Stat 581, Homework Set #9 Solutions

(1.) A = machine ($a=4$), B(A) = operator ($b=3$), y = surface finish

(a) Factors A, B are <u>crossed</u> if the levels of B are the same at each level of A.

Factor B is <u>nested</u> within A if the levels of B are different for each of the levels of A.

(b) model: $y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk}$ $\begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n \end{cases}$

estimates: $\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}$, $\hat{\beta}_{j(i)} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot}$

(c) $F_A = 14.27$ ($P_A = .000$), $F_{B(A)} = 4.17$ ($P_B = .013$)

The experiment finds that machine has an effect on surface finish. Also, the experiment finds that operator within machine has an effect on surface finish.

(d) $\hat{\tau}_1 = 2.83$, $\hat{\tau}_2 = 17.33$ (*), $\hat{\tau}_3 = -3.33$, $\hat{\tau}_4 = -16.83$

$\hat{\beta}_{1(1)} = 1.83$ $\hat{\beta}_{1(2)} = 12.33$ (*) $\hat{\beta}_{1(3)} = 19.0$ $\hat{\beta}_{1(4)} = -4.5$

$\hat{\beta}_{2(1)} = 15.33$ (*) $\hat{\beta}_{2(2)} = -1.67$ $\hat{\beta}_{2(3)} = -8.0$ $\hat{\beta}_{2(4)} = -1.0$

$\hat{\beta}_{3(1)} = -17.17$ $\hat{\beta}_{3(2)} = -11.17$ $\hat{\beta}_{3(3)} = -11.0$ $\hat{\beta}_{3(4)} = 5.5$ (*)

(e) We cannot directly compare operators across machines since we only have data for operator performance with respect to a particular machine.

(2.) A = production process (fixed effect), B(A) = batch (random effect), y = burning rate (n=3)

(a) $MS_A = \dfrac{bn\sum_{i}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2}{a-1}$, $MS_{B(A)} = \dfrac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..})^2}{a(b-1)}$

$$MS_E = \frac{\sum_{i}\sum_{j}\sum_{k}(y_{ijk}-\bar{y}_{ij.})^2}{ab(n-1)}$$

(b) $E(MS_A) = \sigma^2 + n\sigma_\beta^2 + \dfrac{bn}{a-1}\sum_{i}^{a}\tau_i^2$

$E(MS_{B(A)}) = \sigma^2 + n\sigma_\beta^2$, $E(MS_E) = \sigma^2$

(c) $F_A = \dfrac{MS_A}{MS_{B(A)}} = 1.46$, $P_A = .281$

The experiment finds that production process does not have an effect on burning rate.

(d) We think of batch as the experimental unit. The appropriate error term is then a measure of batch variance. Taking repeat measurements from a selected batch does not increase the pertinent sample size.

(e) $MS_A = 338.0$, $MS_{B(A)} = 230.8$, $MS_E = 18.9$

Thus, $F_A < \dfrac{MS_A}{MS_E}$. The evidence in favor of a production process effect will be overstated when the incorrect error term is used.

(f) $\hat{\sigma}^2 = MS_E = 18.9$, $\hat{\sigma}_\beta^2 = \dfrac{MS_{B(A)} - MS_E}{n} = 70.6$

```
> library("readxl")
> setwd("C:/Users/aneath/iCloudDrive/Lexar/stat581 fall2021")
>
> hw9.data = read_excel("handout9data.xlsx")
> str(hw9.data)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':    36 obs. of  12 variables:
 $ supplier : num  1 1 1 1 1 1 1 1 1 1 ...
 $ batch    : num  1 1 1 2 2 2 3 3 3 4 ...
 $ purity   : num  94 92 93 91 90 89 91 93 94 94 ...
 $ machine  : num  1 1 1 1 1 1 1 1 2 2 ...
 $ spindle  : num  1 1 1 1 2 2 2 2 1 1 ...
 $ dimension: num  12 9 11 12 8 9 10 8 14 15 ...
 $ mchine   : num  1 1 1 1 1 1 2 2 2 2 ...
 $ operator : num  1 1 2 2 3 3 1 1 2 2 ...
 $ surface  : num  79 62 94 74 46 57 92 99 85 79 ...
 $ proc     : num  1 1 1 1 1 1 1 1 1 1 ...
 $ batc     : num  1 1 1 2 2 2 3 3 3 4 ...
 $ burn_rate: num  25 30 26 19 28 20 15 17 14 15 ...
>
> library("lme4")
> library("lmerTest")

>
> library("multcomp")
>
> machine = as.factor(na.omit(hw9.data$mchine))
> operator = as.factor(na.omit(hw9.data$operator))
> surface = na.omit(hw9.data$surface)
>
> contrasts(machine)=contr.sum
> contrasts(operator)=contr.sum
>
> nested.mod = aov(surface ~ machine/operator)
> summary(nested.mod)
                Df Sum Sq Mean Sq F value   Pr(>F)
machine          3   3618  1205.9  14.271 0.000291 ***
machine:operator  8   2818   352.2   4.168 0.013408 *
Residuals        12   1014    84.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> estimates = dummy.coef(nested.mod)
> estimates$machine
        1         2         3         4
 2.833333 17.333333 -3.333333 -16.833333
> estimates$`machine:operator`[c(1,5,9)]
      1:1       1:2       1:3
 1.833333 15.333333 -17.166667
> estimates$`machine:operator`[c(2,6,10)]
      2:1       2:2       2:3
12.333333 -1.166667 -11.166667
> estimates$`machine:operator`[c(3,7,11)]
3:1 3:2 3:3
 19  -8 -11
> estimates$`machine:operator`[c(4,8,12)]
 4:1  4:2  4:3
-4.5 -1.0  5.5
>
```

```
> compare.machine = glht(nested.mod,linfct = mcp(machine="Tukey"))
> c.m = summary(compare.machine,test=adjusted("none"))
> c.m

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = surface ~ machine/operator)

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0   14.500      5.307   2.732  0.01819 *
3 - 1 == 0   -6.167      5.307  -1.162  0.26785
4 - 1 == 0  -19.667      5.307  -3.706  0.00300 **
3 - 2 == 0  -20.667      5.307  -3.894  0.00213 **
4 - 2 == 0  -34.167      5.307  -6.438 3.22e-05 ***
4 - 3 == 0  -13.500      5.307  -2.544  0.02576 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- none method)

> cld(c.m)
  1   2   3   4
"b" "c" "b" "a"
>
>
>
> process = as.factor(na.omit(hw9.data$proc))
> batch = as.factor(na.omit(hw9.data$batc))
> rate = na.omit(hw9.data$burn_rate)
>
> contrasts(process)=contr.sum
> random.mod = lmer(rate ~ process + (1|batch:process))
>
> anova(random.mod)
Type III Analysis of Variance Table with Satterthwaite's method
        Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
process   55.4    27.7     2     9  1.4643 0.2815
> summary(random.mod)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModL
merTest']
Formula: rate ~ process + (1 | batch:process)

REML criterion at convergence: 222.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8162 -0.5073 -0.1282  0.4501  1.8625

Random effects:
 Groups         Name        Variance Std.Dev.
 batch:process (Intercept) 70.64    8.405
 Residual                  18.92    4.349
Number of obs: 36, groups:  batch:process, 12

Fixed effects:
            Estimate Std. Error     df t value Pr(>|t|)
(Intercept)   23.806      2.532  9.000   9.401 5.97e-06 ***
process1      -3.972      3.581  9.000  -1.109    0.296
process2      -2.056      3.581  9.000  -0.574    0.580
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
         (Intr) prcss1
process1  0.000
process2  0.000 -0.500
>
> nested.test = function(A,B,y)
+ {
+   av=anova(lm(y~A/B))
+   ss.A = av$`Sum Sq`[1]
+   ss.B = av$`Sum Sq`[2]
+   ss.error = av$`Sum Sq`[3]
+   df.A = av$Df[1]
+   df.B = av$Df[2]
+   df.error = av$Df[3]
+   ms.A = ss.A / df.A
+   ms.B = ss.B / df.B
+   ms.error = ss.error / df.error
+   F.a = ms.A / ms.B
+   p.value = pf(F.a,df1=df.A,df2=df.B,lower.tail = FALSE)
+   table1 = matrix(c(ss.A,ss.B,ss.error,
+                     df.A,df.B,df.error,
+                     ms.A,ms.B,ms.error),nrow = 3)
+   dimnames(table1) = list(c("Fixed Effect A","Random Effect B(A)","Error"),
+                           c("SS","df","MS"))
+   print(table1)
+
+   table2 = matrix(c(F.a,p.value),nrow = 1)
+   dimnames(table2) = list(c(""),c("F-test for fixed effect","p-value"))
+   print(table2)
+
+   a=nlevels(A)
+   b=nlevels(B)
+   n=length(y) / a / b
+
+   var.hat = ms.error
+   var.B.hat = (ms.B - ms.error) / n
+
+   table3 = matrix(c(var.hat,var.B.hat),nrow=1)
+   dimnames(table3) = list(c(""),c("error.var","B.var"))
+   print(table3)
+
+ }
>
>
> nested.test(process,batch,rate)
                          SS df        MS
Fixed Effect A      676.0556  2 338.02778
Random Effect B(A) 2077.5833  9 230.84259
Error               454.0000 24  18.91667
 F-test for fixed effect   p-value
                1.464322 0.2814697
 error.var    B.var
  18.91667 70.64198
```