hw_lect 24-1

Show that for the "sample variance" of $\hat{\tau}_i$, $s_\tau^2 = \frac{1}{a-1}\sum(\hat{\tau}_i - \bar{\hat{\tau}}_\cdot)^2$, where $\hat{\tau}_i$ is defined by $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$, we have $E[s_\tau^2] = \sigma_\tau^2 + \frac{1}{n}\sigma_\epsilon^2$.

Hint: This is nearly trivial if you use what we showed in lect ie. $E[MS_{Tr}] = \sigma_\epsilon^2 + n\sigma_\tau^2$.

$$E[s_\tau^2] = E\left[\frac{1}{a-1}\sum_i(\hat{\tau}_i - \bar{\hat{\tau}}_\cdot)^2\right] = \frac{1}{a-1}\sum_i E\left[(\hat{\tau}_i - \bar{\hat{\tau}}_\cdot)^2\right]$$

defn. of $\hat{\tau}_i$

$$= \frac{1}{a-1}\sum_i E\left[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) - (\cancel{\bar{Y}_{\cdot\cdot} - \bar{Y}_{\cdot\cdot}})^{\to 0}\right]^2$$

$$= \frac{1}{a-1}\sum_i E\left[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right]$$

lecture.

$$= \frac{1}{n} E[MS_{Tr}]$$

$$= \frac{1}{n}\left(\sigma_\epsilon^2 + n\sigma_\tau^2\right) = \sigma_\tau^2 + \frac{1}{n}\sigma_\epsilon^2.$$

Note: Suppose you didn't remember to use the result from lect. Then

$$E[s_\tau^2] = \cdots = \frac{1}{a-1}\sum_i E[(\hat{\tau}_i - \bar{\hat{\tau}}_\cdot)^2] = \cdots = \frac{1}{a-1}\sum_i E[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2]$$

$$= E[(\mu + \tau_i + \bar{\epsilon}_{i\cdot} - \mu - \bar{\tau}_\cdot - \bar{\epsilon}_{\cdot\cdot})^2] = \cdots \text{ as in lect.}$$

Of course, you will get the same answer, but you may get confused because we got rid of $\hat{\tau}_i$ here, and then re-introduced it here.

There is really no such thing happening! The 2 τ's are different. $\hat{\tau}_i$ is <u>defined</u> by $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$, whereas the other τ_i is a r.v.

hw-lect 24-2

Without using the expression we derived in class for $E[MS_{Tr}]$, show that $E[SST] = (an-1)\sigma_\epsilon^2 + (a-1)n\,\sigma_\tau^2$. This calculation is similar to that of the lecture.

$$E[SS_T] = E\Big[\sum_{ij}(y_{ij}-\bar{y}_{..})^2\Big] = E\Big[\sum_{ij}(\cancel{\mu}+\tau_i+\epsilon_{ij}-\cancel{\mu}-\bar{\tau}_{.}-\bar{\epsilon}_{..})^2\Big]$$

$$= \sum_{ij}\underbrace{E[(\epsilon_{ij}-\bar{\epsilon}_{..})^2]} + \sum_{ij}\underbrace{E[(\tau_i-\bar{\tau}_{.})^2]} - 2\sum_{ij}\underbrace{E[(\tau_i-\bar{\tau}_{.})(\epsilon_{ij}-\bar{\epsilon}_{..})]}$$

$\epsilon \perp \tau$: $E[\tau_i-\bar{\tau}_{.}]\,E[\epsilon_{ij}-\bar{\epsilon}_{..}]$

$$\underbrace{V[\epsilon_{ij}-\bar{\epsilon}_{..}]} + \underbrace{E^2[\epsilon_{ij}-\bar{\epsilon}_{..}]}_{0 = E[\epsilon_{ij}]}$$

$$\underbrace{V(\epsilon_{ij})}_{\sigma_\epsilon^2} + \underbrace{V[\bar{\epsilon}_{..}]}_{\frac{\sigma_\epsilon^2}{an}} - 2\underbrace{Cov[\epsilon_{ij},\bar{\epsilon}_{..}]}$$

$$\frac{1}{an}\sum_{kl}\underbrace{Cov[\epsilon_{ij},\epsilon_{kl}]}_{V[\epsilon_{ij}]}$$

only $i=k, j=l$ survives

$$\underbrace{V[\tau_i-\bar{\tau}_{.}]} + \underbrace{E^2[\tau_i-\bar{\tau}_{.}]}_{0 = E[\tau_i]}$$

$$\underbrace{V[\tau_i]}_{\sigma_\tau^2} + \underbrace{V[\bar{\tau}_{.}]}_{\frac{\sigma_\tau^2}{a}} - 2\underbrace{Cov[\tau_i,\bar{\tau}_{.}]}$$

$$\frac{1}{a}\underbrace{\sum_j Cov[\tau_i,\tau_j]}_{V[\tau_i]}$$

$$= an\left[1+\frac{1}{an}-\frac{2}{an}\right]\sigma_\epsilon^2 + an\left[1+\frac{1}{a}-\frac{2}{a}\right]\sigma_\tau^2$$

$$= \underline{(an-1)\,\sigma_\epsilon^2 + (a-1)n\,\sigma_\tau^2}$$

Note: This is the answer we would get if we do use the result derived in lecture for $E[MS_{Tr}]$

$$E[SST] = E[SS_{Tr}] + E[SSE] = (a-1)E[MS_{Tr}] + a(n-1)\,E[MSE]$$

$$= (a-1)\left[\sigma_\epsilon^2 + n\sigma_\tau^2\right] + a(n-1)\,\sigma_\epsilon^2 = (an-1)\sigma_\epsilon^2 + n(a-1)\sigma_\tau^2.$$

hw, lect 24-3

Use the result $E[MS_{Tr}] = \sigma_\epsilon^2 + n\sigma_\tau^2$ to show that if the model of the data is $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, with $\tau_i \sim N(0, \sigma_\tau^2)$ and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, then, the sample variance of y (ie. s_y^2) is a biased estimator of σ_y^2, where $\sigma_y^2 = \sigma_\tau^2 + \sigma_\epsilon^2$ is the pop. variance of y. Hint: start with the decomposition of SST.

$$SST = SS_{Tr} + SS_E$$

$$\frac{1}{an-1} SST = \frac{1}{an-1} SS_{Tr} + \frac{1}{an-1} SS_E \Rightarrow s_y^2 = \frac{1}{an-1} SS_{Tr} + \frac{1}{an-1} SS_E$$

$$E[s_y^2] = \frac{1}{an-1} E[SS_{Tr}] + \frac{1}{an-1} E[SS_E]$$

$$= \frac{a-1}{an-1} E[MS_{Tr}] + \frac{a(n-1)}{an-1} E[MS_E]$$

$$= \frac{a-1}{an-1} \cdot (\sigma_\epsilon^2 + n\sigma_\tau^2) + \frac{a(n-1)}{an-1} \sigma_\epsilon^2$$

$$E[s_y^2] = \sigma_\epsilon^2 + \frac{n(a-1)}{an-1} \sigma_\tau^2 = \sigma_y^2 - \sigma_\tau^2 + \frac{n(a-1)}{an-1} \sigma_\tau^2$$

$$= \sigma_y^2 - \frac{n-1}{an-1} \sigma_\tau^2 \neq \sigma_y^2$$

s_y^2 is a biased estimator of σ_y^2.

In fact, $E[s_y^2] < \sigma_y^2 \Rightarrow s_y^2$ under-estimates σ_y^2.

hw-lect 24-4

In random effects models it is possible to get a negative value for $\hat{\sigma}_\tau^2$.

a) Consider a 1-factor problem, and show that

$$\text{prob}(\hat{\sigma}_\tau^2 < 0) = \text{prob}\left(F_{a-1,a(n-1)} < \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + n\sigma_\tau^2}\right)$$

b) This result is helpful in that it tells how surprised we should be if/when we see a negative $\hat{\sigma}_\tau^2$. For example, if you suspect $\sigma_\tau^2/\sigma_\epsilon^2$ is around 0.5, what is the prob$(\hat{\sigma}_\tau^2 < 0)$, when $n=10$, $a=3$?

You may use R to compute a numerical answer.

a) $$Pr(\hat{\sigma}_\tau^2 < 0) = Pr\left(\frac{MSTr - MS_E}{n} < 0\right) = Pr(MSTr < MS_E)$$

$$= Pr\left(\frac{MSTr}{MS_E} < 1\right)$$

$$= Pr\left(\underbrace{\frac{MSTr/E[MSTr]}{MS_E/E[MS_E]}}_{F_{a-1,a(n-1)}} < \frac{E[MS_E]}{E[MSTr]}\right)$$

$$= Pr\left(F_{a-1,a(n-1)} < \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + n\sigma_\tau^2}\right)$$

b) $$= pr\left(F_{a-1,a(n-1)} < \frac{1}{1+n\,\sigma_\tau^2/\sigma_\epsilon^2}\right)$$

$$= pr\left(F_{2,27} < \frac{1}{1+5}\right) = pr\left(F_{2,27} < \underset{.17}{\overset{\shortparallel}{\frac{1}{6}}}\right)$$

$$= \underset{\uparrow \text{ in R}}{pF(.17, 2, 27)} = \boxed{.155}$$

hw-lect 25-1

For The 2-factor random-effects model, $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$, with $k = 1 \cdots, n$ replications, show That $E[MS_A] = \sigma_\epsilon^2 + n\,\sigma_{\alpha\beta}^2 + bn\,\sigma_\alpha^2$.

You may use This result w/o proof: $Cov[\bar{\epsilon}_{i\cdot\cdot}, \bar{\epsilon}_{\cdot\cdot\cdot}] = \frac{\sigma_\epsilon^2}{abn}$.

$$E[MS_A] = \frac{1}{a-1} E[SS_A] = \frac{bn}{a-1} \sum_i^a E[(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2]$$

$$= \frac{bn}{a-1} \sum_i \left(V[\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}] + \underbrace{E^2[\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}]}_{0} \right)$$

$$= \frac{bn}{a-1} \sum_i V[\cancel{\mu} + \alpha_i + \cancel{\bar{\beta}_\cdot} + (\overline{\alpha\beta})_{i\cdot} + \bar{\epsilon}_{i\cdot\cdot} - \cancel{\mu} - \bar{\alpha}_\cdot - \cancel{\bar{\beta}_\cdot} - (\overline{\alpha\beta})_{\cdot\cdot} - \bar{\epsilon}_{\cdot\cdot\cdot}]$$

$$= \frac{bn}{a-1} \sum_i V[\alpha_i - \bar{\alpha}_\cdot + (\overline{\alpha\beta})_{i\cdot} - (\overline{\alpha\beta})_{\cdot\cdot} + \bar{\epsilon}_{i\cdot\cdot} - \bar{\epsilon}_{\cdot\cdot\cdot}]$$

$$= \frac{bn}{a-1} \sum_i^a \Big\{ \underbrace{V[\alpha_i]}_{\sigma_\alpha^2} + \underbrace{V[\bar{\alpha}_\cdot]}_{\sigma_\alpha^2/a} - 2\,Cov[\alpha_i, \bar{\alpha}_\cdot]$$

$$\hookrightarrow \frac{1}{a}\sum_\ell^a Cov[\alpha_i, \alpha_\ell] = \sigma_\alpha^2/a$$

$$+ \underbrace{V[(\overline{\alpha\beta})_{i\cdot}]}_{\sigma_{\alpha\beta}^2/b} + \underbrace{V[(\overline{\alpha\beta})_{\cdot\cdot}]}_{\sigma_{\alpha\beta}^2/ab} - 2\,Cov[(\overline{\alpha\beta})_{i\cdot}, (\overline{\alpha\beta})_{\cdot\cdot}]$$

$$\hookrightarrow \frac{1}{bab}\sum_{pqr}^{bab} Cov[(\alpha\beta)_{ip}, (\alpha\beta)_{qr}]$$

$$+ \underbrace{V[\bar{\epsilon}_{i\cdot\cdot}]}_{\frac{\sigma_\epsilon^2}{bn}} + \underbrace{V[\bar{\epsilon}_{\cdot\cdot\cdot}]}_{\frac{\sigma_\epsilon^2}{abn}} - 2\,Cov[\bar{\epsilon}_{i\cdot\cdot}, \bar{\epsilon}_{\cdot\cdot\cdot}] \Big\}$$

$$\hookrightarrow \frac{\sigma_\epsilon^2}{abn}$$

$$= \frac{bn}{a-1}\, a \left\{ \sigma_\alpha^2 \left(1 + \frac{1}{a} - \frac{2}{a}\right) + \frac{1}{b}\sigma_{\alpha\beta}^2 \left(1 + \frac{1}{a} - \frac{2}{a}\right) + \frac{1}{bn}\sigma_\epsilon^2 \left(1 + \frac{1}{a} - \frac{2}{a}\right) \right\}$$

$$= \frac{abn}{a-1} \left\{ \frac{a-1}{a}\sigma_\alpha^2 + \frac{a-1}{ab}\sigma_{\alpha\beta}^2 + \frac{a-1}{abn}\sigma_\epsilon^2 \right\}$$

$$\frac{1}{bab}\sum_{pr}^{bb} Cov[(\alpha\beta)_{ip}, (\alpha\beta)_{ir}]$$

$$= bn \left[\sigma_\alpha^2 + \frac{1}{b}\sigma_{\alpha\beta}^2 + \frac{1}{bn}\sigma_\epsilon^2 \right]$$

$$\frac{1}{ab}\sigma_{\alpha\beta}^2 = \frac{1}{bab}\sum_p^b Cov[(\alpha\beta)_{ip}, (\alpha\beta)_{ip}]$$

$$= \sigma_\epsilon^2 + n\,\sigma_{\alpha\beta}^2 + bn\,\sigma_\alpha^2$$

hw-lct 25-2

ByR

```
# Consider the situation in problem 13.1 .
# a) Write code to run a *fixed effects* full model, and identify the numerical values of MSA, MSB, MSAB, and MSE.
# b) In preparation for doing random effects modelling, from these MS values compute the appropriate F-ratios, and
# p-values for testing
# H0: sigma_A^2 = 0 , vs. H1: , vs. H1: sigma_A^2 > 0
# H0; sigma_B^2 = 0 , ...
# H0: sigma_AB^2 = 0 , ...
# c) Using the formulas in lecture compute the estimates of the four variance components.
# d) Is the sum of the four estimates approximately equal to the total variance of y (as it should be)?

 rm(list=ls(all=TRUE))
 y.m = matrix(c(
  50, 49, 50, 50, 48, 51,
  52, 52, 51, 51, 51, 51,
  53, 50, 50, 54, 52, 51,
  49, 51, 50, 48, 50, 51,
  48, 49, 48, 48, 49, 48,
  52, 50, 50, 52, 50, 50,
  51, 51, 51, 51, 50, 50,
  52, 50, 49, 53, 48, 50,
  50, 51, 50, 51, 48, 49,
  47, 46, 49, 46, 47, 48), ncol = 6, byrow=T)
  y = as.vector(t(y.m))
  part = as.factor(rep(1:10,each=6))
  oper = as.factor(rep(rep(1:2,each=3),10))

# a)
  a = 10
  b = 2
  n = 3
  summary.aov(lm(y~ part*oper))
  MSA = 11.002    # part
  MSB = 0.417     # oper
  MSAB = 0.602    # interaction
  MSE = 1.500     # error

# b)
  FA = MSA/MSAB                    # 18.27575
  FB = MSB/MSAB                    # 0.692691
  FAB = MSAB/MSE                   # 0.4013333
  dfA = a-1
  dfB = b-1
  dfAB = (a-1)*(b-1)
  dfE  = a*b*(n-1)
  pf(FA, dfA, dfAB, lower.tail = FALSE)    # 9.39006e-05
  pf(FB, dfB, dfAB, lower.tail = FALSE)    # 0.4267823
  pf(FAB, dfAB, dfE, lower.tail = FALSE)   # 0.9269563

# c)
  ( MSA - MSAB )/(b*n)       # var_alpha = 1.733333
  ( MSB - MSAB )/(a*n)       # var_beta =  -0.006166667
  ( MSAB - MSE )/n           # var_alpha_beta = -0.2993333
   MSE                       # var_epsilon = 1.5

# d)
  1.733333  - 0.006166667  - 0.2993333  + 1.5  # 2.93
# This is pretty close to the sample variance of the ys, but it's not equal to it
  var(y)                              # 2.79
# Note that some of the variance estimates are negative. We have alreadydiscussed this issue.
```