Stat 505 Assignment 2

Due: Sept 12, 2014 Put your name somewhere in the header.

1. Construct a 4 by 4 variance-covariance matrices showing **R, D**, and **V** for each case:

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

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

   (b) so that each variance is .25 and all correlations are .6.

   $$\boldsymbol{D} = \begin{bmatrix} . & 0 & 0 & 0 \\ 0 & . & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & . \end{bmatrix} \quad \boldsymbol{R} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \quad \boldsymbol{V} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

   (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 $\frac{1}{3}$.

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

2. Let $\boldsymbol{z} \sim N_8(\mathbf{0}, 4\boldsymbol{I}_8)$ be a random vector and let $\boldsymbol{\Sigma}$ be a 4 by 8 matrix with these entries:

```
1 1 0 0 1 0 0 0
0 1 1 0 1 1 0 0
0 0 1 1 0 1 1 0
1 0 0 1 0 0 1 1
```

   Describe the distribution of

   $$\begin{bmatrix} 1 \\ 3 \\ 2 \\ 7 \end{bmatrix} + \boldsymbol{\Sigma z}$$

3. Fill in the blanks. You can use the verbatim environment

```
Source            df       SS       MS       F    p-value
---------------------------------------------------


Between groups    __    20.42      ----    ----    ----

Within groups    217    ----     ----
---------------------------------------------------
Total            220   633.56
```

or build a LaTeX table by hand with columns separated by & signs:

Source	df	SS	MS	F	p-value
Between groups		20.42			
Within groups	217				
Total	220	633.56			

or make R create the table for you.

4. Set up the ANOVA model $y_{ij} = \mu + \tau_i + \epsilon ij,\ i = 1, ...4, j = 1, 2$ in R with these data:

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

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

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

(c) Is $\boldsymbol{X}_f$ nonsingular (invertible)? If so, solve the normal equations based on $\boldsymbol{X}_f$ using `crossprod` and `solve` functions.

(d) Repeat (a) using `lm(y ~ X + 0, singular.ok=T)`.

(e) Use the Moore Penrose inverse to find another solution to the normal equations for the full $\boldsymbol{X}$ matrix. Show that $\hat{\mu}$ and $\hat{\tau}_i$ differ from those above, but estimates of the cell means are the same.

(f) Explain how `Xf` and `X` differ and how they are used to create the estimates in parts a, b, d, and e.

R Code

```r
require(xtable, quietly = TRUE)
dfTrt <- NA
dfE <- NA
SSTrt <- NA
MSTrt <- NA
SSE <- NA
MSE <- NA
Fstat <- NA
output <- rbind(c(dfTrt, SSTrt, MSTrt, Fstat, 1 - pf(Fstat, df1 <- NA, df2 <- NA)),
    c(dfE, SSE, MSE, NA, NA), c(220, 633.46, 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))
```

```r
y <- c(3, 4, 1, 0, 7, 9, 6, 5)
f <- factor(rep(LETTERS[1:4], each = 2))
## or use the gl function:
f <- gl(4, 2, labels = LETTERS[1:4])
Xf <- model.matrix(y ~ f)
X <- cbind(1, model.matrix(y ~ f + 0))
```