Stat 208 Homework 2

Jordan Berninger

5/2/2020

Problem 1

Pardon me for the messy screenshot (on the next page), matrices in LaTeX are tedious so I wrote it out by hand. This is a simple regression model with two parameters, μ, α.

Problem 2

We want to show that $lim_{n->\infty} Var(\hat{\beta}) = 0$. From our class notes, we have that $Var(\hat{\beta}) = \sigma^2(X^TX)^{-1}$ and we are assuming a finite variance, which means we can treat it as a constant. Using some of the work from Problem 1, we see that

$$\sigma^2(X^TX)^{-1} = \frac{\sigma^2}{(n-1)\sum t_i^2} = \begin{bmatrix} -\sum t_i^2 & -\sum t_i \\ -\sum t_i & n \end{bmatrix}$$

We can see that as $limn->\infty$ each entry in this matrix converges to the 0 matrix and thus, asymptotically, $Var(\hat{\beta}) = 0$.

Problem 3

(a) This first model was addressed in lecture and we make use of the trigonometric property $cos(a-b) = cos(a)cos(b) + sin(a)sin(b)$. Accordingly, we implement the following equivalent model: $\vec{Y} = \mu + \alpha t + \gamma t^2 + Acos(2\pi t/12) + Bsin(2\pi t/12) + \epsilon_t$.

```
d=60#2018-1959+1
T=12
N=d*T
p=5 # updated this

# Define Observations and Times
# Initialize time (tax) and observation (Y) arrays
tax=1:N
tyr=1959+tax/12
Y=matrix(0,N,1)

# Define the observations by
for(i in 1:N){
Y[i]=loa[i,5]
}
Y=loa$v3
#Plot the data
plot(tax,Y, type="l", main="Mauna Loa Data",
     xlab="Index of Observation",
     ylab="Carbon Dioxide Concentration")
```

Homework 2

① $Y_t = \mu + \alpha t + \varepsilon_t$

$$X = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

$$M = X(X^T X)^{-1} X^T$$

$$= \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \left(\begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix}$$

$$= \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{pmatrix} n & \Sigma t_i \\ \Sigma t_i & \Sigma t_i^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix}$$

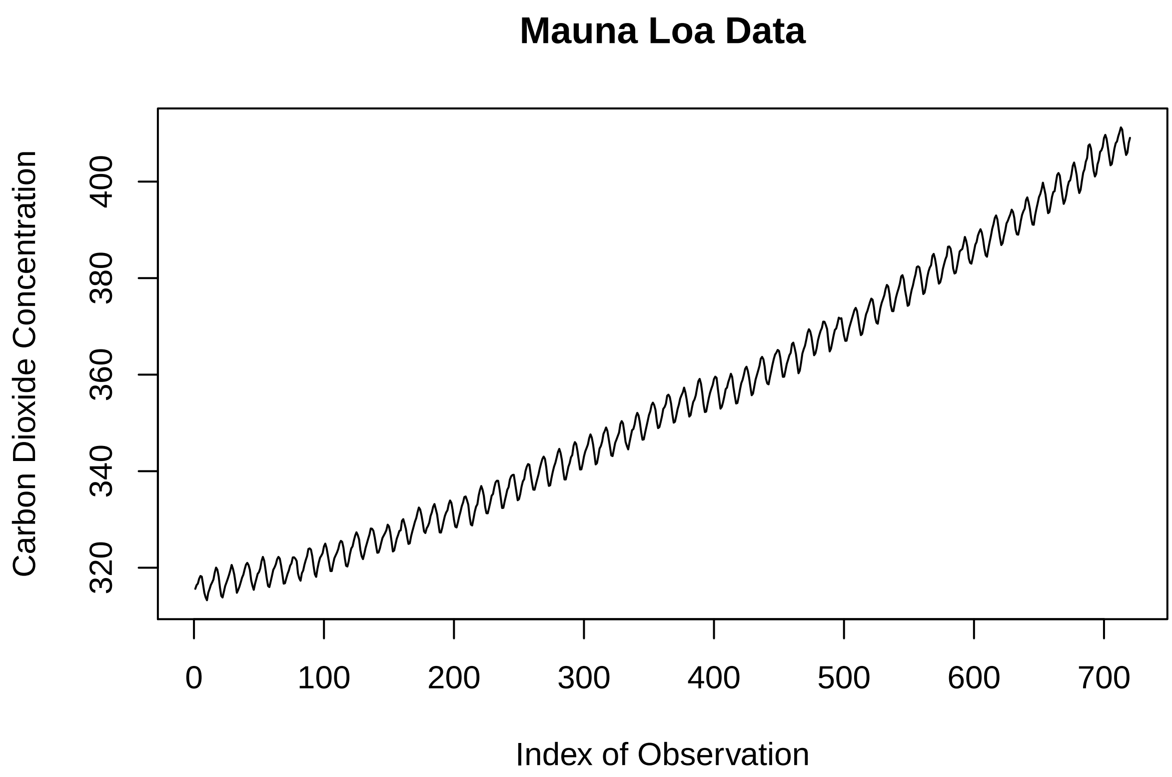
CONSTANT AMONGST MATRICES

$$= \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} \Sigma t_i^2 & -\Sigma t_i \\ -\Sigma t_i & n \end{bmatrix} \left(\frac{1}{n \Sigma t_i^2 - \Sigma t_i^2} \right) \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{bmatrix}$$

$$= \begin{pmatrix} 0 & -\Sigma t_i - n t_1 \\ 0 & -\Sigma t_i - n t_2 \\ 0 & \vdots \\ 0 & \\ 0 & \\ 0 & -\Sigma t_i - n t_n \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} \left(\frac{1}{(n-1)(\Sigma t_i^2)} \right)$$

$$M = \boxed{\left(\frac{1}{(n-1)\Sigma t_i^2} \right) \begin{pmatrix} t_1(-\Sigma t_i - n t_1) & t_2(-\Sigma t_i - n t_1) & \cdots & t_n(-\Sigma t_i - n t_1) \\ \vdots & & & \vdots \\ t_1(-\Sigma t_i - n t_1) & & \cdots & t_n(-\Sigma t_i - n t_n) \end{pmatrix}}$$

Figure 1: Problem 1




```
# Define Design Matrix
D=matrix(0,N,p)
for(i in 1:N){
D[i,1]=1
D[i,2]=i
D[i,3]=i^2
D[i,4]=cos((2*pi*(i))/(12))
D[i,5]=sin((2*pi*(i))/(12))
}

# Compute the OLS estimators
H1=t(D) %*% D
H2=solve(H1)
H3= H2 %*% t(D)
theta_hat= H3 %*% Y

# Compute Estimated Y
Yhat= D %*% theta_hat

# Compute Residuals
Resid=Y-Yhat

# Compute Parameter Standard Errors
se=matrix(0,p,1)
SSE=t(Resid) %*% Resid
sighat=SSE/(N-p)
for(i in 1:p){
se[i]=sighat^(1/2)*H2[i,i]^(1/2)
}
```

```r
R.sq <- 1-SSE/sum((Y-mean(Y))^2)
```

We have the following parameter estimation from this model: $(\hat{\mu}, \hat{\alpha}, \hat{\gamma}, \hat{A}, \hat{B})$:

```r
theta_hat
```

```
##               [,1]
## [1,]  3.149486e+02
## [2,]  6.742288e-02
## [3,]  8.746117e-05
## [4,] -1.656398e+00
## [5,]  2.294233e+00
```

And the SSE and R^2.

```r
SSE
```

```
##          [,1]
## [1,] 644.7252
```

```r
R.sq
```

```
##           [,1]
## [1,] 0.9988068
```

(b) Our second model is $\vec{Y} = \mu + \alpha t + \gamma t^2 + S_t + \epsilon_t$. It will be helpful to reparameterize the model as $\vec{Y} = \mu + \alpha t + \gamma t^2 + aS_1 + \cdots + lS_12 + \epsilon_t$, so each month has its own paramater and $(a + b + c + \cdots + l) = 1$.

```r
d=60#2018-1959+1
T=12
N=d*T
p2=14 # updated this

# Define Observations and Times
# Initialize time (tax) and observation (Y) arrays
tax=1:N
tyr=1959+tax/12

Y=loa$v3
#Plot the data
# Seasonal effect matrix
S=matrix(0,12,11)
diag(S)=rep(1,11)
S[12,]=rep(-1,11)
# Define Design Matrix
D2=matrix(0,N,p2)
for(i in 1:N){
D2[i,1]=1
D2[i,2]=i
D2[i,3]=i^2
```

```r
if(mod(i-1,12) == 0) {
  D2[i:(i+11),4:14] = S
}
}

# Compute the OLS estimators
H1.2=t(D2) %*% D2
H2.2=solve(H1.2)
H3.2= H2.2 %*% t(D2)
theta_hat2= H3.2 %*% Y

# Compute Estimated Y
Yhat2= D2 %*% theta_hat2

# Compute Residuals
Resid2=Y-Yhat2

# Compute Parameter Standard Errors
se2=matrix(0,p2,1)
SSE2=t(Resid2) %*% Resid2
sighat2=SSE2/(N-p2)
for(i in 1:p2){
se2[i]=sighat2^(1/2)*H2.2[i,i]^(1/2)
}

R.sq2 <- 1-SSE2/sum((Y-mean(Y))^2)
```

This model has the parameter estimate $\hat{\beta} = (\hat{\mu}, \hat{\alpha}, \hat{\gamma}, \hat{a}, \ldots, \hat{k}) =$

```r
theta_hat2
```

```
##                [,1]
##  [1,]  3.149523e+02
##  [2,]  6.741279e-02
##  [3,]  8.746040e-05
##  [4,]  6.235220e-02
##  [5,]  6.904217e-01
##  [6,]  1.447816e+00
##  [7,]  2.582703e+00
##  [8,]  3.021581e+00
##  [9,]  2.312617e+00
## [10,]  6.871456e-01
## [11,] -1.475834e+00
## [12,] -3.170156e+00
## [13,] -3.251152e+00
## [14,] -2.040991e+00
```

When we wrote the design matrix, we dropped S_{12} and the associated parameter, l to make the design matrix full rank. However, we can compute $\hat{l}$ since the sum of the seasonal terms must be 1. Thus $\hat{l} = 1-(\hat{a}+\cdots+\hat{k}) =$

```r
1-sum(theta_hat2[3:14])
```

```
## [1] 0.133409
```

We also care about the SSE and the R^2 for this model:

```
SSE2
```

```
##          [,1]
## [1,] 412.5167
```

```
R.sq2
```

```
##           [,1]
## [1,] 0.9992366
```

Both of these models have extremely high R^2 values, which indicates they fit the dataset very well, and the in-sample estimates are very close to the observed values.

Problem 4:

Run an F-test at level 95% between the 2 models.

First, we note that an F-test is appropriate for comparing nested models, a condition which is satisfied here. In class we learned that for an F-test we have the following test statistic:

$F = \frac{SSE_o - SSE_a}{\hat{\sigma^2}(p_a - p_o)}$, where $\hat{\sigma^2} = SSE_a/(n - p_a)$.

We reject H_o if our test statistic, $F > F_{\alpha, p_a - p_o, n - p_a}$.

We implement this test for the two models from problem 3 below.

```
sigma.sq.hat <- SSE2/(N-p2)
F.stat <- (SSE-SSE2)/(sigma.sq.hat*(p2-p))
F.stat
```

```
##          [,1]
## [1,] 44.15693
```

```
qf(p=0.95,df1=p2-p,df2=N-p2)
```

```
## [1] 1.893126
```

The test statistic we computed if much greater than the value from appropriate F-distirubtion at level 95%. This F-test indicates the second model (one with a monthly seasonal terms) is significantly better at explaining the pattern in the dataset than the model with the sinusoidal seasonal patttern.

Problem 5:

First, we note that since $E(\vec{L^T}\vec{Y}) = \vec{\hat{X}}^T\vec{\hat{\beta}}\forall\vec{\beta}$ this implies $\vec{L^T}\vec{X} = \vec{X^T}$. Now, we have that:

$$
\begin{aligned}
Cov\Big(\vec{L^T}\vec{Y} - \hat{\vec{X}^T}\vec{\hat{\beta}}, \hat{\vec{X}^T}\vec{\hat{\beta}}\Big) = & \\
&= E\Big((\vec{L^T}\vec{Y} - \hat{\vec{X}^T}\vec{\hat{\beta}})(\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) - E(\vec{L^T}\vec{Y} - \hat{\vec{X}^T}\vec{\hat{\beta}})E(\hat{\vec{X}^T}\vec{\hat{\beta}}) \\
&= E\Big((\vec{L^T}\vec{Y}\hat{\vec{X}^T}\vec{\hat{\beta}} - \hat{\vec{X}^T}\vec{\hat{\beta}}\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) \\
&- E(\vec{Y^T} - \hat{\vec{X}^T}\vec{\hat{\beta}})E(\hat{\vec{X}^T}\vec{\hat{\beta}}) \\
&= E\Big((\vec{Y^T}\hat{\vec{X}^T}\vec{\hat{\beta}} - \hat{\vec{X}^T}\vec{\hat{\beta}}\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) \\
&- (\vec{Y^T} - E(\hat{\vec{X}^T}\vec{\hat{\beta}}))E(\hat{\vec{X}^T}\vec{\hat{\beta}}) \\
&= E\Big((\vec{Y^T}\hat{\vec{X}^T}\vec{\hat{\beta}} - \hat{\vec{X}^T}\vec{\hat{\beta}}\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) \\
&- (\vec{Y^T} - \vec{Y^T})E(\hat{\vec{X}^T}\vec{\hat{\beta}}) \\
&= E\Big((\vec{Y^T} - \hat{\vec{X}^T}\vec{\hat{\beta}})\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) - 0 \\
&= E\Big((\vec{Y^T} - \hat{\vec{X}^T}\vec{\hat{\beta}})\hat{\vec{X}^T}\vec{\hat{\beta}})\Big) \\
&= E\Big(\vec{Y^T} - \hat{\vec{X}^T}\vec{\hat{\beta}}\Big)E(\hat{\vec{X}^T}\vec{\hat{\beta}}) \\
&= 0
\end{aligned}
$$

Problem 6:

If $\vec{Y}$ is jointly normally distributed, show that its components are independent if and only if its covariance matrix is diagonal in form.

First, assume that the components of $\vec{Y}$ are independent, therefore, for $i \neq j$,

$Cov(Y_i, Y_j) = E(Y_iY_j) - E(Y_i)E(Y_j) = E(Y_i)E(Y_j) - E(Y_i)E(Y_j) = 0.$

Accordingly, all the off diagonal elements of the covariance matrix are zero, and thus the covariance matrix is diagonal.

Now, assume that all pairs of Y_i, Y_j have covariance $= 0$. This means that the covariance matrix is independent, and thus the joint multivaraite normal distribution for $Y_1, \ldots, Y_n$ can be factored into n univaraite normal distributions. Since the joint distribution can be factored into the product of marginal distributions, we know that the $Y_i \perp Y_j \forall i \neq j$.

Problem 7:

We fit the model by the same process from earlier problems. This gives us a estimate $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\gamma})$

```
d=60#2018-1959+1
T=12
N=d*T
p=3 # updated this

# Define Observations and Times
# Initialize time (tax) and observation (Y) arrays
tax=1:N
tyr=1959+tax/12
```

```
Y=loa$v3
#Plot the data

# Define Design Matrix
D=matrix(0,N,p)
for(i in 1:N){
D[i,1]=1
D[i,2]=i
D[i,3]=i^2
# add this extra column to the design matrix, what tau shouls be is not clear - I just left tau out and
# than when i tried subtracting the month
}

# Compute the OLS estimators
H1=t(D) %*% D
H2=solve(H1)
H3= H2 %*% t(D)
theta_hat= H3 %*% Y

# Compute Estimated Y
Yhat= D %*% theta_hat

# Compute Residuals
Resid=Y-Yhat

# Compute Parameter Standard Errors
se=matrix(0,p,1)
SSE=t(Resid) %*% Resid
sighat=SSE/(N-p)
for(i in 1:p){
se[i]=sighat^(1/2)*H2[i,i]^(1/2)
}

R.sq <- 1-SSE/sum((Y-mean(Y))^2)

theta_hat
```

```
##               [,1]
## [1,] 3.149908e+02
## [2,] 6.730812e-02
## [3,] 8.745631e-05
```

Now, we implement the formula from class notes for the $(1-\alpha)$ Confidence Interval for $\hat{\beta} \pm t_{\alpha/2,df=n-p} Var(\hat{})^{1/2}\hat{\beta}$. For $\alpha = 0.05$, we have the confidence interval for γ.

```
alpha <- 0.05
theta_hat_ci_upp <- theta_hat + qt(1-alpha/2, N-p)*sqrt(se)
theta_hat_ci_low <- theta_hat - qt(1-alpha/2, N-p)*sqrt(se)
theta_hat_ci_low[3];theta_hat_ci_upp[3]
```

```
## [1] -0.002784004
```

```
## [1] 0.002958917
```

We note that this 95% confidence interval for γ includes zero, this provides evidence in support of the hypothesis that $\gamma = 0$. In context of this model, I conclude that inclusion of the quadratic term does not significantly improve the model's explanation of the response variance.

Problem 8:

We have the one way anove model $Y_{i,j} = \mu_i + \epsilon_{i,j}$ for L groups and m data points within each group. Accordingly, there are $n = m * L$ data points. Also, we are assuming normal independent errors centered at zero. It follows that $\hat{\mu_i} = \bar{Y_i}$. Now, we want to consider an F-test to determine whether all the μ_i parameters are the same.

Under the null hypothesis, we have $X\vec{\beta} = (1, \ldots, 1_n)^T(\hat{\mu_o})$.

We want the projcetion matrix $M = X(X^TX)^{-1}X^T$. Say $m = 2, L = 5$, this gives us the projection matrix:

```
m=2
L=5
n=m*L

x=matrix(rep(1,n))

Mo=x%*%solve(t(x)%*%x)%*%t(x)
Mo
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [2,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [3,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [4,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [5,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [6,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [7,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [8,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
##  [9,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
## [10,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
```

Under the alternative hypothesis, we are estimating different means for each group. This gives us:

$$X = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

where each block has m rows and $\vec{\beta} = (\mu_1, \mu_2, \ldots, \mu_L)^T$. Again, let us say $m = 2, L = 5$, this gives us the projection matrix under the alternative hypothesis:

```
m=2
L=5
n=m*L

c1=matrix(c(rep(1,m),rep(0,n-m)),ncol=1)
c2=matrix(c(rep(0,m),rep(1,m),rep(0,n-2*m)),ncol=1)
c3=matrix(c(rep(0,2*m),rep(1,m),rep(0,n-3*m)),ncol=1)
c4=matrix(c(rep(0,3*m),rep(1,m),rep(0,n-4*m)),ncol=1)
c5=matrix(c(rep(0,4*m),rep(1,m)),ncol=1)

x=cbind(c1,c2,c3,c4,c5)

Ma=x%*%solve(t(x)%*%x)%*%t(x)
Ma
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]  0.5  0.5  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0
##  [2,]  0.5  0.5  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0
##  [3,]  0.0  0.0  0.5  0.5  0.0  0.0  0.0  0.0  0.0   0.0
##  [4,]  0.0  0.0  0.5  0.5  0.0  0.0  0.0  0.0  0.0   0.0
##  [5,]  0.0  0.0  0.0  0.0  0.5  0.5  0.0  0.0  0.0   0.0
##  [6,]  0.0  0.0  0.0  0.0  0.5  0.5  0.0  0.0  0.0   0.0
##  [7,]  0.0  0.0  0.0  0.0  0.0  0.0  0.5  0.5  0.0   0.0
##  [8,]  0.0  0.0  0.0  0.0  0.0  0.0  0.5  0.5  0.0   0.0
##  [9,]  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.5   0.5
## [10,]  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.5   0.5
```

From class notes we have the following formula for the F-test statistic:

$$\begin{aligned} F - stat &= \frac{SSE_o - SSE_a}{r - 1/\hat{\sigma^2}} \\ &= \frac{\vec{Y^T}(M_a - M_o)\vec{Y}}{r-1} / \frac{SSE_a}{n-r} \\ &= \frac{\vec{Y^T}(M_a - M_o)\vec{Y}}{r-1} \frac{n-r}{\vec{Y^T}(I - M_a)\vec{Y}} \end{aligned}$$

In this F-test, we compare the comuted test statsitic and compare it to the appropriate α level of a F-distribution with $r-1$ and $n-r$ degrees of freedom, where r is the number of parameters in the alternative model.

Problem 9:

We can follow the exact process from the previous problem. We write in the observed data, create the design matrix for the null hypothesis (one μ parameter for all 3 destitute individuals) and the design matrix for the alternative model (each of the 3 pieces of trash gets their own μ parameter). Then, we compute the test statistic and the theoretical 95th percentile from the null hypothesis.

```r
exotic<-c(4,3,7,0,4,2,5)
baskin<-c(3,7,5,2)
rowe<-c(8,4,0,7,3,1)
n<-length(exotic)+length(baskin)+length(rowe)
y<-matrix(c(exotic,baskin,rowe))
r<-3

x.o<-matrix(rep(1,n))
m.o<-x.o%*%solve(t(x.o)%*%x.o)%*%t(x.o)

x.a<-cbind(matrix(c(rep(1,7),rep(0,10)),ncol=1),
           matrix(c(rep(0,7),rep(1,4),rep(0,6)),ncol=1),
           matrix(c(rep(0,10),rep(1,6),ncol=1)))
m.a<-x.a%*%solve(t(x.a)%*%x.a)%*%t(x.a)

F.stat<-((t(y)%*%(m.a-m.o)%*%y)/(r-1))*((n-r)/(t(y)%*%(diag(n)-m.a)%*%y))

F.stat
```

```
##              [,1]
## [1,] -1.357373
```

```r
qf(p=0.95,df1=r-1,df2=n-r)
```

```
## [1] 3.738892
```

The F-test statistic is less than the theoretical 95th quantile, so this F-test fails to reject the null hypothese at $\alpha = 0.05$, so we conclude that the 3 villians all have the same breeding rate.

Problem 10:

Given $X \sim \chi^2_m$ random variable, we know that $E(X) = m$ and $Var(X) = 2m$. Since the degrees of freedom for a random variable is a constant, we know that

$E(X/m) = E(X)/m = 1 \forall m \in 1, \ldots, n$ and

$Var(X/m) = \frac{1}{m^2} Var(X) = \frac{2m}{m^2} = \frac{2}{m}$.

It follows that as $lim_{m\to\infty} Var(X/m) = 0$ and thus $lim_{m\to\infty} P\left(\left|\frac{X_m}{m} - 1\right| > \epsilon\right) = 0 \forall \epsilon > 0$.

In the previous examples, we have seen that the denominator of a F test statistic is $SSE_a/(n-r) \sim \chi^2_{n-r}$. Thus as this converges to 1, the F-test under consideration converges to a Chi-squared test of independence. In other words, the F-test statistic converges to a χ^2_n/n where n is the degrees of freedom of the numerator.