rules on the ppt slides.)

$$\text{Process Capability Ratio} = C_p = \frac{USL - LSL}{6s}$$

sample std dev

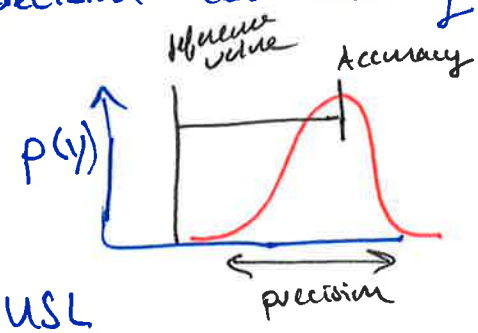
(process variation relative to specifications)

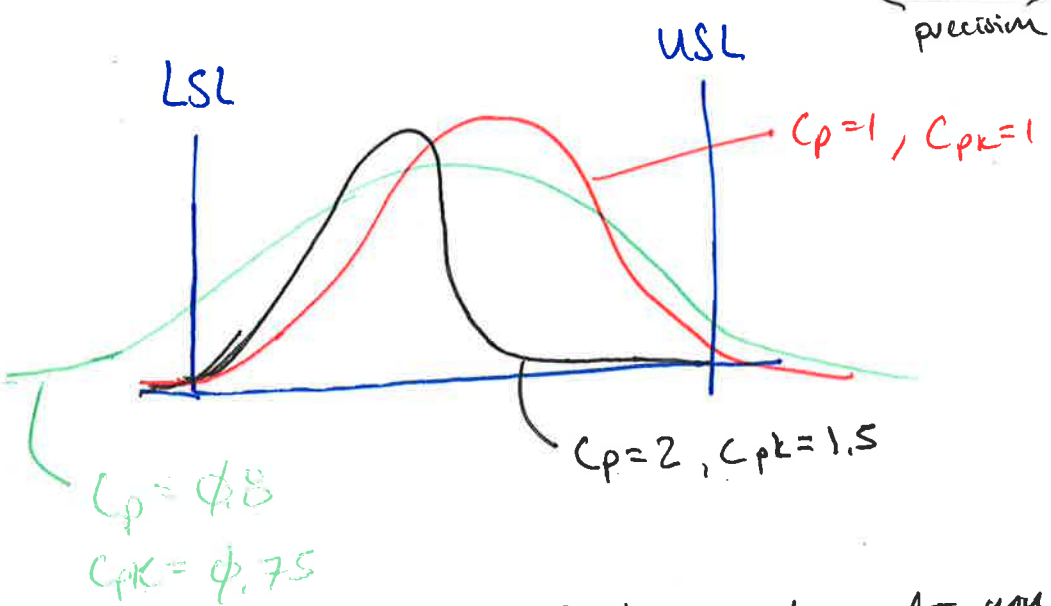
larger = better

$$\text{Process Capability Index} = \frac{\min\left[(USL - \bar{y}) \text{ or } (\bar{y} - LSL)\right]}{3s}$$

(process capability Index)
measures process variation
and centering relative to spec.

Sort of like precision vs. accuracy





What Cp / Cpk values do you want to achieve good performance?

(good Reference: Introduction to Statistical Quality control, Douglas C. Montgomery)