Application of Simulation to Sample Size Calculations and Teaching Statistical Concepts

Fall Technical Conference
Statistics: Powering a Revolution in Quality Improvement
October 4 - 6, 2017 | Philadelphia, PA

Presented by:
Lou Johnson
lou@jmcdataexperts.com

JMC Data Experts
Question: Which Door Should I Take?

Let's Make a Deal:
Monty Knows

1. Contestant choses a door.
2. Monty opens one door with a goat.
3. Monty offers the contestant the remaining door or stay with their door.

One Car and Two Goats
1. Switch to the other door ?
2. Stay with your door?
3. It doesn’t matter?

©JMC 2017
The Right Answer Explained (?)

Since you seem to enjoy coming straight to the point, I’ll do the same. You blew it! Let me explain. If one door is shown to be a loser, that information changes the probability of either remaining choice, neither of which has any reason to be more likely, to 1/2. As a professional mathematician, I’m very concerned with the general public’s lack of mathematical skills. Please help by confessing your error and in the future being more careful.

Robert Sachs, Ph.D.
George Mason University

You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I’ll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don’t need the world’s highest IQ propagating more. Shame!

Scott Smith, Ph.D.
University of Florida
Could a Quick Simulation Answer the Question Definitively?

GMacro
Goat3
Name C1 "Door"
Name C2 "Car"
Name c3 "Guess"
Name c4 "Wins"
Do k1 = 1 : 1000
Sample 1 'Door' 'Car'
Sample 1 'Door' 'Guess'
Let c4[k1] = 'Car' = 'Guess'
Enddo
Name k2 "Stay Wins"
Name k3 "Switch Wins"
Let k2 = Sum(c4)
Let k3 = 1000 - k2
Print k2 k3
endmacro

https://math.ucsd.edu/~crypto/Monty/montybg.html
Agenda

- The Monty Hall Paradox
- Confidence Intervals for the Mean and Standard Deviation
  - Percent variability for the mean and standard deviation
  - Impact on process capability
  - Effect on MSA measures of capability and measuring repeatability
- Sampling Plans for Expanded Gage R & R
  - Simulation results for sources of variation
  - The best sampling plan when variation sources cannot be estimated
- Analysis of Variance Study when Factors are Nested
  - Syringe needle strength – where is the variability coming from?
  - Study results for the complete study
  - Simulation results on the effect of sample size on study results
- Cleaning Process Factorial with a Non-normal Covariate
  - Factorial design for a cleaning process with a non-normal covariate
  - Power and sample size calculations
  - Simulation for the effect of sample size on averaging the covariate
- Questions & Discussion
Non-disclosure Statement

All industrial experiments, results and scenarios are based on the authors’ actual experiences. Data units, variable names, etc have been changed for demonstration purposes only.
The Difference between a Statistician and an Engineer

The engineer takes three measurements (5.5, 6.1, 6.4), takes the average and concludes that the thickness of the fabric is 6.0 /1000”, give or take.

The statistician concludes that the engineer took three samples from a population that is likely normally distributed, centered at 6.0 /1000” and has a standard deviation of 1 /1000”.

The difference is critical. The second interpretation understands that future samples from the population will be different even if nothing has changed.
How do We Teach this Important Concept of Variability

Classroom Exercise

1. Create a population with a given mean and std deviation.
2. Take a random sample of 8 measurements.
3. Calculate the descriptive statistics.

Repeat sampling 10,000 times. What is the distribution of the estimates of the mean?
Conclusions

1. 95% of the estimates of the mean fall between 5.3 and 6.7.
2. Your estimate from 8 samples will likely be within ±12% of the right answer.
3. The confidence interval for the mean is
   \[ \text{Lower limit} = M - Z_{.95} \sigma_M \]
   \[ \text{Upper limit} = M + Z_{.95} \sigma_M \]
   \[ (6 - 1.96 \times .3536, 6, 6 + 1.96 \times .3536) \]
   \[ (5.31, 6, 6.69) \]

What about the standard deviation?
Conclusions

1. 95% of the estimates of the standard deviation between .5 and 1.5.
2. Your estimate from 8 samples will likely be within + / - 50% of the right answer.

How does this affect our perspective on process capability estimates and sample size? How does this affect our perspective on measurement system capability?
Source of Variation in Process Capability Estimates (Cpk)

Cpk = \( \min \left[ \frac{(\text{Upper Spec} - \text{Avg}) \text{ or } (\text{Avg} - \text{Lower Spec})}{(3 \times \text{Standard Deviation})} \right] \)

This is not the problem

This is the main source of variation

To estimate the mean, 5 – 10 samples might be acceptable. Because of the variability in the standard deviation, 30 samples or more are used to estimate Cpk.
Estimating Measurement System Capability

\[
\sigma_{Total}^2 = \sigma_P^2 + \left( \left( \sigma_O^2 + \sigma_{PO}^2 \right) + \sigma_\varepsilon^2 \right)
\]

\[
\sigma_{Total}^2 = \sigma_P^2 + \left( \sigma_{repeatability}^2 + \sigma_{reproducibility}^2 \right)
\]

\[
\sigma_{Total}^2 = \sigma_P^2 + \sigma_{measurement}^2
\]

\[
\% Study Variation = \frac{\sigma_{measurement}}{\sigma_{total}} \times 100\%
\]

\[
\% Tolerance = \frac{6 \times \sigma_{measurement}}{USL - LSL} \times 100\%
\]

\[
\% Process Variation = \frac{\sigma_{measurement}}{\sigma_{Historical}} \times 100\%
\]
Expanded Gage R & R Overview

**Overall Process Variation**

- Part-to-Part Variation
  - Within Gage Variation
  - Gage to Gage Variation
    - Part * Gage Interaction
    - Variation due to Measurement Procedure
  - Measurement System Variation
    - Operator * Gage Interaction
    - Operator * Part Interaction

- Measurement System Variation
  - Variation due to Measurement Procedure

- (Repeatability)
  - Oper-to-Oper Variation
  - Operator * Part Interaction

- (Reproducibility)
  - Operator * Gage Interaction
  - Operator * Part Interaction

©JMC 2017
Sampling Plan for an Expanded Gage R & R Study

Standard Gage Study:

10 Parts X 3 Operators X 2 Repeats


Expanded X 3 Gages = 180 Measurements

\[ \sigma_{Total}^2 = \sigma_P^2 + \left( \sigma_O^2 + \sigma_{OG}^2 + \sigma_{PO}^2 \right) + \left( \sigma_G^2 + \sigma_{PG}^2 + \sigma_{\varepsilon}^2 \right) \]

\[ \sigma_{Total}^2 = \sigma_P^2 + \sigma_{measurement}^2 \]

\[ \% \text{ Process Variation} = \frac{\sigma_{measurement}}{\sigma_{historical}} \times 100\% \]

\[ \% \text{Tolerance} = \frac{6 \times \sigma_{measurement}}{USL - LSL} \times 100\% \]

Sampling Plans for Expanded Gage R & R Studies, Lou Johnson and Daniel Griffith

https://www.youtube.com/watch?v=KtGhnimE6Qw

©JMC 2017
How the Simulation Works

In a gage R&R study, we have the following model:

$$Y_{ijkr} = \mu + P_i + O_j + G_k + \epsilon_{ijkr}$$

1) Specify a SD for Part, Operator, Gage and Error
2) Specify # of Samples for Part, Operator and Gage

Simulate Data

Repeat 5000 times

Histogram of Answers

Analyze Gage R&R – Calculate Variance Components

$\sigma_P^2$

$\sigma_G^2$

$\sigma_e^2$

Reset SD’s and Sample Size
95% Confidence Interval for Repeatability – 4 vs 2 Repeats

Compare

10 Parts X 3 Oper X 2 Repeats = 60 readings
5 Parts X 3 Oper X 4 Repeats = 60 readings

95% CI on Repeatability (2 Repeats)
(1.27 - .75) = .52

95% CI on Repeatability (4 Repeats)
(1.22 - .81) = .41

21% lower variability in your estimate of repeatability just by changing the sampling plan.
Sampling Plan for Chromatograph Peak Height Expanded Study

- 4 Columns used to run the 4 Chromatographs and 4 Techs in parallel
- 2 Repeats
- 96 measurements

This turns out to be the best overall sampling plan for an expanded study when the variation for each component can not be estimated.
Product is made and measured at three manufacturing sites. The third manufacturing site does not always meet specifications whereas the other two do. The goal is to determine which sources of variation (process/site, gage, operator, within lot) are the strongest contributors to the variation and therefore out-of-specification measurements.

A syringe needle manufacturer measures force – till – failure on an Instron mechanical properties gauge.
Sampling Plan to Determine Sources of Variation with Nesting

10 repeat measurements (needles) of 2 Parts crossed with 12 Operators nested in 6 Gages nested in 3 Locations.

Nesting makes it hard to separate sources of variation. More samples (needles) helps, but how many more?
Results of the Study -- Focus on Location and Repeatability

Variance Components

<table>
<thead>
<tr>
<th>Source</th>
<th>VarComp</th>
<th>% Contribution (of VarComp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Gage R&amp;R</td>
<td>0.0979</td>
<td>12.82</td>
</tr>
<tr>
<td>Repeatability</td>
<td>0.8338</td>
<td>10.91</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>0.1459</td>
<td>1.91</td>
</tr>
<tr>
<td>Location</td>
<td>0.0134</td>
<td>1.76</td>
</tr>
<tr>
<td>Operator (Location)</td>
<td>0.0009</td>
<td>0.12</td>
</tr>
<tr>
<td>Instron (Location)</td>
<td>0.0002</td>
<td>0.03</td>
</tr>
<tr>
<td>Part-To-Part</td>
<td>0.6661</td>
<td>87.18</td>
</tr>
<tr>
<td>Part</td>
<td>0.6661</td>
<td>87.18</td>
</tr>
<tr>
<td>Total Variation</td>
<td>0.7641</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Repeatability `is the largest source of variation, but this contains the needle – to – needle variation as well as the measurement variation.

Location is the next largest source of variation, eclipsing Operator and Instron.

So this study was successful, but did we need to use all those samples? 10 per condition?
Sample and Resample the Dataset with Different Sample Sizes

Analysis of Variance Model:

\[ Y_{si,jkr} = \mu + L_s + P_i + O_j + G_k + \varepsilon_{i,jkr} \]

1) Specify the size of a random sample - 2, 4, 6, or 8 data pts.
2) Randomly select that many points from the 10 in the data

Sample the Data Accordingly

Select a New Sample

Repeat 5000 times
Histogram of Answers

Analyze Dataset – Calculate Variance Components

Simulations by Cathy Akritas at Minitab, Inc. cakritas@minitab.com
Results of the Simulations

Results Distribution for Location for Sample Sizes of 2, 4, 6, 8, 10

True Answer = 1.87

Variable
- %Location2
- %Location4
- %Location6
- %Location8

Frequency

Estimated Variance % for Location

©JMC 2017
Results of the Simulations

Results Distribution for Repeatability for Sample Sizes of 2, 4, 6, 8, 10

10.6 = True Answer
Sample Size - 3 Factor Factorial with a Non-Normal Covariate

Incoming Contaminant
Random Lognormal

Three Factor Cleaning
Process Experiment

Clean Product
The power calculation can seem reasonably straightforward. 10 samples per treatment is reasonable.

But the response:

**Turbidity Before - Turbidity After**

is function of *Turbidity Before*.

\[1.1 - .7 = .4\] and \[16.4 - 16.0 = .4\]

have completely different meanings even though they are equal.

Even with random sampling, enough samples are needed to present each condition with the same distribution of incoming contamination.
Sample Size for a 3 Level Factorial with a Non-Normal Covariate

How many samples must be taken from this distribution before the samples will duplicate the parent distribution with reasonable consistency.

Randomly sample 5, 10, 15, 20, ...... etc., twenty times and plot their distribution with the parent.
Sample Size to Consistently Replicate the Distribution
Application of Simulation to Sample Size Calculations and Teaching Statistical Concepts

Questions?

Presented by:
Lou Johnson
lou@jmcdataexperts.com

www.jmcdataexperts.com