Stat 505 Assignment 2

Fall 2014 Solutions

1. Construct a 4 by 4 variance-covariance matrix showing $\boldsymbol{R}$, $\boldsymbol{D}$, and $\boldsymbol{V}$

   (a) so that covariances are all 0 and variances are 1, 4, 9, 16.

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 16 \end{bmatrix},$$

   (b) so that each variance is $\frac{1}{4}$ and all correlations are .6.

$$\boldsymbol{D} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0.6 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{bmatrix}, \quad \boldsymbol{V} = \frac{1}{4}\begin{bmatrix} 1 & .6 & .6 & .6 \\ .6 & 1 & .6 & .6 \\ .6 & .6 & 1 & .6 \\ .6 & .6 & .6 & 1 \end{bmatrix}$$

   (c) so that each variance is 9, neighboring observations have covariance 3, observations 2 steps apart have covariance 1, and the covariance between observations 1 and 4 is $\frac{1}{3}$.

$$\boldsymbol{D} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{9} & \frac{1}{27} \\ \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{3} & 1 & \frac{1}{3} \\ \frac{1}{27} & \frac{1}{9} & \frac{1}{3} & 1 \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 9 & 3 & 1 & \frac{1}{3} \\ 3 & 9 & 3 & 1 \\ 1 & 3 & 9 & 3 \\ \frac{1}{3} & 1 & 3 & 9 \end{bmatrix}$$

   (d) so that correlations are the same as in (c), but variances are $\mu_i^{1.4}$ where the vector of means is $\boldsymbol{\mu} = (2\ 3\ 7\ 6)^T$.

$$\boldsymbol{D} = \begin{bmatrix} 2^{.7} & 0 & 0 & 0 \\ 0 & 3^{.7} & 0 & 0 \\ 0 & 0 & 7^{.7} & 0 \\ 0 & 0 & 0 & 6^{.7} \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{9} & \frac{1}{27} \\ \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{3} & 1 & \frac{1}{3} \\ \frac{1}{27} & \frac{1}{9} & \frac{1}{3} & 1 \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 2^{1.4} & \frac{6^{.7}}{3} & \frac{14^{.7}}{9} & \frac{12^{.7}}{27} \\ \frac{6^{.7}}{3} & 3^{1.4} & \frac{21^{.7}}{3} & \frac{18^{.7}}{9} \\ \frac{14^{.7}}{9} & \frac{21^{.7}}{3} & 7^{1.4} & \frac{42^{.7}}{3} \\ \frac{12^{.7}}{27} & \frac{18^{.7}}{9} & \frac{42^{.7}}{3} & 6^{1.4} \end{bmatrix}$$

2. Let $\boldsymbol{z}$ be a random vector of 8 with distribution $MVN_8(\boldsymbol{0}, 4\boldsymbol{I})$ and let $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$.

   What is the distribution of $\begin{bmatrix} 1 \\ 3 \\ 2 \\ 7 \end{bmatrix} + \boldsymbol{\Sigma z}$?

   multivariate normal: mean $= \begin{bmatrix} 1 \\ 3 \\ 2 \\ 7 \end{bmatrix}$, *and variance* $= \begin{bmatrix} 12 & 8 & 0 & 4 \\ 8 & 16 & 8 & 0 \\ 0 & 8 & 16 & 8 \\ 4 & 0 & 8 & 16 \end{bmatrix}$ $(= 4\boldsymbol{\Sigma\Sigma}^T)$.

	df	SS	MS	F	p-value
Between	3	20.42	6.81	2.41	0.0680
Within	217	613.14	2.83		
Total	220	633.56			

3. Fill in the blanks.

4. Set up the ANOVA model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},\ \ i = 1, \ldots, 4, j = 1, 2$$

in R with these data:

(a) Fit using `lm(y ~ f)`. Give estimates of all estimable cell means, explaining how they relate to the output.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.5000	0.6614	5.29	0.0061
fB	-3.0000	0.9354	-3.21	0.0327
fC	4.5000	0.9354	4.81	0.0086
fD	2.0000	0.9354	2.14	0.0993

The first coefficient is $\overline{y}_A = 3.5$ *which has expectation* $\mu + \tau_1$*, then we have adjustments between the mean of the other groups and the mean of group A. For B:* $\widehat{\tau_2 - \tau_1} = -3$*, C:* $\widehat{\tau_3 - \tau_1} = 4.5$*, and D:* $\widehat{\tau_4 - \tau_1} = 2.0$*. The estimated cell means are then A: 3.5, B: 0.5, C: 8.0, D: 5.5*

(b) Do the same using formula `lm(y ~ f + 0)`.

	Estimate	Std. Error	t value	Pr(>|t|)
fA	3.5000	0.6614	5.29	0.0061
fB	0.5000	0.6614	0.76	0.4918
fC	8.0000	0.6614	12.09	0.0003
fD	5.5000	0.6614	8.32	0.0011

Now we see the estimated cell means directly. It's as if R *set the overall* $\widehat{\mu} = 0$*, and then we are seeing estimates of the* τ_i*'s under that constraint.*

(c) Do the same using `lm(y ~ X + 0, singular.ok=T)`.

In this version, the cell mean for D is taken to be baseline, so the first estimate is $\widehat{\mu + \tau_4} = 5.5 = \overline{y}_D$*, and then we see adjustments the fourth cell mean:* $\widehat{\tau_1 - \tau_4} = -2.0$*,* $\widehat{\tau_2 - \tau_4} = -5.0$*,* $\widehat{\tau_3 - \tau_4} = 2.5$*.*

The coefficient table moved to the top of the next page.

(d) Show that you can obtain the same coefficient estimates as in (a) by solving the normal equations $\boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{\beta} = \boldsymbol{X}_f^T \boldsymbol{y}$ (Use `crossprod` and `solve` functions.)

These match the coefficient estimates from (a).

	Estimate	Std. Error	t value	Pr(>|t|)
X	5.5000	0.6614	8.32	0.0011
XfA	-2.0000	0.9354	-2.14	0.0993
XfB	-5.0000	0.9354	-5.35	0.0059
XfC	2.5000	0.9354	2.67	0.0557

	Beta.hat (a)	Beta.hat (d)	Unique Fits (d)	Unique Fits (e)
Xf(Intercept)	3.50	3.50	3.50	3.50
XffB	-3.00	-3.00	0.50	0.50
XffC	4.50	4.50	8.00	8.00
XffD	2.00	2.00	5.50	5.50

(e) Use the Moore Penrose inverse to find another solution to the normal equations for the full $\boldsymbol{X}$ matrix, $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{y}$ Show that estimates of the cell means are the same.

	1	2	3	4	5
1	3.50	0.00	-3.00	4.50	2.00

Table 1: Coefficient Estimates Differ only in inserted 0

I included the fits in the table for part (d).

(f) How do Xf and X differ?

Xf is missing the second column of X. To get the estimates of cell means, we just need to take the unique rows (1,3,5,7) of either and matrix multiply by the estimated coefficient vector.

R Code

```
D <- diag(c(2, 3, 7, 6)^0.7)
R <- matrix(1/c(1, 3, 9, 27, 3, 1, 3, 9, 9, 3, 1, 3, 27, 9, 3, 1), 4, 4)
V <- D %*% R %*% D
require(xtable)
xtable(V, digits = 3)
```

```
sigma <- rbind(c(1, 1, 0, 0, 1, 0, 0, 0), c(0, 1, 1, 0, 1, 1, 0, 0), c(0, 0, 1, 1,
    0, 1, 1, 0), c(1, 0, 0, 1, 0, 0, 1, 1))
crossprod(sigma) * 4
```

```
require(xtable, quietly = TRUE)
dfE <- 217
dfTrt <- 220 - dfE
SSTrt <- 20.42
MSTrt <- SSTrt/dfTrt
SSE <- 633.56 - SSTrt
MSE <- SSE/dfE
Fstat <- MSTrt/MSE
output <- rbind(c(dfTrt, SSTrt, MSTrt, Fstat, 1 - pf(Fstat, dfTrt, dfE)), c(dfE, SSE,
    MSE, NA, NA), c(220, 633.56, NA, NA, NA))
dimnames(output) <- list(c("Between", "Within", "Total"), c("df", "SS", "MS", "F",
    "p-value"))
xtable(output, digits = c(0, 0, 2, 2, 2, 4))
```

	1	2	3	4
1	2.639	1.168	0.705	0.211
2	1.168	4.656	2.808	0.840
3	0.705	2.808	15.245	4.562
4	0.211	0.840	4.562	12.286

```
y <- c(3, 4, 1, 0, 7, 9, 6, 5)
f <- gl(4, 2, labels = LETTERS[1:4])
Xf <- model.matrix(y ~ f)
X <- cbind(1, model.matrix(y ~ f + 0))
```

```
xtable(lm(y ~ f))
```

```
xtable(summary(lm(y ~ f + 0)))
```

```
print(xtable(lm(y ~ X + 0, singular.ok = TRUE)), table.placement = "hbt")
```

```
beta.hat <- solve(crossprod(Xf), crossprod(Xf, y))
beta.hat2 <- MASS::ginv(crossprod(X)) %*% crossprod(X, y)
output <- cbind(coef(lm(y ~ Xf + 0)), beta.hat, Xf[c(1, 3, 5, 7), ] %*% beta.hat, X[c(1,
    3, 5, 7), ] %*% beta.hat2)
colnames(output) <- c("Beta.hat (a)", "Beta.hat (d)", "Unique Fits (d)", "Unique Fits (e)")
xtable(output)
```

```
xtable(t(  beta.hat2 <-  MASS::ginv(crossprod(X)) %*% crossprod(X,y)),
       caption="Coefficient Estimates Differ only in inserted 0",only.contents=TRUE,
       include.rownames=FALSE, include.colnames=FALSE)
out2 <-   X[c(1,3,5,7),] %*% beta.hat2
##dimnames(out2) <- list(c("meanA","meanB","meanC","meanD"),NULL)
##xtable( t(out2),floating = "F")
```