Approximate Variance - Covariance Matrix for $\underset{\sim}{Y}$?

Variance-covariance matrix for $\underset{\sim}{Y}$ $\approx \underset{\sim}{D} \, \underset{\sim}{\Sigma} \, \underset{\sim}{D}'$

where $\underset{\sim}{\Sigma}$ = covariance matrix for $\begin{pmatrix} S_y^2 \\ s^2 \end{pmatrix}$

$$= \text{diag}\left(\frac{2\sigma_y^4}{n-1}, \frac{2\sigma_{measurement}^4}{m-1} \right)$$

and

$$\underset{\sim}{D} = \begin{pmatrix} \left.\frac{\partial g}{\partial t_1}\right|_{\sigma_y^2, \sigma_{meas}^2} & \left.\frac{\partial g}{\partial t_2}\right|_{\sigma_y^2, \sigma_{meas}^2} \\ \left.\frac{\partial f}{\partial t_1}\right|_{\sigma_y^2, \sigma_{meas}^2} & \left.\frac{\partial f}{\partial t_2}\right|_{\sigma_y^2, \sigma_{meas}^2} \end{pmatrix}$$

Usefulness?

$\underset{\sim}{D} \underset{\sim}{\Sigma} \underset{\sim}{D}'$ measures variability in $\begin{pmatrix} \hat{\sigma}_x \\ s \end{pmatrix}$

- The "bigger" this is the worse off I am in estimating $\begin{pmatrix} \sigma_x \\ \sigma_{meas} \end{pmatrix}$

- $\det \underset{\sim}{D} \underset{\sim}{\Sigma} \underset{\sim}{D}' = K(\sigma_x^2, \sigma_{meas}^2, n, m)$ is a one-number summary of how well I'm going to do using $\underset{\sim}{Y}$ to estimate

- Useful in planning data collection: with planning values for variances plugged in I have a method for comparing various (n, m) choices.

- Post-data I can use $K(\sigma_x^2, \sigma_{meas}^2, n, m)$ to get an estimate of the determinant of the variance-covariance matrix as $K(\hat{\sigma}_x^2, s^2, n, m)$.

Range-Based Version of All This

$X_1, X_2, \cdots, X_n$ iid $N(\mu, \sigma^2)$

$$E\,\frac{R}{d_2(n)} = \sigma$$

In our scenario, n widgets produce $y_1, \cdots, y_n$ and a single additional widget produces $y_1', \cdots, y_m'$. We can make

$$R = \max y - \min y$$

$$R' = \max y' - \min y'$$

Note that

$$E\,\frac{R}{d_2(n)} = \underbrace{\sigma_y = \sqrt{\sigma_x^2 + \sigma_{meas.}^2}}$$

$$E\,\frac{R'}{d_2(m)} = \sigma_{measurement}.$$

So perhaps I might

estimate σ_y with $\dfrac{R}{d_2(n)}$

estimate σ_{meas} with $\dfrac{R'}{d_2(m)}$

and create an estimator of σ_x as

$$\tilde{\sigma}_x = \sqrt{\max\left(0, \left(\frac{R}{d_2(n)}\right)^2 - \left(\frac{R'}{d_2(m)}\right)^2\right)}$$

and I can go through steps parallel to what I did for $\hat{\sigma}_x$ to find an approximate standard deviation for $\tilde{\sigma}_x$ and (upon plugging estimates in for σ_x^2 and $\sigma_{measurement}^2$) a standard error.

A more conventional approach to estimating σ_x and σ_{meas} is through balanced 1-way ANOVA (using a simple "random effects model")

widget 1   widget 2 $\cdots$ widget r

$y_{11}\ y_{12} \cdots y_{1m}$ $y_{21}\ y_{22} \cdots y_{2m}$ $y_{r1}\ y_{r2} \cdots y_{rm}$

$$y_{ij} = \text{measurement } j \text{ on widget } i$$

Model $\quad y_{ij} = \underset{\text{iid } N(\mu_x,\, \sigma_x^2)}{x_i} + \underset{\text{iid } N(0,\, \sigma^2)}{\varepsilon_{ij}}$

$\sigma^2 = \sigma^2_{measurement}$

x_i's and ε_{ij}'s are independent.

Let $\alpha_i = x_i - \mu_x$

then $y_{ij} = \mu_x + \underset{\text{iid } N(0,\, \sigma_x^2)}{\alpha_i} + \varepsilon_{ij}$

Let $\overline{y_i} = \frac{1}{m} \sum_j y_{ij}$, $\quad s_i^2 = \frac{1}{m-1} \sum_j (y_{ij} - \overline{y_i})^2$.

Then $\overline{y}_1, s_1^2, \overline{y}_2, s_2^2, \cdots, \overline{y}_r, s_r^2$ all are independent random variables, and

$$\frac{(m-1)s_i^2}{\sigma^2} \sim \chi^2_{m-1}.$$

Since sums of independent χ^2 r.v.'s are χ^2 (where d.f.'s add)

$$\frac{r(m-1)}{\sigma^2} \cdot \underbrace{\frac{1}{r} \sum_{i=1}^{r} s_i^2} \sim \chi^2_{r(m-1)}.$$

So $E\left(\frac{1}{r}\sum_{i=1}^{r} s_i^2\right) = \sigma^2$ and $\frac{1}{r}\sum_{i=1}^{r} s_i^2$ is a plausible estimator of $\sigma^2 = \sigma^2_{measurement}$.

$\frac{1}{r}\sum_{i=1}^{r} s_i^2$: MSE from one-way ANOVA

— like a sample variance from a normal population with variance σ^2 where the sample size is $r(m-1)+1$.

$MSE = \frac{1}{r}\sum_{i=1}^{r} s_i^2$ is independent of a "sample variance" computed from $\overline{y}_i$'s:

$$\frac{\sum_{i=1}^{r}(\overline{y}_i - \overline{y}_.)^2}{r-1} = \frac{MSTr}{m}$$ in ANOVA notation

where $$\overline{y}_. = \frac{1}{r}\sum_{i=1}^{r} \overline{y}_i .$$

Note that

$$\underbrace{\frac{(r-1)\left(\frac{MSTr}{m}\right)}{\sigma_x^2 + \frac{\sigma^2}{m}}}\sim \chi^2_{r-1}$$

$\bar{Y}_i$ have variance $\sigma_x^2 + \frac{\sigma^2}{m}$.

So $\frac{MSTr}{m}$ is a plausible estimator of $\sigma_x^2 + \frac{\sigma^2}{m}$.

So another estimator of σ_x is

$$\sqrt{\max\left(0, \frac{MSTr}{m} - \frac{MSE}{m}\right)} = \check{\sigma}_x .$$

$\check{\sigma}_x$ has essentially the same structure as $\hat{\sigma}_x$. So the same arguments that produced delta method approximate variances for $\hat{\sigma}_x$ and eventually std. errors can be applied here too!

General Structure of What We've Been Doing

$S_1^2, S_2^2, \cdots, S_K^2$ independent

with $\frac{\nu_i S_i^2}{ES_i^2} \sim \chi^2_{\nu_i}$ $\quad i=1,2,\cdots,K$

What's of interest is $\theta = g(ES_1^2, ES_2^2, \cdots, ES_k^2)$

We've been estimating θ by

$$U = g(S_1^2, S_2^2, \cdots, S_K^2).$$

Note that the Δ method gives

$$\text{Var}\, U \approx \sum_{i=1}^{K} \left(\frac{\partial g}{\partial x_i}\Big|_{ES_1^2, \cdots, ES_K^2} \right)^2 \cdot \underbrace{\frac{2(ES_i^2)^2}{\nu_i}}_{\text{Var}(S_i^2)}.$$

Therefore

$$\widehat{\text{Var}\, U} = \sum_{i=1}^{K} \left(\frac{\partial g}{\partial x_i}\Big|_{S_1^2, \cdots, S_K^2} \right)^2 \cdot \frac{2(S_i^2)^2}{\nu_i},$$

which is formula (1.3) of the notes.

Confidence intervals for θ?

See the book "Confidence Intervals on Variance Components" by Burdick and Graybill:

Modified Large Sample (MLS) Method

- Section 1.5 of the notes reproduces the MLS method for linear g

- See program of Brandon Paris on Professor Vardeman's Stat 531 Web page for implementing

- Andy Chiang has developed a replacement for the MLS stuff.

All of this is an amplification of pages 19 & 20 of V&J.

Now proceed to a 531 version of §2.2.2 of V&J — "Gage R&R Studies"

Observed variation

"real" or "widget" or "unit-to-unit" variation

"measurement" variation

reproducibility variation

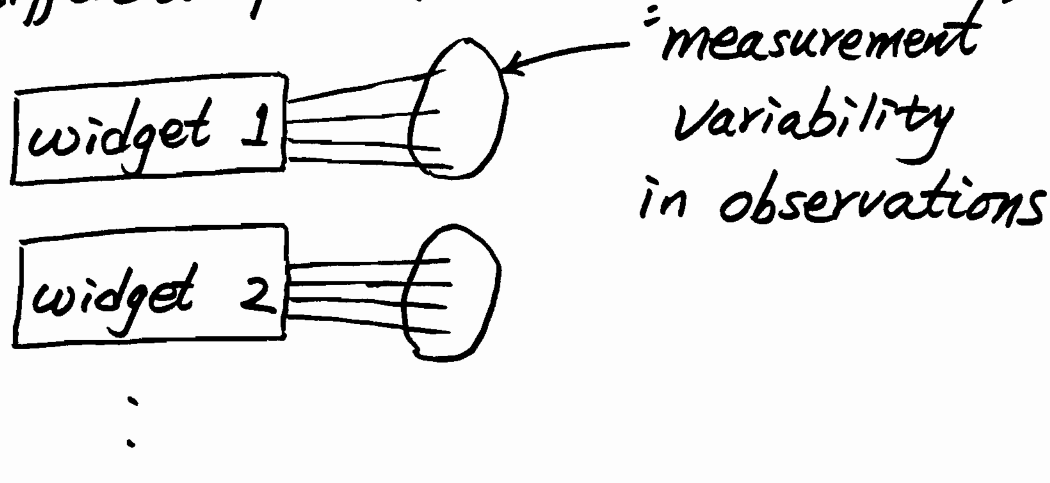
"between operator" variation

repeatability variation

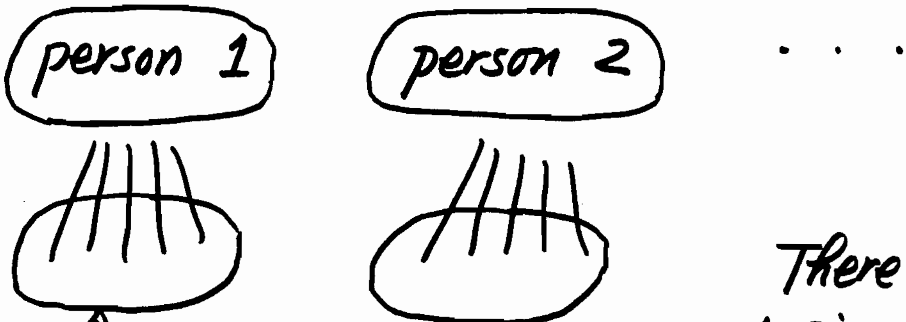
"within operator" variation

Motivation

— If a single person measures different parts:



— If multiple people measure the same part:



variation here is what we've been calling σ or $\sigma_{\text{measurement}}$

There will also typically be person-to-person variability in a set-up like this.

It is "clear" how to take the "one-person / many parts" material just discussed and make estimates (based on "one-part / many-people" data) of

$\sigma_{\text{repeatability}}$ and $\sigma_{\text{reproducibility}}$

σ, $\sigma_{\text{measurement}}$

between operators

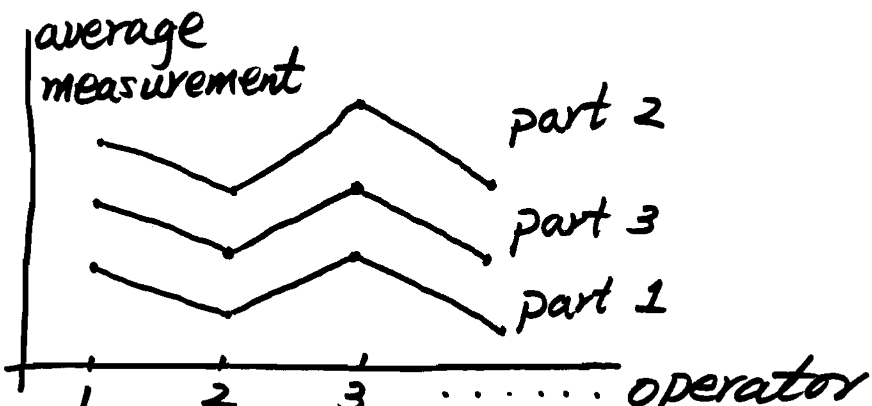
analog of σ_x

What's not completely clear is how to integrate these notions into a coherent single story.

One might hope I could model things this way:

$$\text{measured value} = \text{an overall average} + \underset{\text{0 mean}}{\text{"part" effect}} + \underset{\text{0 mean}}{\text{"operator" effect}} + \underset{\text{0 mean}}{\text{"measurement error"}}$$

But this implies that if I have I parts and J operators and suppose that measurement error is small.



This is a "no-interaction" picture between parts and operators

To allow for lack of parallelism (often empirically present) in your description of measurement, include in your model a term peculiar to a given part×operator combination — simplest version: (2.4) of V&J:

y_{ijk} = kth observation on ith part by jth operator

$$y_{ijk} = \mu + \underbrace{\alpha_i}_{\text{part effect}} + \underbrace{\beta_j}_{\text{operator effect}} + \underbrace{\alpha\beta_{ij}}_{\text{part}\times\text{operator interaction}} + \underbrace{\varepsilon_{ijk}}_{\text{measurement error}}$$

where all of the $\underbrace{\alpha_i, \beta_j, \alpha\beta_{ij} \text{ and } \varepsilon_{ijk}}_{\text{random effects}}$ are independent normal r.v.'s all with mean 0 and variances

$$\boxed{\begin{matrix} \operatorname{Var}\alpha_i = \sigma_\alpha^2 & \operatorname{Var}\beta_j = \sigma_\beta^2 \\ \operatorname{Var}\alpha\beta_{ij} = \sigma_{\alpha\beta}^2 & \operatorname{Var}\varepsilon_{ijk} = \sigma^2 \end{matrix}} \leftarrow \text{variance components}$$

parametric functions of these are of interest

— e.g. $\sigma^2 \leftarrow$ measures "repeatability"

		Operator 1	Operator 2	. . .
part	1	$y_{11k}=\mu+\alpha_1+\beta_1+\alpha\beta_{11}+\varepsilon_{11k}$	$y_{12k}=\mu+\alpha_1+\beta_2+\alpha\beta_{12}+\varepsilon_{12k}$	
	2	$y_{21k}=\mu+\alpha_2+\beta_1+\alpha\beta_{21}+\varepsilon_{21k}$		
	⋮			

Thinking of Column 1 as fixed (β_1 is fixed)

$$\underbrace{\text{Cond. Var } y_{i11}} = \underbrace{\overbrace{\sigma_\alpha^2+\sigma_{\alpha\beta}^2}^{\sigma^2_{parts}}}_{\sigma_x^2} + \underbrace{\sigma^2}_{\sigma^2_{measurement}}$$

This is what a sample variance from 1st column (1 measurement per cell) would estimate) σ_y^2

Thinking of Row 1 as fixed (α_1 is fixed)

$$\underbrace{\text{Cond. Var } y_{1j1}} = \overbrace{\underbrace{\sigma_\beta^2 + \sigma_{\alpha\beta}^2}_{\sigma^2_{\text{reproducibility}}} + \underbrace{\sigma^2}_{\sigma^2_{\text{measurement}}}}^{\sigma^2_{\text{overall (R\&R)}}}$$

operator to operator variation

This is what a sample variance from the 1st row (1 measurement per cell) would estimate.

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

$$\sigma_x^2 \rightarrow \sigma^2_{\text{parts}} = \sigma_\alpha^2 + \sigma_{\alpha\beta}^2$$

Others prefer to call σ_α^2 (only) "part" or "process" variation (and thus charge $\sigma_{\alpha\beta}^2$ to "operators" only)

Vardeman & VanValkenburg (1999. Technometrics) note that an argument can be made for the standard point of view.