# Statistical Models for Gage R&R Studies

David Steinberg

Tel Aviv University

# Outline

- Terminology and goals
- The standard GR&R plan
- Statistical models
  - Sample of operators
  - All operators
- Study design
  - Sample sizes
  - Augmented plans
  - Baseline data
  - Leveraged plan
- Summary

# Terminology

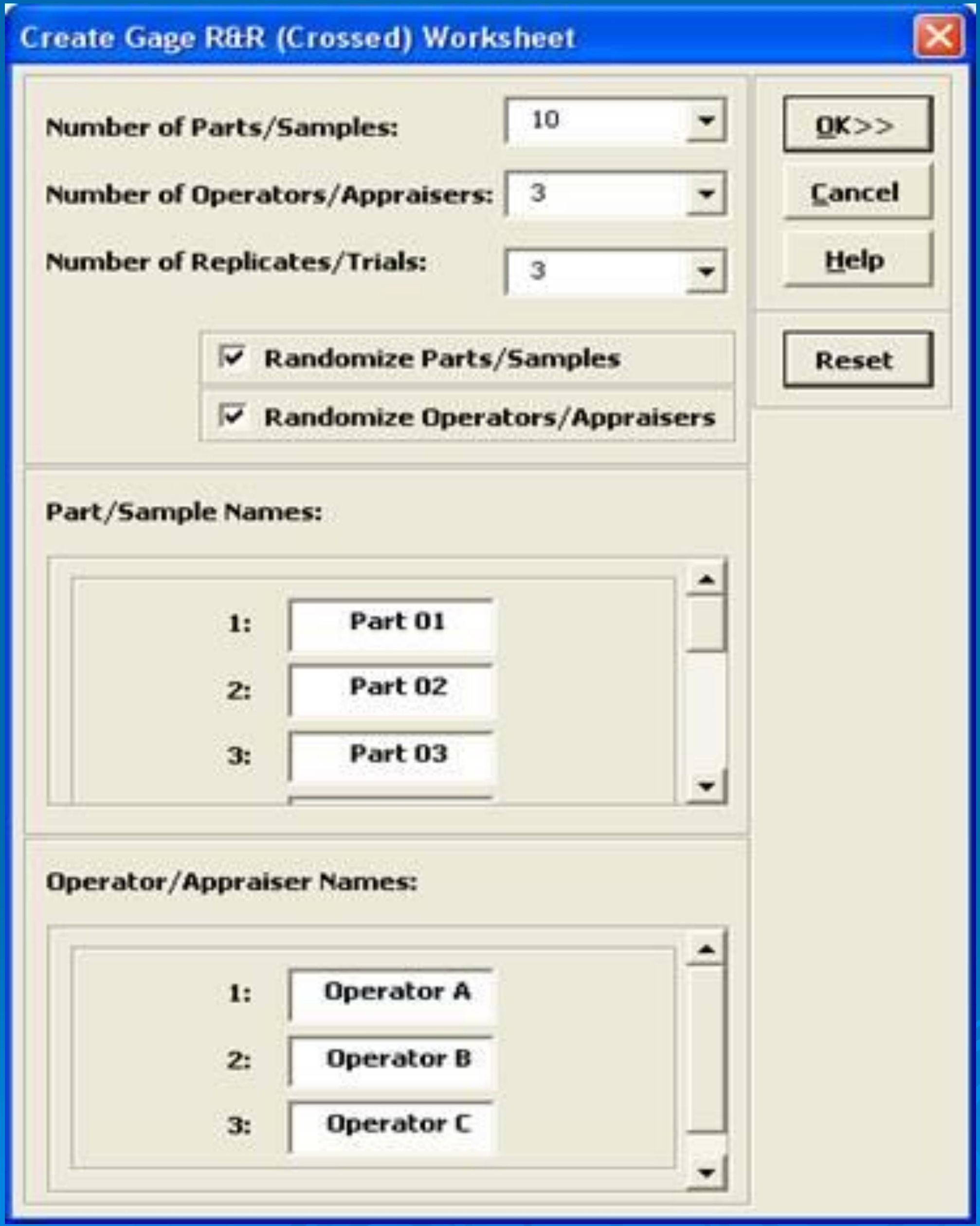
- Repeatability – the variation associated with repeat readings of the same item by the same operator.
- Reproducibility – the variation associated with differences among operators.
- Process variation – the actual variation among different parts produced by the process.

# Goals of Gage R&R

- Determine the repeatability of a measurement process.
- Determine the reproducibility of a measurement process.
- Compare R and R to the overall process variation.
- Improve the measurement process.

# The Standard Gage R&R

- Use a random sample of $n$ parts.
- Each part is measured by $J$ operators.
- Each operator measures each part $K$ times.
- Typically $n$=10, $J$,$K$=2 or 3.


Create Gage R&R (Crossed) Worksheet
Number of Parts/Samples:
10
Number of Operators/Appraisers:
3
Number of Replicates/Trials:
3
Randomize Parts/Samples
Randomize Operators/Appraisers
OK>>
Cancel
Help
Reset
Part/Sample Names:
1:
Part 01
2:
Part 02
3:
Part 03
Operator/Appraiser Names:
1:
Operator A
2:
Operator B
3:
Operator C

# The Standard Gage R&R

- Variation of mean values across operators is used to assess Reproducibility.
- Variance within the sets of *K* repeat measurements is used to estimate Repeatabiltiy.
- Use Analysis of Variance or means and ranges of the sets of *K* repeats.

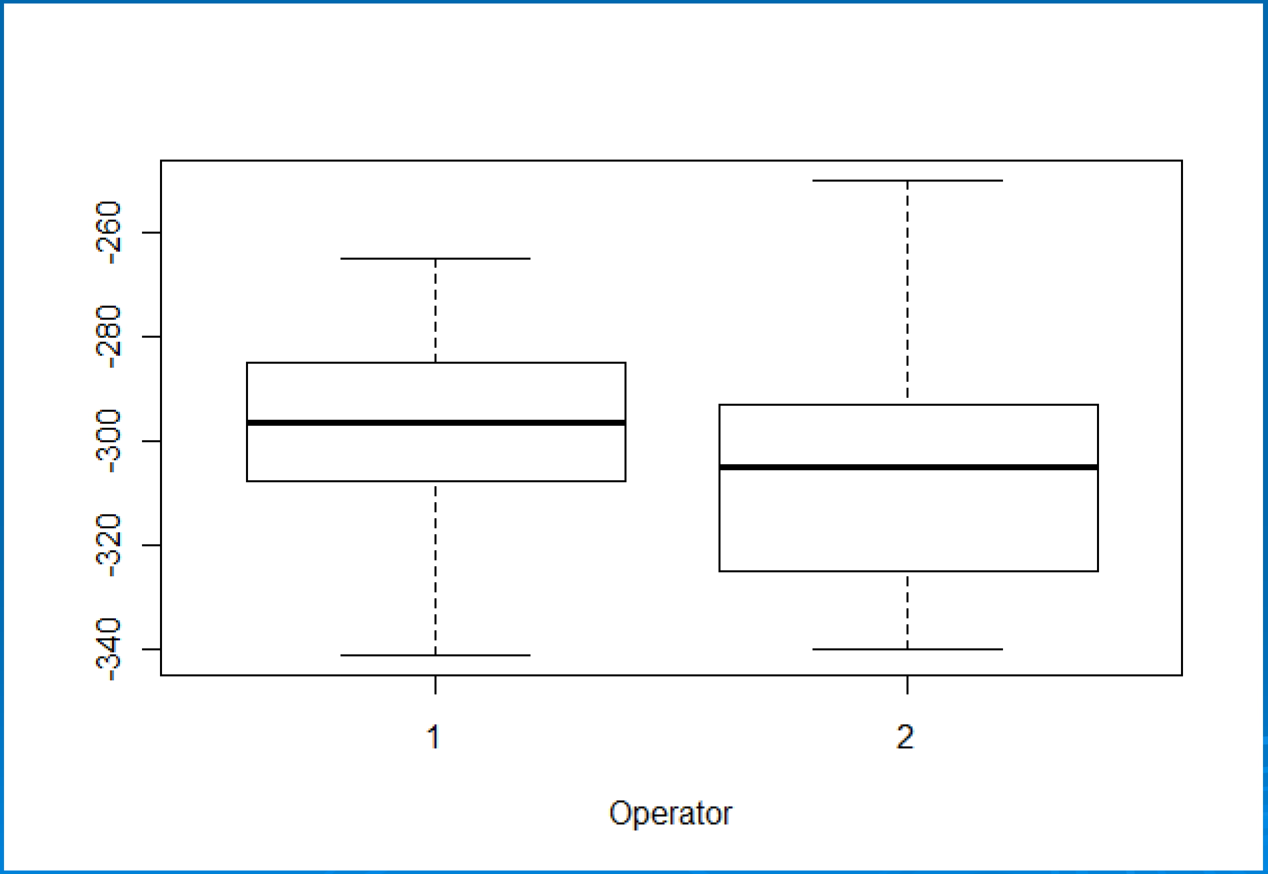
# Gage R&R – Questions

- How are the parts sampled?
- Is 10 a good sample size?
- Are the operators a sample or a population?
- How many operators should be included?
- What is a good choice for *K*?
- Should all parts be measured the same number of times?

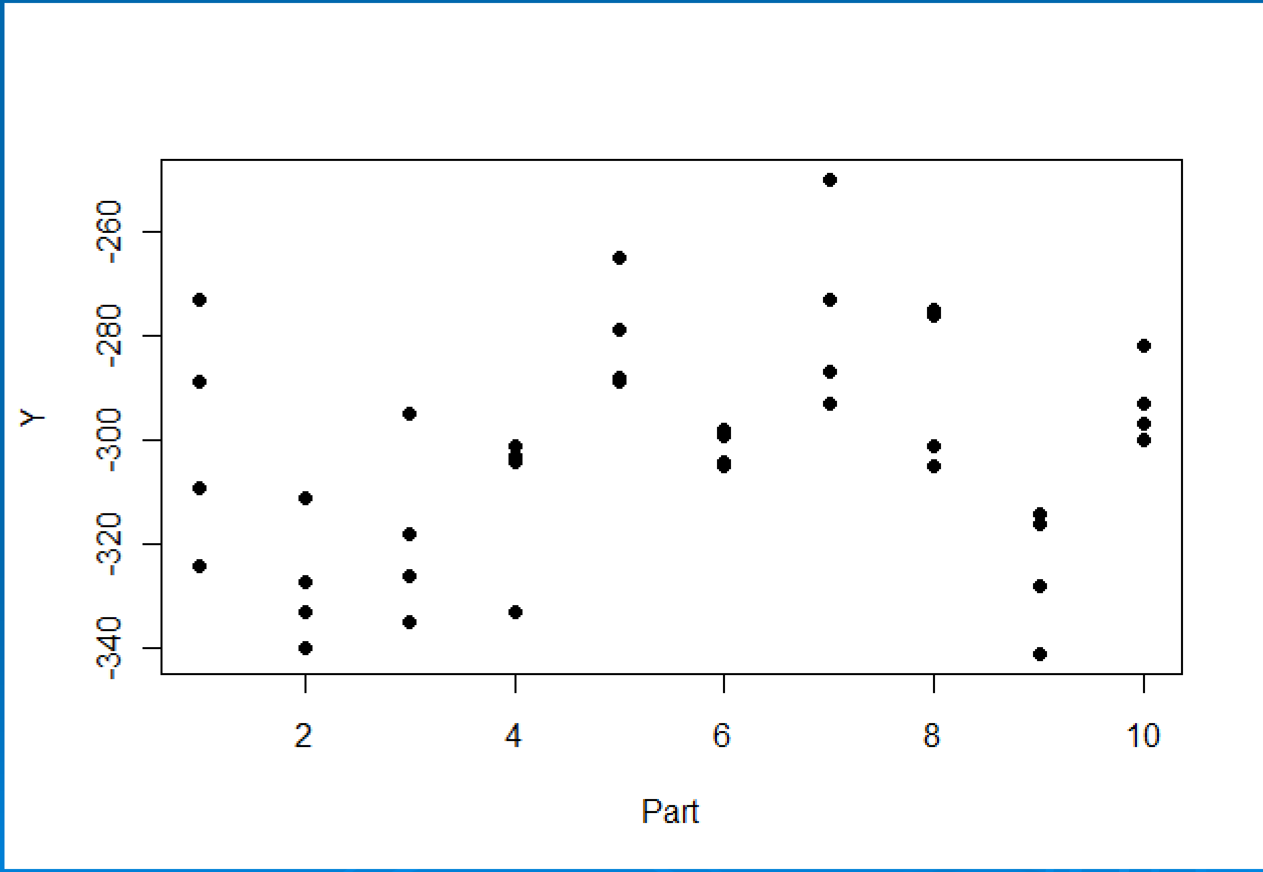
# Standard Gage R&R – Example

- Tsai (1988, Quality Engineering) reported on a study on an injection-molded part in a tool shop.  The specs were 685 ± 0.5 mm.
- The data show the difference between measurement and 685, in .001 mm units.

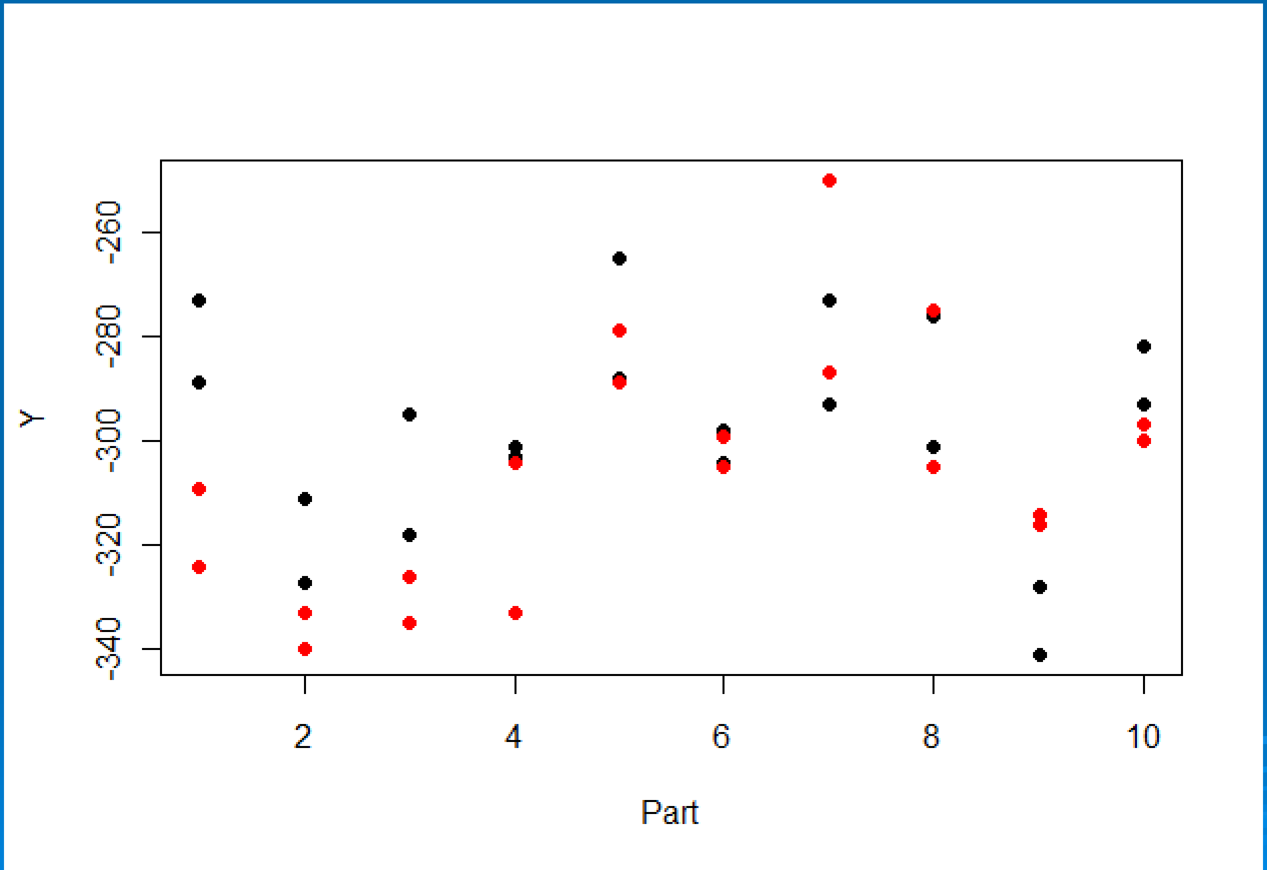
# Standard Gage R&R – Example



# Standard Gage R&R – Example



# Standard Gage R&R – Example



# Standard Gage R&R – Example

The Analysis of Variance table from this experiment:

| Source | SSq | df | MSq |
|---|---|---|---|
| Parts | 11,750 | 9 | 1306 |
| Operators | 648 | 1 | 648 |
| Parts by Op | 2,538 | 9 | 282 |
| Error | 3,270 | 20 | 163 |

# Gage R&R – Model

Let $y_{ijk}$ be the *k*'th measurement by the *j*'th operator of the *i*'th part.

A common statistical model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}.$$

Mean part size

Inter-part variation $\sigma_P^2$

Inter-operator variation $\sigma_O^2$

Part by operator interaction $\sigma_{PO}^2$

Repeatability variation $\sigma^2$

# Gage R&R – Model

Reproducibility: $\sigma_O^2 + \sigma_{PO}^2$

Repeatability: $\sigma^2$

Gage R&R: $\sigma_O^2 + \sigma_{PO}^2 + \sigma^2$

$$\sigma_{R\&R} = (\sigma_O^2 + \sigma_{PO}^2 + \sigma^2)^{0.5}$$

# Gage R&R – Sample of Operators

The study should describe ***all*** operators.

If it includes a ***sample*** of the possible operators, then we need to think how the sample represents the population.

A common model is

$$\beta_j \sim N(0, \sigma_O^2)$$

$$\alpha\beta_{ij} \sim N(0, \sigma_{PO}^2)$$

This is a *Random Effects* model.

# Gage R&R – Sample of Operators

With the random effects model, we can use ANOVA mean squares or sample ranges to estimate the variances.

| Mean Square | Expected Value |
|---|---|
| MSE | $\sigma^2$ |
| MSParts by Operators | $K\sigma_{PO}^2 + \sigma^2$ |
| MSOperators | $nK\sigma_O^2 + K\sigma_{PO}^2 + \sigma^2$ |
| MSParts | $JK\sigma_P^2 + K\sigma_{PO}^2 + \sigma^2$ |

# Gage R&R – Sample of Operators

The standard range estimators are based on the within cell ranges $R_{ij}$ and the range of the per operator averages,

$$R_O = \max \bar{y}_{.j} - \min \bar{y}_{.j}.$$

# Gage R&R – Sample of Operators

The basis for the range estimators is as follows: if $Z_1, \ldots, Z_n$ are a random sample from $N(\mu, \tau^2)$, then $E\{\text{Range}\} = d_2(n)\,\tau$.

So it is easy to "correct" the range into an unbiased estimator.

The factor $d_2(n)$ is a standard term available in software and tables.

# Gage R&R – Sample of Operators

The range estimator of reproducibility is $R_0/d_2(J)$.

The expected value of this estimator is

$$\left( \sigma_O^2 + \frac{1}{n}\sigma_{PO}^2 + \frac{1}{nK}\sigma^2 \right)^{1/2}$$

It can be corrected to be consistent only if there is no interaction.

# Gage R&R – All Operators

The study might include all operators.  In that case all we need is to characterize that specific team of operators.

The sampling assumption for operators is no longer reasonable; the operator terms are *Fixed Effects*.

The part by operator interaction still involves sampling, so remains a random effect.

This leads to a *Mixed Effects* model.

# Gage R&R – All Operators

In the *Mixed Effects* model,

$$E\{MSOperators\} = \frac{IK}{J-1}\sum \beta_j^2 + K\sigma_{PO}^2 + \sigma^2$$

$$\sigma_O^2 = \frac{1}{J-1}\sum \beta_j^2$$

There is an implicit assumption that $\sum \beta_j = 0$ .

# Standard Gage R&R – Example

The operators here were considered to be a sample. Equating mean squares with their expectations to get estimates:

| Variance | Estimate |
|---|---|
| $\sigma^2$ | 163 |
| $\sigma^2_{PO}$ | 59 |
| $\sigma^2_O$ | 18 |
| Reproducibility | 77 |
| $\sigma^2_P$ | 256 |

# Gage R&R – Study Design

How does the design affect the quality of the estimators?

See Vardeman and Van Valkenburg (*Technometrics*, 1999, 202-211).

The ability to estimate $\sigma_O^2$ depends heavily on *J*, the number of operators sampled. For a fixed total sample size, using only 2 or 3 operators is not efficient. It is better to increase *J* and decrease *n* and *K*.

# Gage R&R – Study Design

When the sole goals are repeatability and reproducibility, Vardeman and Van Valkenburg found that the best designs typically use just 1 part, measured 2 or 3 times by many operators.

More repeats are desirable when repeatability is much larger than reproducibility.

# Gage R&R – Augmented Plans

The standard plan is balanced – each part by operator pair has the same number of measurements.

Stevens, Browne, Steiner and MacKay (2010, *JQT*, 388-399) look at the benefits of unbalanced plans.

# Gage R&R – Augmented Plans

They assume the goal is to estimate the ratio of measurement variance (R&R) to total variance and that all operators are included.

Their plan:

1. Begin with a small standard plan.
2. Augment the plan with further data.

# Gage R&R – Augmented Plans

**Type A augmentation:**

Each operator measures a new set of parts once each.

**Type B augmentation:**

Each operator measures the same set of parts once each.

The benefits of both options are that measuring more parts leads to better estimates of the inter-part variance.

# Gage R&R – Augmented Plans

With just one operator (e.g. automated measurement systems), the best plans are standard plans with 2 measurements per part.

With more than one operator, and no part by operator interaction, type A plans are best, typically by 7%-20% relative to the best standard plan.

When there is part by operator interaction, the best plans depend on the number of operators.

With two operators, plans of type B with a very small standard plan are most efficient.

With 3 or 4 operators, plans of type A with a small standard plan are most efficient.

# Gage R&R – Baseline Data

Stevens, Browne, Steiner and MacKay (2012, *IIE Transactions*, 1166-1175) discuss the use of ***baseline data*** – ongoing process data that can be used to estimate the total degree of variability.

Combining baseline data with a standard plan is just like the type A augmented plan, except that the baseline data are already in hand.

# Gage R&R – Baseline Data

Using baseline data offers substantial gains in precision for estimating R&R.

The best plans to use along with baseline data often have very few parts.

# Gage R&R – Leveraged Plans

Browne, Steiner and MacKay (2009, *Technometrics*, 239-249) ask whether the parts should be sampled at random.

They focus on a system with just one operator and on estimation of the ratio of measurement variance to total variance.

# Gage R&R – Leveraged Plans

The study has two phases.  The first is a random sample of *b* parts, each measured once.

The second phase uses a sample of *n* parts from the first phase, chosen to include parts with extreme values, but an average that is similar to the phase 1 average.  Each part is measured *K* times.

# Gage R&R – Leveraged Plans

Why sample extreme parts?

The idea is that the variance ratio of interest can be thought of as a regression coefficient, and it can be estimated most accurately when the regression problem focuses on parts with extreme values.

# Gage R&R – Leveraged Plans

Simulation studies show that leveraged plans are often much more efficient than standard plans in this setting.

# Gage R&R – Summary

- Don’t use GR&R plans blindly.
- Think about whether your study includes all operators; represents all operators.
- Be wary of range-based estimates. Modern software makes it easy to use better estimators.
- Don’t adopt the standard design without a close look at the goals of your study and the alternatives for design.