STAT 8004, Assignment 3

David Dobor

February 12, 2015

Question 1

Suppose that we are working under Gauss-Markov model

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where $E(\epsilon) = \mathbf{0}$ and $\text{var}(\epsilon) = \sigma^2\mathbf{I}$. Let $\hat{\mathbf{Y}}$ be the ordinary least square estimator of $\mathbf{Y}$.

(a) Show that $\hat{\mathbf{Y}}$ and $\mathbf{Y} - \hat{\mathbf{Y}}$ are uncorrelated.

(b) Show that
$$E\{(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}})\} = \sigma^2\{n - \text{rank}(\mathbf{X})\}.$$

Answer to Question 1

(a) First, establish orthogohality of $\mathbf{P}_X\mathbf{Y} = \hat{\mathbf{Y}}$ and $(\mathbf{I} - \mathbf{P}_X)\mathbf{Y} = \mathbf{Y} - \hat{\mathbf{Y}}$. This follows because

$$(\mathbf{Y} - \hat{\mathbf{Y}})^T\hat{\mathbf{Y}} = ((\mathbf{I} - \mathbf{P}_X)\mathbf{Y})^T\hat{\mathbf{Y}} = \hat{\mathbf{Y}}^T((\mathbf{I} - \mathbf{P}_X)^T\mathbf{P}_X)\mathbf{Y},$$

and the middle term in the last expression is identically zero (i.e., $(\mathbf{I} - \mathbf{P}_X)^T\mathbf{P}_X = \mathbf{P}_X - \mathbf{P}_X^T\mathbf{P}_X = \mathbf{P}_X - \mathbf{P}_X = 0$ by idempotence and symmetry of $\mathbf{P}_X$).

Therefore

$$(\mathbf{Y} - \hat{\mathbf{Y}})^T\hat{\mathbf{Y}} = 0.$$

Since in the Gauss-Markov model we also have that $E(\mathbf{Y}) = \hat{\mathbf{Y}}$, it then follows that

$$E(\hat{\mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}})) - E(\hat{\mathbf{Y}})E(\mathbf{Y} - \hat{\mathbf{Y}}) = E(0) - E(\hat{\mathbf{Y}}) \times 0 = 0.$$

(b) By Theorem 5.2 (a) we have that (taking the 'any symmetric' $\mathbf{A}$ to be $\mathbf{I}$ in that theorem):

$$E(\mathbf{y}^T\mathbf{y}) = tr(\Sigma) + \mu^T\mu.$$

We apply this theorem with $\mathbf{Y}-\hat{\mathbf{Y}}$ for $\mathbf{y}$, whose 'μ' in the Gauss-Markov model is 0. Thus since

$$\text{var}(\mathbf{Y}-\hat{\mathbf{Y}}) = \text{var}(\ (\mathbf{I}-\mathbf{P}_X)\mathbf{Y}\ ) = \sigma^2(\mathbf{I}-\mathbf{P}_X),$$

and

$$tr(\Sigma) = tr(\sigma^2(\mathbf{I}-\mathbf{P}_X)) = \sigma^2(tr(\mathbf{I}) - tr(\mathbf{P}_X)) = \sigma^2(n-\text{rank(X)}),$$

the result follows.

Question 2

Consider the one-way ANOVA model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ for the jth individual of the ith group. Suppose there are 4 treatments (groups) and the sample sizes are respectively 2,1,1,2 for treatments. Now suppose that $\mathbf{Y} = (y_{11}, y_{12}, y_{21}, y_{31}, y_{41}, y_{42})^T = (2,1,4,6,3,5)^T$ contains the observations. Use R and weighted generalized least squares to find appropriate estimate for

$$E(\mathbf{Y}) \text{ and } \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \beta$$

in the Aiken model with $\text{var}(\epsilon) = \mathbf{V}$ for two cases where

(a) $\mathbf{V} = \mathbf{V}_1 = \text{diag}(1,9,9,1,1,9)$ and

(b)

$$\mathbf{V} = \mathbf{V}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 9 \end{bmatrix}$$

Answer to Question 2

In what follows we transform the following Aiken model

$$\begin{bmatrix}2\\1\\4\\6\\3\\5\end{bmatrix} = \begin{bmatrix}1&1&0&0&0\\1&1&0&0&0\\1&0&1&0&0\\1&0&0&1&0\\1&0&0&0&1\\1&0&0&0&1\end{bmatrix}\begin{bmatrix}\mu\\\tau_1\\\tau_2\\\tau_3\\\tau_4\end{bmatrix} + \begin{bmatrix}\epsilon_{11}\\\epsilon_{12}\\\epsilon_{21}\\\epsilon_{31}\\\epsilon_{41}\\\epsilon_{42}\end{bmatrix}$$

to the Gauss-Markov model with an appropriate linear transformation. We do the `OLS` estimation as usual in the resulting Gauss-Markov model. We then transform our estimates back to the 'Aiken-space' and report the results.

(a)

```
library(MASS) # function ginv() is in the MASS library
# data given in problem 2:
X <- matrix(
    c(1, 1, 1, 1, 1, 1,
      1, 1, 0, 0, 0, 0,
      0, 0, 1, 0, 0, 0,
      0, 0, 0, 1, 0, 0,
      0, 0, 0, 0, 1, 1 ),
    nrow=6,
    ncol=5)

y <- matrix(c(2, 1, 4, 6, 3, 5), 6, 1)

# error variance
V1 <- diag(c(1, 9, 9, 1, 1, 9));
V1.inv <- solve(V1);
# transformation to convert Aiken to Gauss-Markov:
V1.inv.sqrt = sqrt(V1.inv);

# transform data with V1.inv.sqrt:
u <- V1.inv.sqrt %*% y
W <- V1.inv.sqrt %*% X

# beta.hat in the transformed model:
beta.hat <- ginv(t(W) %*% W) %*% t(W) %*% u

# compute u.hat
u.hat <- W %*% beta.hat

# back to y.hat
y.hat <- solve(V1.inv.sqrt) %*% u.hat

## The estimate of c^T \beta
c.t <- matrix(
    c(1, 1, 1, 1,
      1, 0, 0, 0,
      0, 1, 0, 0,
      0, 0, 1, 0,
      0, 0, 0, 1),
    nrow=4,
    ncol=5)

c.t %*% ginv(t(W) %*% W) %*% t(W) %*% u
```

The output for $\hat{y}$:

```
> y.hat
     [,1]
[1,]  1.9
[2,]  1.9
[3,]  4.0
[4,]  6.0
[5,]  3.2
[6,]  3.2
```

The output for the estimate of $c^T\beta$:

```
     [,1]
[1,]  1.9
[2,]  4.0
[3,]  6.0
[4,]  3.2
```

(b)

```
V2 <- matrix(
    c(1,  1,   0,   0,  0,  0,
      1,  9,   0,   0,  0,  0,
      0,  0,   9,  -1,  0,  0,
      0,  0,  -1,   1,  0,  0,
      0,  0,   0,   0,  1, -1,
      0,  0,   0,   0, -1,  9),
    nrow=6,
    ncol=6)

# first, compute the inverse square root of V2
Q <- eigen(V2)$vectors
# D contains the inverse square roots of of V's eigenvalues
# on the diagonal, zeros elsewhere:
D <- diag(1/sqrt(eigen(V2)$values))

# The desired `inverse square root` matrix.
# It's the transformation to convert Aiken to Gauss-Markov:
V2.inv.sqrt <- Q %*% D %*% t(Q)

# transform data with V2.inv.sqrt:
u <- V2.inv.sqrt %*% y
W <- V2.inv.sqrt %*% X

# beta_hat in the transformed model:
beta.hat <- ginv(t(W) %*% W) %*% t(W) %*% u

# compute u.hat
```

```
u.hat <- W %*% beta.hat

# back to y.hat
y.hat <- solve(V2.inv.sqrt)%*% u.hat

## The estimate of c^T \beta
c.t %*% ginv(t(W) %*% W) %*% t(W) %*% u
```

Output for $\hat{y}$:

```
> y.hat
      [,1]
[1,] 2.000
[2,] 2.000
[3,] 4.000
[4,] 6.000
[5,] 3.333
[6,] 3.333
```

The output for the estimate of $c^T\beta$:

```
      [,1]
[1,] 2.000
[2,] 4.000
[3,] 6.000
[4,] 3.333
```

Question 3

The `lm` function in `R` allows one to do weighted least squares with the form $\sum w_i (y_i - \hat{y}_i)^2$ for positive weights w_i. For $\mathbf{V}_1$ in the last question, find the `BLUE`s of the 4 cell means using `lm` and an appropriate vector of weights.

Answer to Question 3

```
X <- matrix(
    c(1, 1, 1, 1, 1, 1,
      1, 1, 0, 0, 0, 0,
      0, 0, 1, 0, 0, 0,
      0, 0, 0, 1, 0, 0,
      0, 0, 0, 0, 1, 1 ),
    nrow=6,
    ncol=5)

y <- matrix(c(2, 1, 4, 6, 3, 5), 6, 1)

lm(y ~ X[,2:5] -1, weights=c(1, 1/9, 1/9, 1, 1, 1/9))
```

output agrees with question 2:

```
Call:
lm(formula = y ~ X[, 2:5] - 1, weights = c(1, 1/9, 1/9, 1, 1,
    1/9))

Coefficients:
X[, 2:5]1  X[, 2:5]2  X[, 2:5]3  X[, 2:5]4
      1.9        4.0        6.0        3.2
```

Question 4

Running

```
library (MASS)
data(Boston)
```

will load the Boston housing data intoR. Use ?Boston to see the information on the variables. Now create two matrices **Y** and **X** that will be used to fit a regression model to some of these data.

```
Y=as.matrix(Boston$medv)
X=as.matrix(Boston[,c('crim','nox','rm','age','dis')])
X=cbind(rep(1,dim(Boston)[1]),X)
```

(a) Make a scatterplot matrix for $y, x_1, \ldots, x_5$. If you had to guess based on this plot, which single predictor do you think is probably the best predictor of Price? Do you see any evidence of multicollinearity (correlation among the predictors) in this graphic?

(b) Use qr() function to find the rank of **X**.

(c) Use R matrix operations on the **X** matrix and **Y** vector to find the estimated regression coefficient vector $\hat{\beta}$, the estimated mean vector $\hat{\mathbf{Y}}$, and the vector of residuals $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$.

(d) Plot the residuals against the fitted means.

(e) Create a normal plot from the values in the residual vector.

(f) Compute the sum of squared residuals and the corresponding estimate of σ^2

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}})}{n - \text{rank}(\mathbf{X})}$$

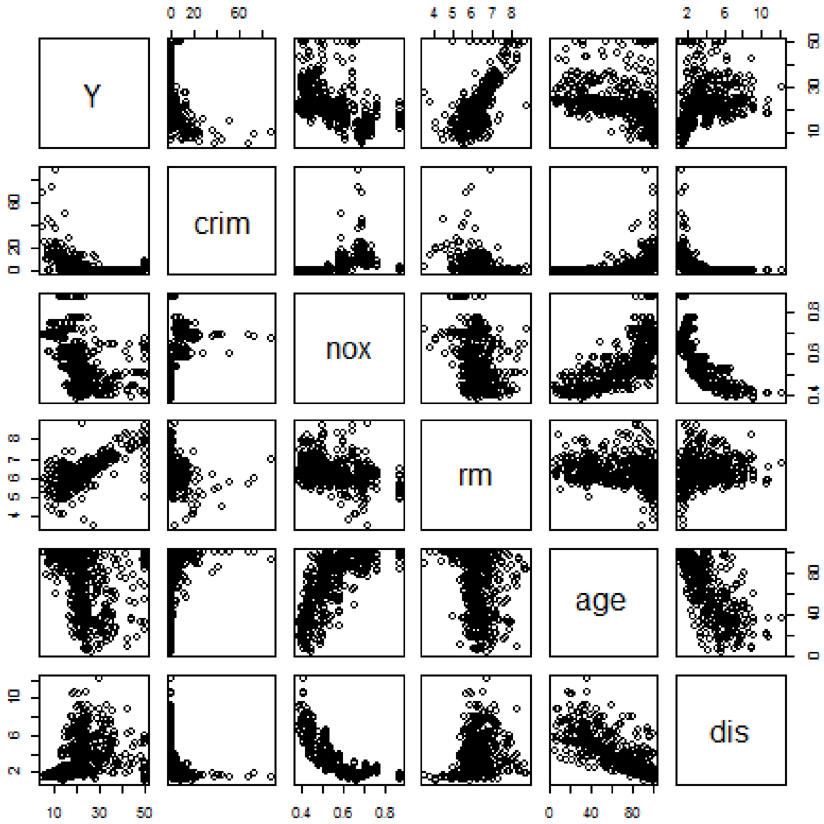
(g) Call the lm function in R and confirm your answers, and note that ?lm gives you various information such as the outputs of the function.

```
m1=lm(medv~crim+nox+rm+age+dis,data=Boston)
```

Answer to Question 4

(a) The average number of rooms (`rm`) has the highest correlation with the price (`medv`).

Although there is no sign of perfect multicollinearity here (after all X is of full rank), some variables appear to be highly correlated. For example, `age` and `dis` are highly negatively correlated (perhaps the discerning eye of an expert statistician might do better at detecting such patterns).



(b)

```
> qrx <- qr(X)
> qrx$rank
[1] 6
```

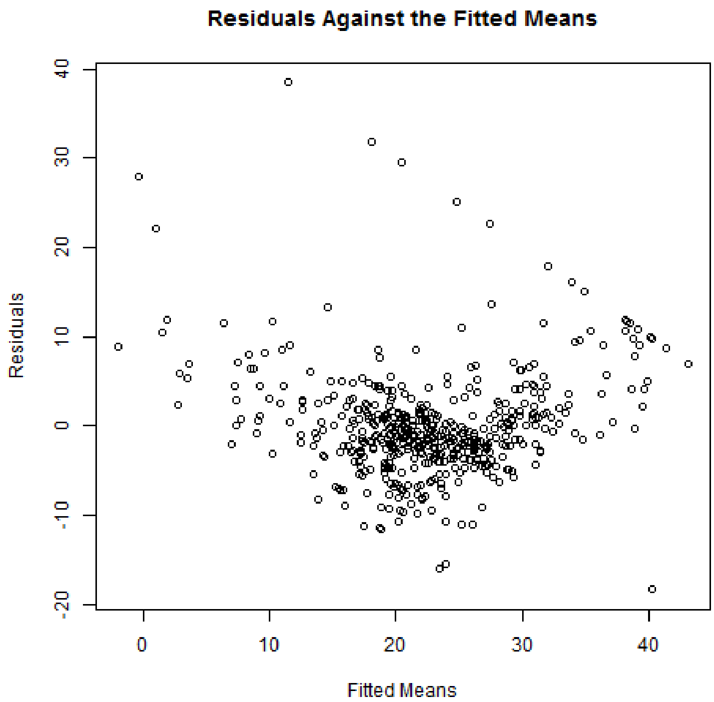
(c) Output for $\hat{\beta}$ is given below. Outputs for $\hat{\mathbf{Y}}$ and $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ are 506 elements long, so only the first few are shown.

```
> beta.hat <- solve(t(X) %*% X) %*% t(X) %*% Y
> beta.hat
              [,1]
       -6.22734381
crim   -0.20808251
nox   -18.05089222
rm      7.73531226
age    -0.06662411
dis    -1.19104483
```

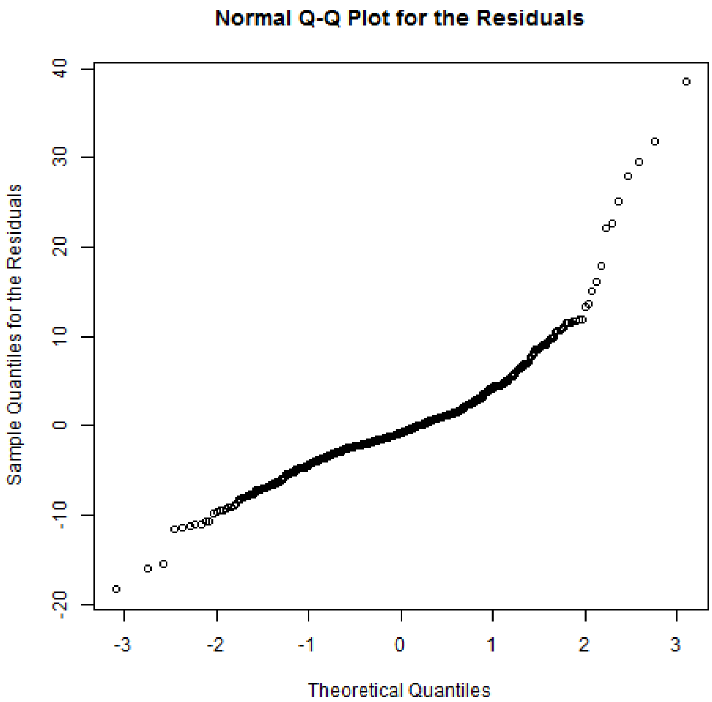
```
> Y.hat <- X %*% beta.hat
> head(Y.hat)  # 506 long - show only first few
      [,1]
1 25.70437
2 23.79686
3 30.89256
4 29.35859
5 29.94388
6 24.10601
```

```
> e <- Y - Y.hat
> head(e)  # 506 long - show only first few
       [,1]
1 -1.704374
2 -2.196864
3  3.807444
4  4.041409
5  6.256122
6  4.593993
```

(d) We note the following pattern in the residuals: the points that seem to lie on a straight line from the upper left to the bottom right (result of censoring `medv` above a certain value):



(e) The residuals clearly deviate from normality:



(f)

```
> sigma.squared.hat <- (t(e) %*% e) / (506 - 6)
> sigma.squared.hat
         [,1]
[1,] 34.82387
```

(g) All the quantities computed earlier based on matrix operations can be obtained from a call to `lm`. For example:

```
m1=lm(medv~crim+nox+rm+age+dis,data=Boston)
m1$coeff              # agrees with beta.hat computed earlier
m1$residuals          # agrees with e computed earlier
m1$df.residual        # agrees with 506 - rank(X)
# etc., etc., etc..
```