IE 361 Module 8

Repeatability and Reproducibility for "0/1" (or Go/No-Go) Contexts

Reading: Section 2.7 of Revised *SQAME*

Prof. Steve Vardeman and Prof. Max Morris

Iowa State University

August 2008

Go/No-Go Inspection

Ideally, observation of a process results in quantitative measurements. But there are some contexts in which all that is determined is whether or not an item or process condition is of one of two types, that we will for the present call "conforming" and "non-conforming." It is, for example, common to check the conformance of machined metal parts to some engineering requirements via the use of a "go/no-go gauge." (A part is conforming if a critical dimension fits into the larger of two check fixtures and does not fit into the smaller of the two.) And it is common to task human beings with making visual inspections of manufactured items and producing a "OK/Not-OK" call on each.

Engineers are sometimes then called upon to apply the qualitative "repeatability" and "reproducibility" concepts of metrology to such Go/No-Go or "0/1" contexts. One wants to separate some measure of overall inconsistency in 0/1 "calls" on items into pieces that can be mentally charged to inherent inconsistency in the equipment or method, and the remainder that can be charged to differences between how operators use it. Exactly how to do this is, in fact, presently not well-established. The best available statistical methodology for this kind of problem is more complicated than can be presented here (involving so-called "generalized linear models"). What we can present is a rational way of making point estimates of what might be termed repeatability and reproducibility components of variation in 0/1 calls. (The estimates presented here are based on reasoning similar to that employed in *SQAME* to find correct range-based estimates in usual measurement R&R contexts.) We will then remind you of elementary methods of estimating differences in population proportions and point to their relevance in the present situation.

Some Simple Probability Modeling

To begin, think of coding a "non-conforming" call as "1" and a "conforming" call as "0," and having J operators each make m calls **on a fixed part**. Suppose that J operators have individual probabilities $p_1, p_2, \ldots p_J$ of calling the part non-conforming on any single viewing, and that across m viewings

$X_j =$ the number of non-conforming calls among the m made by operator j

is Binomial (m, p_j). **We'll assume that the p_j are random draws from some population with mean π and variance v.**

The quantity

$$p_j \left(1 - p_j\right)$$

is a kind of "per call variance" associated with the declarations of operator j, and might serve as a kind of repeatability variance for that operator.

(Given the value of p_j, elementary probability says that the variance of X_j is $mp_j(1-p_j)$.) The biggest problem here is that unlike what is true in the usual case of Gauge R&R for measurements, this variance is not constant across operators. But its expected value, namely

$$\begin{aligned} \mathrm{E}\left(p_j\left(1-p_j\right)\right) &= \pi - \mathrm{E}p_j^2 \\ &= \pi - \left(v + \pi^2\right) \\ &= \pi\left(1-\pi\right) - v \end{aligned}$$

can be used as a sensible measure of variability in conforming/non-conforming classifications chargeable to repeatability sources. The variance v serves as a measure of reproducibility variance.

Some Simple Probability Modeling

This ultimately points to

$$\pi(1-\pi)$$

as the "total R&R variance" in this context. That is, we make definitions for 0/1 contexts

$$\sigma^2_{\text{R\&R}} = \pi(1-\pi)$$

$$\sigma^2_{\text{repeatability}} = \pi(1-\pi) - v$$

and

$$\sigma^2_{\text{reproducibility}} = v$$

Some Simple R&R Point Estimates for 0/1 Contexts

Still thinking of a single fixed part, we'll let

$$\hat{p}_j = \frac{\text{the number of "non-conforming" calls made by operator } j}{m} = \frac{X_j}{m}$$

and define the (sample) average and (sample) variance of these

$$\bar{\hat{p}} = \frac{1}{J}\sum_{j=1}^{J}\hat{p}_j \quad \text{and} \quad s_{\hat{p}}^2 = \frac{1}{J-1}\sum_{j=1}^{J}\left(\hat{p}_j - \bar{\hat{p}}\right)^2$$

It is possible to argue that

$$\mathrm{E}\bar{\hat{p}} = \pi$$

and that

$$\begin{aligned} \mathrm{E}s_{\hat{p}}^2 &= \mathrm{Var}\hat{p}_j \\ &= \mathrm{Var}p_j + \mathrm{E}\frac{p_j\left(1-p_j\right)}{m} \\ &= \frac{m-1}{m}v + \frac{\pi\left(1-\pi\right)}{m} \end{aligned}$$

Some Simple R&R Point Estimates for 0/1 Contexts

so that

$$v = \frac{m}{m-1}\mathrm{E}s_{\hat{p}}^2 - \frac{\pi(1-\pi)}{m-1}$$

This suggests the simple estimators (still based on a single part)

$$\hat{\sigma}^2_{\mathrm{R\&R}} = \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$$

$$\hat{\sigma}^2_{\text{reproducibility}} = \max\left(0, \frac{1}{m-1}\left(ms_{\hat{p}}^2 - \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)\right)\right)$$

and

$$\hat{\sigma}^2_{\text{repeatability}} = \hat{\sigma}^2_{\mathrm{R\&R}} - \hat{\sigma}^2_{\text{reproducibility}}$$

(On rare occasions, $s_{\hat{p}}^2$ will exceed $\bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$, leading to a value of $\hat{\sigma}^2_{\text{reproducibility}}$ above larger than $\hat{\sigma}^2_{\mathrm{R\&R}}$. In those cases, we will reduce $\hat{\sigma}^2_{\text{reproducibility}}$ to $\hat{\sigma}^2_{\mathrm{R\&R}} = \bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$.)

Some Simple R&R Point Estimates for 0/1 Contexts

What to do based on multiple parts (say I of them) is not completely obvious. For our purposes in IE 361, we will simply average estimates made one part at a time across multiple parts, presuming that parts in hand are sensibly thought of as a random sample of parts to be checked, and that this averaging is a sensible way to combine information across parts.

In order for any of this to have a chance of working, m is going to have to be fairly large. The usual Gauge R&R "$m = 2$ or 3" just isn't going to produce informative results in the present context. And in order for this to work in practice (so that an operator isn't just repeatedly looking at the same few parts over and over and remembering how he's called them in the past) this seems like it's going to require a large value of I as well as m.

Suppose that $I = 5$ parts are inspected by $J = 3$ operators, $m = 10$ times apiece, and that in the table below are sample fractions of "non-conforming" calls made by the operators and a few summary statistics.

	Operator 1	Operator 2	Operator 3	$\bar{\hat{p}}$	$\bar{\hat{p}}\left(1-\bar{\hat{p}}\right)$	$s^2_{\hat{p}}$
Part 1	.2	.4	.2	.2667	.1956	.0133
Part 2	.6	.6	.7	.6333	.2322	.0033
Part 3	1.0	.8	.7	.8333	.1389	.0233
Part 4	.1	.1	.1	.1	.0900	0
Part 5	.1	.3	.3	.2333	.1789	.0133

The entries in the next to last column of the table above are $\hat{\sigma}^2_{\text{R\&R}}$ values for the 5 parts. Estimates of $\hat{\sigma}^2_{\text{reproducibility}}$ are, for example, computed as for Part 1

$$\begin{aligned}\hat{\sigma}^2_{\text{reproducibility}} &= \max\left(0, \frac{1}{10-1}\left(10\left(.0133\right) - .1956\right)\right) \\ &= 0\end{aligned}$$

leaving estimates of $\hat{\sigma}^2_{\text{repeatability}}$ computed as for Part 1

$$\begin{aligned}\hat{\sigma}^2_{\text{repeatability}} &= .1956 - 0 \\ &= .1956\end{aligned}$$

The whole set of estimates and their averages are collected in another table below.

	$\hat{\sigma}^2_{\text{R\&R}} = \bar{\hat{p}}\left(1 - \bar{\hat{p}}\right)$	$\hat{\sigma}^2_{\text{reproducibility}}$	$\hat{\sigma}^2_{\text{repeatability}}$
Part 1	.1956	0	.1956
Part 2	.2322	0	.2322
Part 3	.1389	.0105	.1284
Part 4	.0900	0	.0900
Part 5	.1789	0	.1789
Average	.1671	.0021	.1650

Then, for example, a fraction of only

$$\frac{.0021}{.1671} = 1.3\%$$

of the inconsistency in conforming/non-conforming calls seen in the original data seems to be attributable to clear differences in how the operators judge the parts (differences in the binomial "success probabilities" p_j). Rather, the bulk of the variance seems to be attributable to unavoidable binomial variation. The p's are not close enough to either 0 or 1 to make the calls tend to be consistent. So the variation seen in the $\hat{p}$'s in a given row is not clear evidence of large operator differences.

Of course, we need to remember that the computations above are on the *variance* (and not standard deviation) scale. On the (more natural) standard deviation scale, reproducibility variation

$$\sqrt{.0021} = .05$$

and repeatability variation

$$\sqrt{.1650} = .41$$

are not quite so strikingly dissimilar.

Application of Inference for the Difference in Two Proportions

The question of whether call rates for two operators on the same part are really detectably different brings up the Stat 231 topic of estimating the difference in two binomial parameters, say p_1 and p_2. Recall that an elementary large sample approximate confidence interval for $p_1 - p_2$ has endpoints

$$\hat{p}_1 - \hat{p}_2 \pm z\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}$$

As it turns out, this formula can fail badly for small sample sizes one typically would meet in an R&R study. So rather than use it as taught in Stat 231, we will use a slight modification that is more reliable, namely

$$\hat{p}_1 - \hat{p}_2 \pm z\sqrt{\frac{\tilde{p}_1\left(1-\tilde{p}_1\right)}{n_1} + \frac{\tilde{p}_2\left(1-\tilde{p}_2\right)}{n_2}}$$

where

$$\tilde{p}_i = \frac{n_i \hat{p}_i + 2}{n_i + 4}$$

That is, under the square root of the usual formula one essentially replaces the $\hat{p}$ values with $\tilde{p}$ values derived by adding 2 "successes" in 4 "additional trials" to the counts used to make up the $\hat{p}$ values. (This has the effect of making the standard large sample interval a bit wider and correcting the problem that for small sample sizes and extreme values of p it can fail to hold its nominal confidence level.)

Consider again Part 1 from the earlier example, and in particular consider the question of whether Operator 1 and Operator 2 have clearly different probabilities of calling that part non-conforming on a single call. With $\hat{p}_1 = .2$ and $\hat{p}_2 = .4$,

$$\tilde{p}_1 = \frac{2+2}{10+4} = .2857 \text{ and } \tilde{p}_2 = \frac{4+2}{10+4} = .4286$$

so that approximate 95% confidence limits for the difference $p_1 - p_2$ are

$$.2 - .4 \pm 1.96\sqrt{\frac{.2857\,(1-.2857)}{10} + \frac{.4286\,(1-.4286)}{10}}$$

i.e.

$$-.2 \pm .49$$

These limits cover 0 and there is no clear evidence in the $\hat{p}_1 = .2$ and $\hat{p}_2 = .4$ values from the relatively small samples of sizes $m = 10$ that Operators 1 and 2 have different probabilities of calling Part 1 non-conforming.