ST512-R
Homework #6 - solutions

1. Rao 14.16

   (a) $Y_{ij} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + E_{ijk}$ where α_i are fixed diet effects and $B, (\alpha B)$ and E denote random location, interaction and error terms w/ varcomps $\sigma^2_B, \sigma^2_{\alpha B}, \sigma^2$, resp. (A mixed, crossed model.)

   (b)

Cov Parm	Estimate
location	0.03304
diet*location	0.005054
Residual	0.004779

   Interpretation: big location-to-location variability

   (c)

```
                     The SAS System                          1
                   The Mixed Procedure
                Type 3 Analysis of Variance
```

Source	DF	Sum of Squares	Mean Square
diet	2	0.011108	0.005554
location	3	0.639313	0.213104
diet*location	6	0.089325	0.014887
Residual	12	0.057350	0.004779

Source	Expected Mean Square	Error Term
diet	Var(Residual) + 2 Var(diet*location) + Q(diet)	MS(diet*location)
location	Var(Residual) + 2 Var(diet*location) + 6 Var(location)	MS(diet*location)
diet*location	Var(Residual) + 2 Var(diet*location)	MS(Residual)
Residual	Var(Residual)	.

Source	Error DF	F Value	Pr > F
diet	6	0.37	0.7035
location	6	14.31	0.0038
diet*location	12	3.12	0.0445

   - The F-tests for fixed diet and for random location effects use $MS[AB]$ as the denominator error term on $(3-1)(4-1) = 6$ df.
   - The F-test for random diet× location uses $MS[E]$ as the denominator error term on 12 df.

   (d) 90% c.i. for $\mu + \alpha_1$ given by

$$\bar{y}_{i++} \pm t(0.05, \hat{df})\sqrt{\frac{1}{nb}(\hat{\sigma}^2 + n\hat{\sigma}^2_B + n\hat{\sigma}^2_{\alpha B})}$$

where $\hat{df}$ approx'd by Satterthwaite (Box 14.9) as $\hat{df} = 3.86$.

$$\begin{aligned} SE(\bar{y}_{i++}) &= \sqrt{\frac{1}{nb}(\hat{\sigma}^2 + n\hat{\sigma}^2_B + n\hat{\sigma}^2_{\alpha B})} \\ &= \sqrt{\frac{1}{anb}(MS[AB](a-1) + MS[B])} \\ &= \sqrt{\frac{1}{24}((0.0149)(3-1) + 0.2131)} \\ &= \sqrt{0.0101} \\ &= 0.1006 \end{aligned}$$

This is a linear combo of $MS[AB]$ and $MS[B]$ terms and Box 14.9 gives

$$\begin{aligned} \hat{df} &= \frac{(c_1 MS_1 + c_2 MS_2 + \cdots + c_k MS_k)^2}{(c_1 MS_1)^2/df_1 + (c_2 MS_2)^2/df_2 + \cdots + (c_k MS_k)^2/df_k} \\ &= \frac{(0.1006)^4}{\frac{1}{24^2}\left(\frac{((3-1)0.0149)^2}{6} + \frac{0.2131^2}{3}\right)} \\ &= 3.86 \end{aligned}$$

In the output, the c.i. for diet 1 mean appears as $\boxed{(1.87, 2.3)}$:

Least Squares Means

Effect	diet	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
diet	1	2.0825	0.1006	3.86	20.70	<.0001	0.1
diet	2	2.0812	0.1006	3.86	20.69	<.0001	0.1
diet	3	2.1275	0.1006	3.86	21.15	<.0001	0.1

Effect	diet	Lower	Upper
diet	1	1.8658	2.2992
diet	2	1.8645	2.2980
diet	3	1.9108	2.3442

(e) Simultaneous Bonferroni 90% c.i.s for differences among diet means are "clean" w/ $\widehat{SE} = \sqrt{2/(nb)MS[AB]} = 0.061$ on 6 *df*.

Differences of Least Squares Means

Effect	diet	_diet	Estimate	Standard Error	DF	t Value	Pr > |t|
diet	1	2	0.001250	0.06101	6	0.02	0.9843
diet	1	3	-0.04500	0.06101	6	-0.74	0.4886
diet	2	3	-0.04625	0.06101	6	-0.76	0.4771

Effect	diet	_diet	Adjustment	Adj P	Alpha	Lower	Upper
diet	1	2	Bonferroni	1.0000	0.1	-0.1173	0.1198
diet	1	3	Bonferroni	1.0000	0.1	-0.1635	0.07355
diet	2	3	Bonferroni	1.0000	0.1	-0.1648	0.07230

Effect	diet	_diet	Adj Lower	Adj Upper
diet	1	2	-0.1665	0.1690
diet	1	3	-0.2127	0.1227
diet	2	3	-0.2140	0.1215

2. Rao 14.18ab

   (a)

$$\begin{aligned}\hat{\sigma}^2 &= MS[E] \\ &= 122 \\ \hat{\sigma}^2_{B(A)} &= \frac{MS[B(A)] - MS[E]}{n} \\ &= 95 \\ \hat{\sigma}^2_A &= \frac{MS[A] - MS[B(A)]}{nb} \\ &= \frac{2240 - 312}{6} \\ &= 321\end{aligned}$$

   On inspection, the proportion of variability in a hospital stay due to hospital variability is $321/(321 + 95 + 122) \approx 60\%$ the proportion due to day-to-day variability (within a hospital) is $95/(321+95+122) \approx 18\%$. Fluctuation between hospitals is much bigger than fluctuations between days.

   (b) 95% c.i. for μ given by

$$\bar{y}_{+++} \pm t(0.025, 3)\sqrt{\frac{1}{nab} MS[A]}$$

   which yields

$$380.24 \pm 3.18\sqrt{\frac{1}{24}2240}$$

   or

$$380.24 \pm 3.18(9.66)$$

   or

$$\$380.24 \pm \$30.7$$

3. Rao 14.19 (use 95% in part d)

   (a) $Y_{ijk} = \mu + \alpha_i + B_{j(i)} + E_{ijk}$ (mixed, nested model) w/ random tech. effect nested in fixed lab effect.

   (b) Since

$$\widehat{SE}(\bar{y}_{i++} - \bar{y}_{j++}) = \sqrt{\frac{2}{nb} MS[B(A)]} = \sqrt{\frac{2}{40}1.63} = 0.29$$

   The lab means are

Lab	1	2	3	4
Mean	3.8	3.6	4.6	4.4

   Using Bonferroni, with $t(.025/6, 12)SE = 3.15(0.29) = 0.91$ we see that the only significant difference is between labs 2 and 3.

(c) $\hat{\sigma}^2 = MS[E] = 1.02$ and $\hat{\sigma}^2_{B(A)} = (MS[B(A)] - MS[E])/10 = 0.061$ so that

$$\hat{\rho} = \frac{0.061}{0.061 + 1.02} = 0.056.$$

Not much intra-technician correlation. The variability due to random technician effect is small.

(d)

$$\begin{aligned} SE(\hat{\theta}_1) &= \sqrt{\frac{2}{40} MS[B(A)]} \\ &= 0.285 \\ SE(\hat{\theta}_2) &= \sqrt{\frac{1}{40} MS[B(A)]} \\ &= 0.20 \\ SE(\hat{\theta}_3) &= \sqrt{\frac{2}{40} MS[B(A)]} \\ &= 0.285 \end{aligned}$$

each on $df = 12$ with $t(0.025/3, 12) = 2.78$ yielding the c.i.'s
$0.2 \pm 2.78(0.285)$ and
$0.8 \pm 2.78(0.20)$ and
$0.2 \pm 2.78(0.285)$ respectively for $\theta_1, \theta_2, \theta_3$.

4. Rao 14.20

(a) τ_i are fixed fertilizer treatment effects, $R_{j(i)}$ are random pot effects, nested in the fertilizer factor. We're assuming independent, normally distributed random effects.

(b) (table below)

Source	DF	Sum of Squares	Mean Square	EMS	F	p-value
trt	3	394.440000	131.480000	$\sigma^2 + 3\sigma^2_{B(A)} + 9\psi^2_A$	129.61	< 0.0001
pot(trt)	8	8.115556	1.014444	$\sigma^2 + 3\sigma^2_{B(A)}$	3.63	0.0066
Residual	24	6.700000	0.279167	σ^2		

(c) Yup. $p < 0.0001$

(d)

```
                  Differences of Least Squares Means

                                          Standard
Effect    trt    _trt    Estimate          Error     DF    t Value    Pr > |t|

trt        1      2        7.9333         0.4748      8      16.71      <.0001
trt        1      3        7.3556         0.4748      8      15.49      <.0001
trt        1      4        7.6000         0.4748      8      16.01      <.0001
trt        2      3       -0.5778         0.4748      8      -1.22      0.2583
trt        2      4       -0.3333         0.4748      8      -0.70      0.5026
trt        3      4        0.2444         0.4748      8       0.51      0.6206

                  Differences of Least Squares Means
```

Effect	trt	_trt	Adjustment	Adj P
trt	1	2	Tukey	<.0001
trt	1	3	Tukey	<.0001
trt	1	4	Tukey	<.0001
trt	2	3	Tukey	0.6342
trt	2	4	Tukey	0.8935
trt	3	4	Tukey	0.9532

(The "none" fertilizer treatment differs significantly from each of the other three, among which there are no sig diffs.)

(e) (Using `alpha=0.025`, we can trick SAS into giving us Bonferroni adjusted simultaneous 95% confidence intervals for θ_1 (6.6,9.2) and θ_2 (6.6,8.7).)

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
theta1	7.9333	0.4748	8	16.71	<.0001	0.025
theta2	7.6296	0.3877	8	19.68	<.0001	0.025

Label	Lower	Upper
theta1	6.6269	9.2397
theta2	6.5629	8.6963

(f) Using the output below,

Cov Parm	Estimate
pot(trt)	0.2451
Residual	0.2792

we have

$$\frac{\hat{\sigma}^2_{B(A)}}{\hat{\sigma}^2 + \hat{\sigma}_{B(A)}} = \frac{0.25}{0.25 + 0.28} = 0.47$$

The random pot effect is big, it accounts for almost half the variability in a wheat yield measurement for a fixed fertilizer treatment.

5. Rao 14.14

(a) μ is the mean of all a treatment means, $\mu = \frac{1}{a}\sum_{i=1}^{a}(\mu + \alpha_i)$, α_i is the effect of the i^{th} level of factor A, a factor with effects which are modelled as fixed. $\bar{B}_+$ is the average of the b effects of the b levels of factor B that were randomly sampled for this experiment. $(\alpha B)_{1+}$ is the average of the b interaction effects corresponding to the first level of a. Interaction effects allow for the effects of B in the experiment to change with levels of a. (The expression given in part (a) is "derived" below):

$$\begin{aligned}
\text{Var}(\bar{Y}_{1++}) &= \frac{1}{nb}\sum_j \sum_k Y_{1jk} \\
&= \frac{1}{nb}\sum_j \sum_k (\mu + \alpha_1 + B_j + (\alpha B)_{1j} + E_{1jk}) \\
&= \mu + \bar{B}_+ + \overline{(\alpha B)}_{1+} + \overline{E}_{1++}
\end{aligned}$$

(b)

$$\begin{aligned} Var(\bar{Y}_{1++}) &= Var(\mu + \bar{B}_{+} + \overline{(\alpha B)}_{1+} + E_{1++}) \\ &= \frac{\sigma_B^2}{b} + \frac{\sigma_{alphaB}}{n} + \frac{\sigma^2}{nab} \end{aligned}$$

(c)

$$\widehat{Var}(\bar{Y}_{1++} = \frac{MS(AB)}{nb} + \frac{MS(B) - MS(AB)}{nab} = \frac{1}{nab}\left((a-1)MS(AB) + MS(B)\right)$$

(d) Let's not argue about it!

6. Rao 14.24

(a)

Source	*df*	EMS	F
A	2	$8\sigma_{\alpha B}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2 + 16\Psi_A^2$	$MS(A)/MS(AB)$
B	1	$24\sigma_B^2 + 8\sigma_{\alpha B}^2 + 12\sigma_{B\gamma}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2$	complicated
C	1	$12\sigma_{B\gamma}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2 + 24\Psi_C^2$	$MS(C)/MS(BC)$
AB	2	$8\sigma_{\alpha B}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2$	$MS(AB)/MS(ABC)$
AC	2	$4\sigma_{\alpha B\gamma}^2 + \sigma^2 + 8\Psi_{AC}^2$	$MS(AC)/MS(ABC)$
BC	1	$12\sigma_{B\gamma}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2$	$MS(BC)/MS(ABC)$
ABC	2	$4\sigma_{\alpha B\gamma}^2 + \sigma^2$	$MS(ABC)/MS(E)$

(b) See F-ratios above. For the one labelled complicated, note that under $H_0 : \sigma_B^2 = 0$,

$$EMS(B) = \underbrace{8\sigma_{\alpha B}^2}_{EMS(AB)-EMS(ABC)} + \underbrace{12\sigma_{B\gamma}^2 + 4\sigma_{\alpha B\gamma}^2 + \sigma^2}_{EMS(BC)}$$

so that under H_0, $EMS(B) = EMS(AB) - EMS(ABC) + EMS(BC)$. The distribution of the F-ratio

$$F = \frac{MS(B)}{MS(AB) + MS(BC) - MS(ABC)}$$

under H_0 may be approximated using the Satterthwaite formula for the denominator df.