STAT 511 **Solutions to Homework 10** **Spring 2003**

1. Use a small set of fake unbalanced two-way factorial data from 2 (unbalanced) blocks.

   (a) An estimate of σ is the residual standard error (= 0.394).

```
> summary(lmf.out)
lm(formula = y ~ A * B + FBlock)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.5929     0.1040  92.245 2.13e-13 ***
A1            1.2849     0.1040  12.356 1.72e-06 ***
B1           -0.2476     0.1473  -1.681    0.131
B2            1.5571     0.1541  10.104 7.86e-06 ***
FBlock1      -1.0286     0.1053  -9.768 1.01e-05 ***
A1:B1         0.2603     0.1487   1.751    0.118
A1:B2        -0.2349     0.1541  -1.524    0.166

Residual standard error: 0.394 on 8 degrees of freedom
Multiple R-Squared: 0.9793,     Adjusted R-squared: 0.9638
F-statistic: 63.07 on 6 and 8 DF,  p-value: 2.665e-06
```

   (b) The inferences for the fixed effects is the same in both analyses. They produce the same point estimates for the fixed effects and their standard errors.

```
> summary(lmef.out)
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | Block
        (Intercept)  Residual
StdDev:    1.446976 0.3940027

Fixed effects: y ~ A * B
                Value Std.Error DF   t-value p-value
(Intercept)  9.593456 1.0284378  8  9.328183  <.0001
A1           1.284322 0.1039918  8 12.350224  <.0001
B1          -0.246421 0.1472967  8 -1.672959  0.1329
B2           1.556544 0.1541129  8 10.100021  <.0001
A1:B1        0.262713 0.1486697  8  1.767092  0.1152
A1:B2       -0.234322 0.1541129  8 -1.520455  0.1669
```

   The point estimates for σ are the same in both analyses, but from `lmf.out` fitted model we obtained $\left(\frac{0.394\sqrt{8}}{\sqrt{\chi^2_{.975,8}}}, \frac{0.394\sqrt{8}}{\sqrt{\chi^2_{.025,8}}}\right) = (0.266, 0.755)$ as a 95% confidence interval for σ, and for `lmef.out` fitted model we obtained $(0.241, 0.643)$, a narrower interval.

```
> intervals(lmef.out)
Approximate 95% confidence intervals

 Random Effects:
  Level: Block
                    lower     est.    upper
sd((Intercept)) 0.3565873 1.446976 5.871606

 Within-group standard error:
    lower      est.     upper
0.2413784 0.3940027 0.6431317
```

   In the fixed effects analysis the estimate of the block effect is -1.0286. In the mixed effect model τ is assumed normally distributed with mean zero and the estimate of σ_τ is 1.447. From this distribution, the value -1.0286 is within one standard deviation from its mean and, therefore consistent with the estimate of σ_τ.

(c) The BLUE of $\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \tau_1$ is $9.592857 + 1.284920 - 0.2476190 + 0.2603175 - 1.028571 = 9.862$.
The BLUP of $\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \tau_1$ is $9.593456 + 1.284322 - 0.2464212 + 0.2627131 - 1.017791 = 9.876$.
The BLUP of $\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \tau_3$ is $9.593456 + 1.284322 - 0.2464212 + 0.2627131 + 0 = 10.894$.
In the fixed effects model we require information from the third block in order to estimate its effect and to predict a new observation from this block.

```
> random.effects(lmef.out)
  (Intercept)
1   -1.017791
2    1.017791
```

2. Use a very small set of fake unbalanced two-level nested data.

(a) The pooled variance from the 5 samples (of sizes 2,4,2,2, and 3) is 1.089 (standard error = 1.044). An exact 95% confidence interval for σ is $\left(\frac{1.044\sqrt{8}}{\sqrt{\chi^2_{.975,8}}}, \frac{1.044\sqrt{8}}{\sqrt{\chi^2_{.025,8}}}\right) = (0.705, 2.000)$. An approximate 95% confidence interval for σ is $(0.640, 1.745)$. The algorithm implemented in R produces a narrower confidence interval.

```
> intervals(lmef.out2)
Approximate 95% confidence intervals

 Within-group standard error:
    lower      est.     upper
0.6397994 1.0565890 1.7448912
```

(b) Note that the estimate of σ (residual standard error = 1.044) produced by `lmf.out2` is exactly the one based on the mean square error referred in 2(a).

```
> summary(lmf.out2)
Residual standard error: 1.044 on 8 degrees of freedom
Multiple R-Squared: 0.6286,     Adjusted R-squared: 0.4429
F-statistic: 3.385 on 4 and 8 DF,  p-value: 0.06686
```

The predictions for the fixed effects model correspond to the cell sample means ($\bar{y}_{ij.}$)

```
> round(predict(lmf.out2),3)
    1     2     3     4     5     6     7     8     9    10    11    12    13
6.050 6.050 7.400 7.400 7.400 7.400 9.650 9.650 8.500 8.500 8.067 8.067 8.067
```

The mixed effects estimate of μ is $\hat{\mu} = 7.793$ under the mixed effects model. The distance between the mixed effects predictions and $\hat{\mu}$ is closer than the distance between the fixed effects predictions and $\hat{\mu}$, for the combinations 1/1 and 2/1.

```
> round(predict(lmef.out2),3)
  1/1   1/1   1/2   1/2   1/2   1/2   2/1   2/1   2/2   2/2   2/3   2/3   2/3
6.739 6.739 7.210 7.210 7.210 7.210 8.881 8.881 8.530 8.530 8.354 8.354 8.354
```

3. Consider the scenario of the example beginning on page 1190 of Neter *et. al.*

(a)
```
> anova(lmf1.out)
Analysis of Variance Table

Response: y
             Df  Sum Sq Mean Sq F value Pr(>F)
time          2   67073   33537                     # SSA    =   67073
ad            1  168151  168151                     # SSB    =  168151
time:ad       2     391     196                     # SSAB   =     391
ad:test       8 1833681  229210                     # SSC(B) = 1833681
time:ad:test 16    5727     358                     # SSE    =    5727
Residuals     0       0
```

```
> anova(lmf2.out)
Analysis of Variance Table

Response: y
             Df  Sum Sq Mean Sq F value Pr(>F)
time          2   67073   33537                # SSA    =   67073
ad            1  168151  168151                # SSB    =  168151
test          4  397490   99373
time:ad       2     391     196                # SSAB   =     391
time:test     8    2475     309
ad:test       4 1436191  359048
time:ad:test  8    3252     407
Residuals     0       0
```

SSC(B) = SS(test) + SS(ad:test) and Df(C(B)) = Df(test) + Df(ad:test).

SSE = SS(time:test) + SS(time:ad:test) and Df(Error) = Df(time:test) + Df(time:ad:test).

(b) An approximate 95% confidence interval for σ_γ is $(169.076, 451.181)$ and an approximate 95% confidence interval for σ is $(13.380, 26.754)$.

```
> intervals(lmef.out)
Approximate 95% confidence intervals
 Random Effects:
  Level: loc
                    lower     est.    upper
sd((Intercept)) 169.0764 276.1957 451.1809

 Within-group standard error:
   lower     est.    upper
13.37975 18.92001 26.75436
```

(c) Exact 95% confidence interval for σ is $\left(\sqrt{\frac{5727}{\chi^2_{.975,16}}}, \sqrt{\frac{5727}{\chi^2_{.025,16}}}\right) = (14.090, 28.794)$.

$\hat{\sigma}^2_\gamma = \frac{MSC(B)-MSE}{3} = 229210 - 3583 = 76284$ and, $\nu = \frac{76284^2}{\frac{((1/3)229210)^2}{8} + \frac{((-1/3)358)^2}{16}} = 7.975$. Therefore, an approximate 95% confidence interval for σ_γ is $\left(\sqrt{\frac{76284\nu}{\chi^2_{.975,\nu}}}, \sqrt{\frac{76284\nu}{\chi^2_{.025,\nu}}}\right) = (186.467, 529.839)$.

These intervals are wider than those obtained in 3(b).

(d) A 95% confidence interval for the difference in Time 1 and Time 2 main effects is $(\bar{y}_{1..} - \bar{y}_{2..}) \pm t_{.975,16}\sqrt{\frac{2}{10}MSE} = (-98.4, -62.5)$. A 95% confidence interval for the difference in Ad Campaign 1 and Ad Campaign 2 main effects is $(\bar{y}_{.1.} - \bar{y}_{.2.}) \pm t_{.975,8}\sqrt{\frac{2}{15}MSC(B)} = (-253.4, 552.9)$. This last interval can also be obtained using `intervals(lmef.out)` and, doubling the limits of the interval for the fixed effect `ad1`.

```
> intervals(lmef.out)
Approximate 95% confidence intervals

 Fixed effects:
                  lower        est.      upper
(Intercept)  479.234485 664.5333333 849.832182
time1        -26.489335 -16.1333333  -5.777331
time2         53.910665  64.2666667  74.622669
ad1         -126.698897  74.8666667 276.432230
time1:ad1    -15.022669  -4.6666667   5.689335
time2:ad1     -9.822669   0.5333333  10.889335
```

(e) We can predict $Y_{22k} = 664.5333 + 64.2667 - 74.8667 - 0.5333 + 0 = 653.4$ for $k = new$.

```
> summary(lmef.out)
Fixed effects: y ~ 1 + time * ad
                 Value Std.Error DF   t-value p-value
(Intercept) 664.5333  87.40902 16  7.602571  <.0001
time1       -16.1333   4.88512 16 -3.302543  0.0045
time2        64.2667   4.88512 16 13.155583  <.0001
ad1          74.8667  87.40902  8  0.856510  0.4166
time1:ad1    -4.6667   4.88512 16 -0.955281  0.3536
time2:ad1     0.5333   4.88512 16  0.109175  0.9144
```

4. Use the data set in "Statistical Quality Control in the Chemical Laboratory" by G. Wernimont.

(a) A 95% confidence interval for σ is $\left(\sqrt{\frac{.0351}{\chi^2_{.975,22}}}, \sqrt{\frac{.0351}{\chi^2_{.025,22}}}\right) = (0.0309, 0.0565)$.

$\hat{\sigma}^2_\beta = \frac{MSB(A)-MSE}{2} = \frac{\frac{1.9003}{11} - \frac{.0351}{22}}{2} = 0.08558$ and, $\nu = \frac{0.08558^2}{\frac{((1/2)0.172754545)^2}{11} + \frac{((-1/2)0.001595455)^2}{22}} = 10.8$. Therefore, an approximate 95% confidence interval for σ_β is $\left(\sqrt{\frac{0.08558\nu}{\chi^2_{.975,\nu}}}, \sqrt{\frac{0.08558\nu}{\chi^2_{.025,\nu}}}\right) = (0.2066, 0.4998)$.

$\hat{\sigma}^2_\alpha = \frac{MSA-MSB(A)}{2(2)} = \frac{\frac{3.2031}{10} - \frac{1.9003}{11}}{4} = 0.03689$ and, $\nu = \frac{0.03689^2}{\frac{((1/4)0.32031)^2}{10} + \frac{((-1/4)0.172754545)^2}{11}} = 1.68$. Therefore, an approximate 95% confidence interval for σ_α is $\left(\sqrt{\frac{0.03689\nu}{\chi^2_{.975,\nu}}}, \sqrt{\frac{0.03689\nu}{\chi^2_{.025,\nu}}}\right) = (0.0963, 1.6366)$.

The point estimates of sources of variation suggest that the largest part of the variation in measured copper content comes from differences between specimens from a given casting (σ_β). However, the confidence intevals suggest that the variation between castings (σ_α) is estimated with higher uncertainty and the variation from this part could be even higher that the one between specimens.

(b) A 95% confidence interval for μ is $\bar{y}_{...} \pm t_{.975,10}\sqrt{\frac{MSA}{11(2)(2)}} = (85.378, 85.758)$.