Module 4

The analysis of data from standard gauge R&R studies

Prof. Stephen B. Vardeman
Statistics and IMSE
Iowa State University

March 4, 2008

Parts, Operators, and Repeat Measurements

Gauge R&R studies are meant to quantify sizes of sources of measurement *imprecision* (in particular, the size of operator differences). Data are typically m measurements made on each of I parts/measurands by each of J operators using a fixed gauge, measurement protocol, etc.

Gauge R&R for a 1-Inch Micrometer Caliper. Heyde, Kuebrick, and Swanson conducted a gauge R&R study on a micrometer caliper. Below are what the $J = 3$ (student) operators obtained, each making $m = 3$ measurements of the heights of $I = 4$ steel punches.

Table: Measured Heights of 10 Steel Punches in 10^{-3} Inch

	Student 1	**Student 2**	**Student 3**
Punch 1	496, 496, 499	497, 499, 497	497, 498, 496
Punch 2	498, 497, 499	498, 496, 499	497, 499, 500
Punch 3	498, 498, 498	497, 498, 497	496, 498, 497
Punch 4	497, 497, 498	496, 496, 499	498, 497, 497

"Two-Way Random Effects Models" for Gauge R&R Data

With

$$y_{ijk} = \text{the } k\text{th measurement made by operator } j \text{ on part } i \ ,$$

the model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \ ,$$

where μ is an (unknown) constant and the other terms are (independent) normal variables with means 0 and variances σ_α^2 (for the α's), σ_β^2 (for the β's), $\sigma_{\alpha\beta}^2$ (for the $\alpha\beta$'s),and σ^2 (for the ϵ's). In this model

- the unknown constant μ is an "average" measurement
- the α's are (random) effects of different parts
- the β's are (random) effects of different operators
- the $\alpha\beta$'s are (random) joint effects peculiar to particular part$\times$operator combinations, and
- the ϵ's are (random) "repeatability" errors.

$\sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}^2$, and σ^2 are "variance components" and their sizes govern how much variability is seen in the measurements y_{ijk}.

A Very Small Hypothetical Case

For $I = 2$, $J = 2$, and $m = 2$

	Operator 1	Operator 2
Part 1	$y_{111} = \mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \epsilon_{111}$ $y_{112} = \mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \epsilon_{112}$	$y_{121} = \mu + \alpha_1 + \beta_2 + \alpha\beta_{12} + \epsilon_{121}$ $y_{122} = \mu + \alpha_1 + \beta_2 + \alpha\beta_{12} + \epsilon_{122}$
Part 2	$y_{211} = \mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \epsilon_{211}$ $y_{212} = \mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \epsilon_{212}$	$y_{221} = \mu + \alpha_2 + \beta_2 + \alpha\beta_{22} + \epsilon_{221}$ $y_{222} = \mu + \alpha_2 + \beta_2 + \alpha\beta_{22} + \epsilon_{222}$

Interpreting Effects and Model Parameters

- The only differences between measurements for a fixed part$\times$operator combination are the errors ϵ. The variability of these is governed by the parameter σ. That is, σ is a measure of "repeatability" variation in this model
- For fixed "part i" (row i), the quantity $\mu + \alpha_i$ is constant. This can be interpreted as the value of the ith measurand (these vary across parts/rows because the α_i vary).
- For a fixed part i, the values $\beta_j + \alpha\beta_{ij}$ vary column/operator to column/operator and function as *part-i-specific operator biases*.
  - Since these change with the measurand because the $\alpha\beta_{ij}$ change row-to-row of a fixed column, this model allows for (random) "nonlinearity" of the operators through these "interaction" terms. Only when $\sigma^2_{\alpha\beta} \approx 0$ (so all $\alpha\beta \approx 0$) are the "devices" linear.

Interpreting Effects and Model Parameters (cont.)

- The variance of $\beta_j + \alpha\beta_{ij}$ is $\sigma^2_\beta + \sigma^2_{\alpha\beta}$, so an appropriate measure of "reproducibility" variation is

$$\sigma_{\text{reproducibility}} = \sqrt{\sigma^2_\beta + \sigma^2_{\alpha\beta}}$$

  - This standard deviation would be experienced by many operators making single measurements on the same part *if there were no repeatability component to the variation in measurements*.
  - It is also the standard deviation that would be experienced computing with "long-run average measurements" for many operators on the same part. That is, *this is a measure of variability in operator bias for a fixed part*.

Interpreting Effects and Model Parameters (cont.)

- The quantity

$$\sigma_{\text{R\&R}} = \sqrt{\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2 + \sigma^2} = \sqrt{\sigma_{\text{reproducibility}}^2 + \sigma^2}$$

  is the standard deviation for many operators each making a single measurement on the same part. It is a measure of the combined imprecision in measurement attributable to *both* repeatability and reproducibility sources.

- And one might think of

$$\frac{\sigma^2}{\sigma_{\text{R\&R}}^2} = \frac{\sigma^2}{\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2 + \sigma^2} \quad \text{and} \quad \frac{\sigma_{\text{reproducibility}}^2}{\sigma_{\text{R\&R}}^2} = \frac{\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2}{\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2 + \sigma^2}$$

  as the fractions of total measurement variance due respectively to repeatability and reproducibility.

Interpreting Effects and Model Parameters (cont.)

- It is common to treat some multiple of $\sigma_{\text{R\&R}}$ (often the multiplier is six, but sometimes 5.15 is used) as a kind of uncertainty associated with a measurement made using the gauge or measurement system in question. And when a gauge is being used to check conformance of a measured part characteristic to engineering specifications (say, some lower specification L and some upper specification U) this multiple is compared to the spread in specifications. So it is common to call the quantity

$$GCR = \frac{6\sigma_{\text{R\&R}}}{U - L}$$

  a **gauge capability (or precision-to-tolerance) ratio,** and require that it be no larger than .1 (and preferably as small as .01) before using the gauge for checking conformance to such specifications. (In practice, one will only have an estimate of $\sigma_{\text{R\&R}}$ upon which to make an empirical approximation of a gauge capability ratio.)

Analysis of Gauge R&R Data

Range-based estimation of R&R quantities (based on AIAG or similar forms) is common. But it is less than ideal because:

- the methods in common use are *wrong* at least as far as estimation of $\sigma_{\text{reproducibility}}$ and $\sigma_{\text{R\&R}}$ is concerned. (They estimate something other than these quantities.) The first edition of Vardeman and Jobe has correct range-based methods.
- they are necessarily inefficient (better estimates are available even after their correction).
- they have no associated confidence interval methods. (Single-number estimates with no indication of their precision are of little real use.)

- The best existing technology for analysis of two-way random effects data is "restricted maximum likelihood" methodology. This is implemented in some widely available commercial packages (like ™JMP and in the high quality open source R software), but since the R&R application is a bit specialized, what we'd most like from the software is not directly available. (Intervals for the variance components one at a time *are* directly available, but not intervals for our more specialized quantities. These can be made, but describing how to piece together what these programs provide in order to get them is not simple.)
- It's possible to give some "not-impossible-to-use-'by-hand'" formulas for intervals based on ANOVA computations and Satterthwaite approximations. We'll illustrate what these give.

This is implemented in all sensible statistical packages. (If you want to live dangerously, this also can *possibly* be done using ^{TM}`EXCEL`. ^{TM}`Google` will help you find pages like
`http://www.cvgs.k12.va.us/digstats/main/Guides/g_2anovx.html`
showing how to try this.)

We'll suppose that we've been able to get the ANOVA sums of squares. That is, with

$$\bar{y}_{i.} = \frac{1}{J}\sum_{j} \bar{y}_{ij} \quad \text{and} \quad \bar{y}_{.j} = \frac{1}{I}\sum_{i} \bar{y}_{ij} \quad \text{and} \quad \bar{y}_{..} = \frac{1}{IJ}\sum_{ij} \bar{y}_{ij} \ .$$

we'll suppose that some sort of statistical package has been employed to produce

Two-Way ANOVA (cont.)

$$
\begin{aligned}
SSTot &= \sum_{ijk} \left(y_{ijk} - \bar{y}_{..} \right)^2 \ , \\
SSE &= \sum_{ijk} \left(y_{ijk} - \bar{y}_{ij} \right)^2 \ , \\
SSA &= mJ \sum_{i} \left(\bar{y}_{i.} - \bar{y}_{..} \right)^2 \ , \\
SSB &= mI \sum_{j} \left(\bar{y}_{.j} - \bar{y}_{..} \right)^2 \ , \text{ and} \\
SSAB &= m \sum_{ij} \left(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \right)^2 \\
&= SSTot - SSE - SSA - SSB
\end{aligned}
$$

"Degrees of freedom" and corresponding "mean squares" for these sums of squares are

$$\begin{aligned} dfE &= (m-1)\,IJ \;\text{ and }\; MSE = SSE/\,(m-1)\,IJ\;, \\ dfA &= I-1 \;\text{ and }\; MSA = SSA/\,(I-1)\;, \\ dfB &= J-1 \;\text{ and }\; MSB = SSB/\,(J-1)\;, \text{ and} \\ dfAB &= (I-1)\,(J-1) \;\text{ and }\; MSAB = SSAB/\,(I-1)\,(J-1)\;. \end{aligned}$$

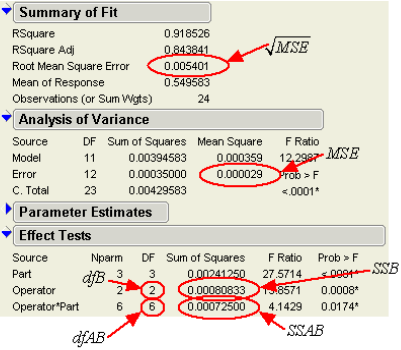
Two-Way ANOVA (cont.)

All of these calculations are often summarized in so-called "ANOVA tables" like

Source	**SS**	**df**	**MS**
Part	SSA	$I-1$	$MSA = \frac{SSA}{I-1}$
Operator	SSB	$J-1$	$MSB = \frac{SSB}{J-1}$
Part×Operator	$SSAB$	$(I-1)(J-1)$	$MSAB = \frac{SSAB}{(I-1)(J-1)}$
Error	SSE	$IJ(m-1)$	$MSE = \frac{SSE}{IJ(m-1)}$
Total	$SSTot$	$IJm-1$	

Example 4-2 Gauge R&R on a Cheap Plastic Caliper (TM JMP)

	Part	Operator	Peanut Measurement
1	1	1	0.52
2	1	1	0.52
3	1	2	0.54
4	1	2	0.53
5	1	3	0.55
6	1	3	0.55
7	2	1	0.56
8	2	1	0.55
9	2	2	0.54
10	2	2	0.54
11	2	3	0.55
12	2	3	0.56
13	3	1	0.57
14	3	1	0.56
15	3	2	0.55
16	3	2	0.56
17	3	3	0.57
18	3	3	0.57
19	4	1	0.55
20	4	1	0.55
21	4	2	0.54
22	4	2	0.55
23	4	3	0.56
24	4	3	0.55



Summary of Fit

RSquare	0.918526
RSquare Adj	0.843841
Root Mean Square Error	0.005401
Mean of Response	0.549583
Observations (or Sum Wgts)	24

Analysis of Variance

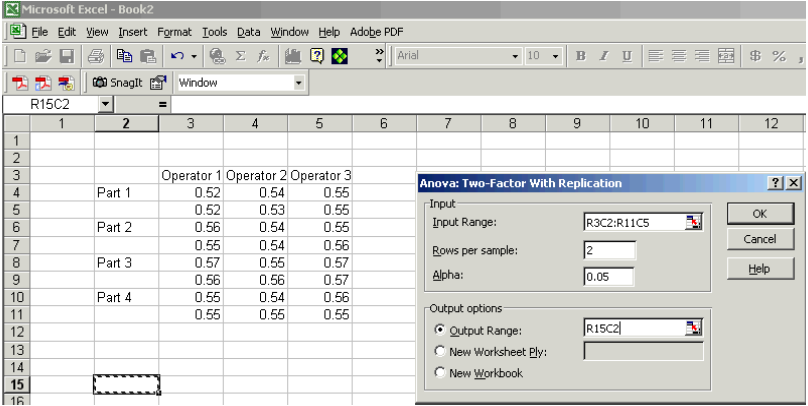
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	0.00394583	0.000359	12.2987
Error	12	0.00035000	0.000029	Prob > F
C. Total	23	0.00429583		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Part	3	3	0.00241250	27.5714	<.0001*
Operator	2	2	0.00080833	13.8571	0.0008*
Operator*Part	6	6	0.00072500	4.1429	0.0174*

Example 4-2 (^{TM}EXCEL)



ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.002412	3	0.000804	27.57143	1.15E-05	3.4903
Columns	0.000808	2	0.000404	13.85714	0.000761	3.88529
Interaction	0.000725	6	0.000121	4.142857	0.017388	2.996117
Within	0.00035	12	2.92E-05			
Total	0.004296	23				

Estimates and Intervals

Estimates based on ANOVA are

$$\hat{\sigma}_{\text{repeatability}} = \widehat{\sigma} = \sqrt{MSE} \; ,$$

$$\hat{\sigma}_{\text{reproducibility}} = \sqrt{\max\left(0, \frac{MSB}{mI} + \frac{(I-1)}{mI}MSAB - \frac{1}{m}MSE\right)} \; ,$$

and

$$\hat{\sigma}_{\text{R\&R}} = \sqrt{\frac{1}{mI}MSB + \frac{I-1}{mI}MSAB + \frac{m-1}{m}MSE} \; .$$

Intervals are

$$\left(\hat{\sigma}\sqrt{\frac{\hat{\nu}}{\chi^2_{\hat{\nu},\text{upper}}}}, \hat{\sigma}\sqrt{\frac{\hat{\nu}}{\chi^2_{\hat{\nu},\text{lower}}}}\right) \; .$$

Satterthwaite Approximate Degrees of Freedom

Degrees of freedom are

$$\nu_{\text{repeatability}} = IJ\left(m-1\right) \;,$$

$$\begin{aligned}
\hat{\nu}_{\text{reproducibility}} &= \frac{\hat{\sigma}^4_{\text{reproducibility}}}{\dfrac{\left(\dfrac{MSB}{mI}\right)^2}{J-1} + \dfrac{\left(\dfrac{(I-1)\,MSAB}{mI}\right)^2}{(I-1)(J-1)} + \dfrac{\left(\dfrac{MSE}{m}\right)^2}{IJ\,(m-1)}} \\
&= \frac{\hat{\sigma}^4_{\text{reproducibility}}}{\dfrac{1}{m^2}\left(\dfrac{MSB^2}{I^2\,(J-1)} + \dfrac{(I-1)\,MSAB^2}{I^2\,(J-1)} + \dfrac{MSE^2}{IJ\,(m-1)}\right)} \;,
\end{aligned}$$

and

Satterthwaite Approximate Degrees of Freedom (cont.)

$$\hat{\nu}_{\text{R\&R}} = \frac{\hat{\sigma}^4_{\text{R\&R}}}{\frac{\left(\frac{MSB}{mI}\right)^2}{J-1} + \frac{\left(\frac{(I-1)\,MSAB}{mI}\right)^2}{(I-1)(J-1)} + \frac{\left(\frac{(m-1)\,MSE}{m}\right)^2}{IJ(m-1)}}$$

$$= \frac{\hat{\sigma}^4_{\text{R\&R}}}{\frac{1}{m^2}\left(\frac{MSB^2}{I^2(J-1)} + \frac{(I-1)\,MSAB^2}{I^2(J-1)} + \frac{(m-1)\,MSE^2}{IJ}\right)} .$$

Example 4-2 (cont.)

Some arithmetic produces

$$\hat{\sigma}_{\text{repeatability}} = .005401 \qquad \nu_{\text{repeatability}} = 12$$
$$\hat{\sigma}_{\text{reproducibility}} = .009014 \qquad \hat{\nu}_{\text{reproducibility}} = 4.035$$
$$\hat{\sigma}_{\text{R\&R}} = .011 \qquad \hat{\nu}_{\text{R\&R}} = 7.452$$

Then (rounding approximate df down) one gets 95% limits

$$\begin{array}{lll} .0039\,\text{in} & \text{and} \quad .0089\,\text{in} & \text{for } \sigma = \sigma_{\text{repeatability}} \\ .0054\,\text{in} & \text{and} \quad .0259\,\text{in} & \text{for } \sigma_{\text{reproducibility}} \\ .0073\,\text{in} & \text{and} \quad .0224\,\text{in} & \text{for } \sigma_{\text{R\&R}} \end{array}$$

None of these standard deviations is terribly well-determined (degrees of freedom are small and intervals are wide). If better information is needed, more data would have to be collected. But there is at least some indication that $\sigma_{\text{repeatability}}$ and $\sigma_{\text{reproducibility}}$ are of the same order of magnitude. The caliper used to make the measurements was fairly crude, *and* there were detectable differences in the way the student operators used that caliper.