We've dealt with MS's before.

$$SSA = JK\sum_i (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 \qquad (d.f. = I-1)$$

$$SSB(A) = K\sum_{i,j} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 \qquad (d.f. = I(J-1))$$

$$\underset{\underset{SSE}{\parallel}}{SSC(B(A))} = \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij\cdot})^2 \qquad (d.f. = IJ(K-1))$$

Under the random effects model

$$\frac{SSA}{EMSA}, \frac{SSB(A)}{EMSB(A)}, \frac{SSC(B(A))}{EMSC(B(A))}$$

are independent χ^2 r.v. with d.f.

and $EMSC(B(A)) = \sigma^2$

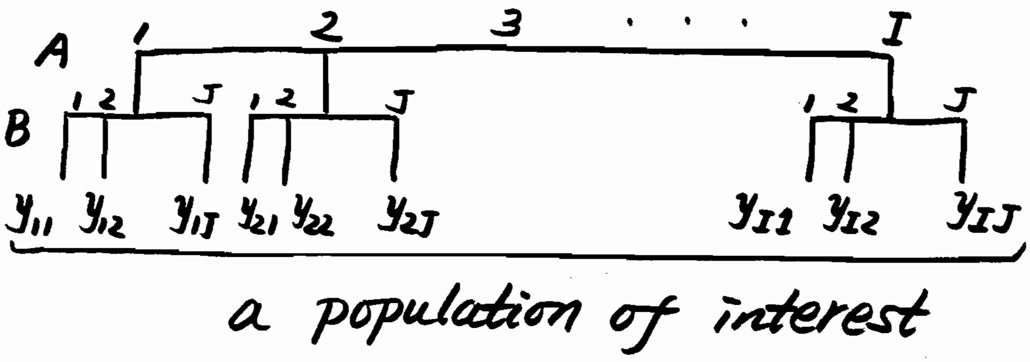
$$EMSB(A) = \sigma^2 + K\sigma_\beta^2$$

$$EMSA \quad = \sigma^2 + K\sigma_\beta^2 + JK\sigma_\alpha^2.$$

Obviously the variance components $\sigma^2, \sigma_\beta^2, \sigma_\alpha^2$ are simple (linear) functions of EMS's. Thus all the stuff from early in the term (introduced in context of gage R&R studies and 2-way model) is relevant here.

- Direct inference for variance components is aimed at the process that generates the y_{ijk}'s.
- Another possible use of this material is in trying to make statements about variation in an existing structured data set.

A simple 2-level illustration:



I might want some idea of variability in that population without having to look at all IJ y's.

— Suppose I sample $I' < I$ levels of A and $J' < J$ levels of B within each of these.

<u>Note</u>: s^2 = grand sample variance for all IJ members of the population

$$= \frac{1}{IJ-1}\left(\text{population "total sum of squares"}\right)$$

$$= \frac{1}{IJ-1}\left(\text{pop. "SSTr"} + \text{pop. "SSE"}\right).$$

Assume that the hierarchical random effects model describes the <u>IJ y's</u>.

$$Es^2 = \frac{1}{IJ-1}\left[(I-1)E\Big(\underbrace{\text{population } MSA}_{\text{"}MSTr\text{"}}\Big) + I(J-1)E\Big(\underbrace{\text{pop. } MSB(A)}_{\text{"}MSE\text{"}}\Big)\right]$$

$$= \frac{1}{IJ-1}\left[(I-1)(\sigma^2 + J\sigma_\alpha^2) + I(J-1)\sigma^2\right].$$

From $I'J'$ data values I can estimate σ^2 and σ_α^2 and thus come up with a sensible predictor for s^2 (by plugging into right hand side above).

Another way to look at the problem is to think of the IJ y's as fixed unknown numbers.

— Drop the assumption that they are generated from the hierarchical random effects model.

— Use simple random sampling of levels of A and levels of B within A and apply finite population sampling theory. (This is "subsampling" in Cochran terminology)

— An unbiased estimator of S^2 is

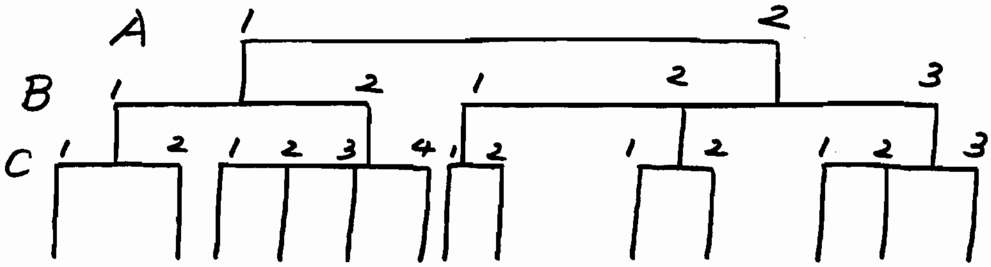
$$\hat{S}^2 = \frac{J(I-1)}{IJ-1}\left(\frac{\text{sample MSA}}{J'}\right) \qquad \text{"MSTr"}$$

$$+\left[\frac{I(J-1)}{IJ-1} - \frac{J(I-1)}{IJ-1}\left(\frac{1}{J'}-\frac{1}{J}\right)\right]\left(\begin{array}{c}\text{sample}\\ \text{MSB(A)}\end{array}\right)$$

"MSE"

Some additional observations having to do with an unbalanced hierarchical data structure (sample and/or population)

Content of problems 4.4 & 4.8 of the notes.



• How to estimate $\sigma_\alpha^2, \sigma_\beta^2, \sigma^2$ from the hierarchical random effects model?

Fact: $X_1, X_2, \cdots, X_n$ independent r.v.'s with mean μ.

$$S^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$$

$$ES^2 = \frac{1}{n} \sum_{i=1}^{n} Var X_i$$

This enables simple-minded (from 1st principles) estimation of $\sigma_\alpha^2, \sigma_\beta^2, \sigma^2$ from unbalanced data.

Let S_{ij}^2 = sample variance at level i of A and level j of B within A

So $ES_{ij}^2 = \sigma^2$ and thus, e.g.,

$$\text{"MSE"} = S^2_{pooled} = \frac{1 \cdot S_{11}^2 + 3 \cdot S_{12}^2 + 1 \cdot S_{21}^2 + 1 \cdot S_{22}^2 + 2 \cdot S_{23}^2}{1 + 3 + 1 + 1 + 2}$$

is a perfectly sensible estimator of σ^2.

How about estimating σ_β^2?

Let $\overline{y}_{ij}$ = average of observations at level i of A and level j of B within A.

S_i^2 = sample variance of $\overline{y}_{ij}$'s.

Note that

$$\bar{y}_{11} = \mu + \alpha_1 + \beta_{11} + \bar{\varepsilon}_{11} \longleftarrow \text{an average of 2 } (\varepsilon_{111}, \varepsilon_{112})$$

$$\bar{y}_{12} = \mu + \alpha_1 + \beta_{12} + \bar{\varepsilon}_{12} \longleftarrow \text{an average of 4}$$

differ only in β's and $\bar{\varepsilon}$'s.

Thus $$ES_1^2 = \frac{1}{2}\left(\sigma_\beta^2 + \frac{\sigma^2}{2} + \sigma_\beta^2 + \frac{\sigma^2}{4}\right)$$

$$= \sigma_\beta^2 + \frac{3}{8}\sigma^2.$$

Similarly we can obtain ES_2^2 as follows.

Note that

$$\bar{y}_{21} = \mu + \alpha_2 + \beta_{21} + \bar{\varepsilon}_{21} \text{ — an average of 2}$$

$$\bar{y}_{22} = \mu + \alpha_2 + \beta_{22} + \bar{\varepsilon}_{22} \text{ — an average of 2}$$

$$\bar{y}_{23} = \mu + \alpha_2 + \beta_{23} + \bar{\varepsilon}_{23} \text{ — an average of 3}$$

So $$ES_2^2 = \frac{1}{3}\left(\sigma_\beta^2 + \frac{\sigma^2}{2} + \sigma_\beta^2 + \frac{\sigma^2}{2} + \sigma_\beta^2 + \frac{\sigma^2}{3}\right)$$

$$= \sigma_\beta^2 + \frac{4}{9}\sigma^2.$$

So a convex combination of S_1^2 and S_2^2 can be used to estimate

$$\sigma_\beta^2 + (\text{constant}) \cdot \sigma^2. \quad \left(\text{Recall that } ES_1^2 = \sigma_\beta^2 + \frac{3}{8}\sigma^2.\right)$$

Thus after subtracting (constant)·MSE one can have an estimator of σ_β^2.

That is, for a given $\rho \in (0,1)$.

$$E(\rho s_1^2 + (1-\rho)s_2^2) = \sigma_\beta^2 + (\rho \cdot \tfrac{3}{8} + (1-\rho)\cdot \tfrac{4}{9})\sigma^2$$

$$\Rightarrow E\left((\rho s_1^2 + (1-\rho)s_2^2) - (\rho \cdot \tfrac{3}{8} + (1-\rho)\cdot\tfrac{4}{9})\cdot S_{pooled}^2\right)$$

$$= \sigma_\beta^2$$

$$\Rightarrow \rho s_1^2 + (1-\rho)s_2^2 - [\rho \cdot \tfrac{3}{8} + (1-\rho)\cdot \tfrac{4}{9}]\cdot S_{pooled}^2$$

is a sensible (unbiased) estimator of σ_β^2.

A similar story can be told for estimating σ_α^2.

Let $\bar{y}_1 = \frac{1}{2}(\bar{y}_{11} + \bar{y}_{12})$

$$= \mu + \alpha_1 + \bar{\beta}_1 + \tfrac{1}{2}(\bar{\varepsilon}_{11} + \bar{\varepsilon}_{12})$$

an average of 2

$$\bar{y}_2 = \tfrac{1}{3}(\bar{y}_{21} + \bar{y}_{22} + \bar{y}_{23})$$

$$= \mu + \alpha_2 + \bar{\beta}_2 + \tfrac{1}{3}(\bar{\varepsilon}_{21} + \bar{\varepsilon}_{22} + \bar{\varepsilon}_{23})$$

an average of 3

s^2 = sample variance of $\bar{y}_1$ and $\bar{y}_2$.

Note that

$$Es^2 = \frac{1}{2}\left(\sigma_\alpha^2 + \frac{1}{2}\sigma_\beta^2 + \frac{3}{16}\sigma^2 + \sigma_\alpha^2 + \frac{1}{3}\sigma_\beta^2 + \frac{4}{27}\sigma^2\right)$$

$$= \sigma_\alpha^2 + \frac{5}{12}\sigma_\beta^2 + \frac{145}{864}\sigma^2.$$

Thus $s^2 - \frac{5}{12}\hat{\sigma}_\beta^2 - \frac{145}{864}s^2_{pooled}$

is a plausible/possible estimator of σ_α^2.

<u>Problem 4.4</u> of the notes:

estimation of population means or proportions when there is lack of balance.

sub population	N_1	N_2	$\cdots$	N_r
sample	n_1	n_2		n_r
mean for sub population	$\overline{Y_1}$	$\overline{Y_2}$		$\overline{Y_r}$
0-1 case	P_1	P_2		P_r

Grand mean $\bar{Y} = \sum N_i \bar{Y}_i / \sum N_i$

Grand proportion $p = \sum N_i p_i / \sum N_i$

The "correct" way to estimate $\bar{Y}$ or p is by weighting sample means or proportions by $(N_i/\sum N_j)$'s.

$$\begin{array}{cccc} \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_r \\ \hat{p}_1 & \hat{p}_2 & \cdots & \hat{p}_r \end{array}$$

$$\hat{\bar{Y}} = \sum \left(\frac{N_i}{\sum N_j}\right) \bar{y}_i \qquad \hat{p} = \sum \left(\frac{N_i}{\sum N_j}\right) \hat{p}_i$$

Ch. 8 of V&J / Ch. 5 of the Notes /

"The Legitimate Role of Inspection ..."

What does statistics have to say about product-oriented inspection?

<u>sampling inspection</u> vs. <u>100% inspection</u>

statistical theory and methods deal with uncertainty introduced by looking at only "some" products	statistics is relevant only because inspection (measurement) is a noisy affair

A fundamental issue is:

When does product-oriented inspection on a sampling basis make sense?

The answer:

Only when quality is unknown/varies.

Vander Wiel/Vardeman argument:

Suppose we face N items, which are iid in terms of G/D (Good/Defective) natures.

O O Ø O O . . . Ø O

Let p = probability item 1 is D.

Suppose further that inspection can be faulty:

w_G = probability a G item fails

w_D = probability a D item passes

Further assume that costs are accrued on a per-item basis according to:

	Item Condition: G	D
Pass	K_I	$K_I + K_{DP}$
Fail	$K_I + K_{GF}$	$K_I + K_{DF}$
No Inspection	0	K_{DU}

Consider (for purposes of exposition) "random inspection policies" that inspect any item with probability π.

The mean cost incurred over N items is the same as the expected cost suffered for one item:

$$
\begin{aligned}
ECost &= p(1-\pi)K_{DU} + \pi\left[K_I + p\,W_D K_{DP} \right.\\
&\qquad \left. + (1-p)\,W_G K_{GF} + p(1-W_D)K_{DF}\right] \\
&= p K_{DU} + \pi\left[K_I + W_G K_{GF} - p\underline{\underline{K}}\right]
\end{aligned}
$$

where

$$
\begin{aligned}
K &= (1-W_D)(K_{DU} - K_{DF}) + W_D(K_{DU} - K_{DP}) \\
&\qquad + W_G K_{GF}.
\end{aligned}
$$

If $K<0$ ECost is minimized for $\pi=0$;

if $K>0$ ECost is minimized

for $\pi=1$ if $p \geq \dfrac{K_I + W_G K_{GF}}{K}$

for $\pi=0$ if $p \leq \dfrac{K_I + W_G K_{GF}}{K}$.

Define

$$P_c = \begin{cases} \infty & \text{if } K \le 0 \\ \dfrac{K_I + W_G K_{GF}}{K} & \text{if } K > 0. \end{cases}$$

The best random inspection policy has

$\pi = 0$ (no inspection) if $P < P_c$

$\pi = 1$ (100% inspection) if $P \ge P_c$.

Note that implementing this requires knowledge of P.