R

Question 2 (setup)

```
library(Matrix)
```

```
## Warning: package 'Matrix' was built under R version 3.5.2
```

```
y = c(43,45,47,46,48,33,37,38,35,56,54,57)
X = matrix(c(rep(1,12),rep(1,5),rep(0,7),rep(0,5),rep(1,4),rep(0,3),rep(0,9),rep(1,3)),12,4)
n = dim(X)[1]
r = rankMatrix(X)[1]
```

Question 2. a)

```
xtx = t(X)%*%X
xtxc = diag(c(0,1/5,1/4,1/3))

xtxc
```

```
##      [,1] [,2] [,3]      [,4]
## [1,]    0  0.0 0.00 0.0000000
## [2,]    0  0.2 0.00 0.0000000
## [3,]    0  0.0 0.25 0.0000000
## [4,]    0  0.0 0.00 0.3333333
```

Question 2. b) (helper)

```
b = xtxc%*%t(X)%*%y

b
```

```
##          [,1]
## [1,]  0.00000
## [2,] 45.80000
## [3,] 35.75000
## [4,] 55.66667
```

Question 2. c)

```
tt = c(4,2,1,1)

tt == round(tt%*%xtxc%*%xtx,3)
```

```
##      [,1] [,2] [,3] [,4]
## [1,] TRUE TRUE TRUE TRUE
```

Question 2. d)

```
tt1 = c(1,1,0,0)
e = y - X%*%b
s2 = sum(e^2)/(n-r)
ta = qt(0.975, df=(n-r))
CI = c(tt1%*%b) + c(-1,1)*c(ta*sqrt(s2)*sqrt(1+t(tt1)%*%xtxc%*%tt1))

CI
```

```
## [1] 40.96818 50.63182
```

Question 2. e)

```
C = matrix(c(0,0,1,-1),1,4)
m = rankMatrix(C)[1]
SS = t(C%*%b)%*%solve(C%*%xtxc%*%t(C))%*%C%*%b
Fstat = (SS/m)/s2

pf(Fstat, m, n-r, lower=F) < 0.05
```

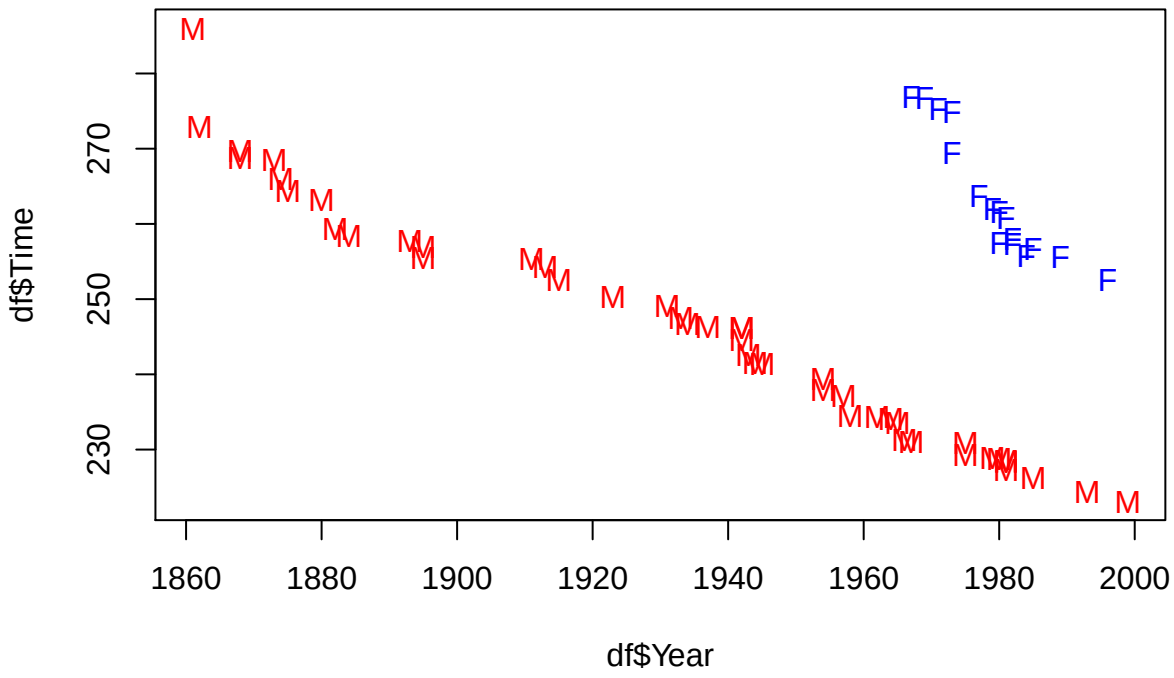
```
##      [,1]
## [1,] TRUE
```

Question 4 (setup)

```
setwd("C:\\Users\\akira\\Dropbox\\University\\Linear Statistical Models\\Lab Data")
df = read.csv("mile.csv")
```

Question 4. a)

```
palette(c("blue","red"))
plot(df$Time~df$Year, pch=array(df$Gender), col=df$Gender)
```



Question 4. b)

```
amodel = lm(Time ~ Gender+Year, df)
imodel = lm(Time ~ Gender*Year+Gender+Year, df)

anova(amodel, imodel)
```

```
## Analysis of Variance Table
##
## Model 1: Time ~ Gender + Year
## Model 2: Time ~ Gender * Year + Gender + Year
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     59 895.62
## 2     58 518.03  1    377.59 42.276 2.001e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 4. c)

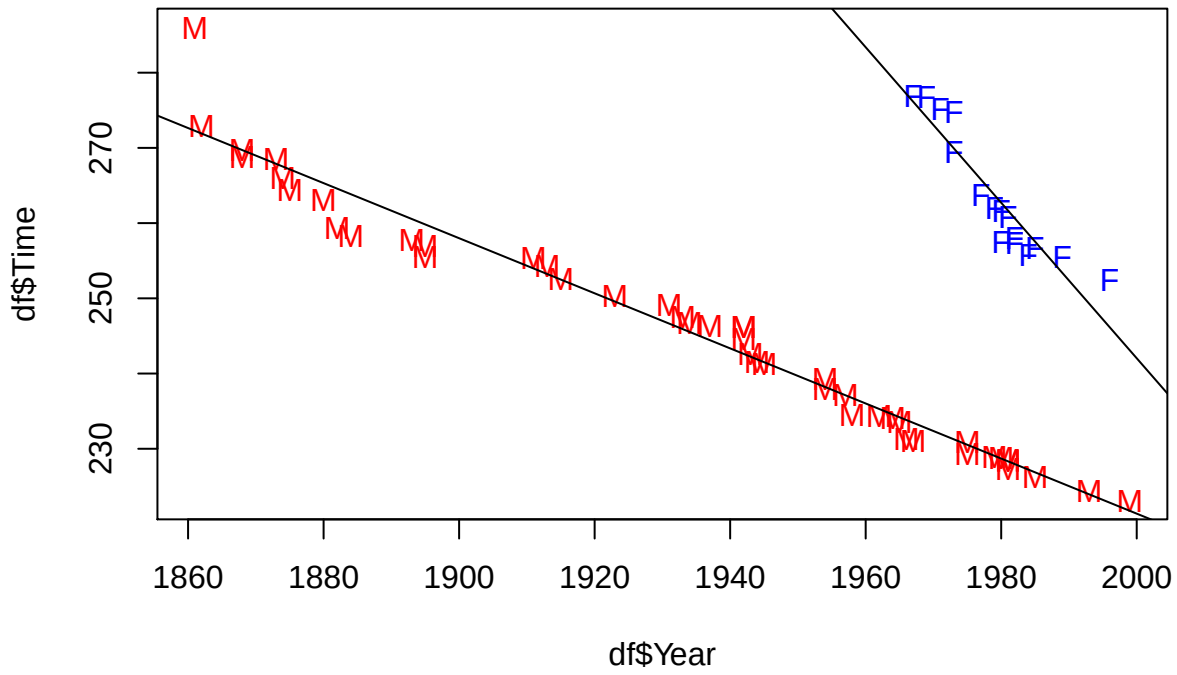
```
summary(imodel)
```

```
##
## Call:
## lm(formula = Time ~ Gender * Year + Gender + Year, data = df)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4512 -1.6160 -0.1137  1.1784 13.7265
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2309.4247   202.0583  11.429  < 2e-16 ***
## GenderMale      -1355.6778   203.1441  -6.673 1.03e-08 ***
## Year               -1.0337     0.1021 -10.126 1.95e-14 ***
## GenderMale:Year     0.6675     0.1027   6.502 2.00e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.989 on 58 degrees of freedom
## Multiple R-squared:  0.9663, Adjusted R-squared:  0.9645
## F-statistic: 553.8 on 3 and 58 DF,  p-value: < 2.2e-16
```

```
male = c(imodel$coefficients[1] + imodel$coefficients[2], imodel$coefficients[3] + imodel$coefficients[4])
female = c(imodel$coefficients[1], imodel$coefficients[3])

plot(df$Time~df$Year, pch=array(df$Gender), col=df$Gender)
abline(male)
abline(female)
```



Question 4. d)

```
point_estimate = -imodel$coefficients[2]/imodel$coefficients[4]

point_estimate
```

```
## GenderMale
##    2030.95
```

Question 4. e)

```
tt = c(0,1,-1, 0, 2031, -2031)
n = nrow(df)
p = length(df)
X = matrix(0, n, p)
y = df$Time

X[,1] = 1
mapper = unlist(Map({function(i) if (i=="Male") 1 else 2}, df$Gender))
X[cbind(1:n, mapper+1)] = 1
X[,4] = df$Year
X[cbind(1:n, mapper+4)] = df$Year

xtx = t(X)%*%X
xtxc = matrix(0, dim(X)[2], dim(X)[2])
xtxc[c(2:3,5:6),c(2:3,5:6)] = t(solve(xtx[c(2:3,5:6),c(2:3,5:6)]))
A = t(xtxc)%*%xtx

tt == round(tt%*%A)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE
```

Question 4. f)

```
ci = gmodels::estimable(imodel, c(0,0,0,1), conf.int=0.95)
c(ci$Lower, ci$Upper)
```

```
## [1] 0.4620087 0.8730100
```

Question 4. g)

```
car::linearHypothesis(imodel, c(0,0,1,1), -0.3)
```

```
## Linear hypothesis test
##
## Hypothesis:
## Year  + GenderMale:Year = - 0.3
##
## Model 1: restricted model
## Model 2: Time ~ Gender * Year + Gender + Year
```

```
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     59 850.63
## 2     58 518.03  1     332.6 37.238 9.236e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 5. a)

```
n3 = 100/(5/sqrt(10) + 1 + 1)
n2 = round(n3*1/2)
n1 = round(n3*1/sqrt(10))
n3 = round(n3) - 1 # for rounding
```

Question 5. b)

```
n = c(n1,n2,n3)
nsum = sum(n)
x = sample(nsum, nsum)
j1 = x[1:n[1]]
j2 = x[(n[1]+1):(n[1]+n[2])]
j3 = x[(n[1]+n[2]+1):nsum]

print("Treatment 1 Patients - $5000")
```

```
## [1] "Treatment 1 Patients - $5000"
```

```
(j1)
```

```
## [1] 36 27 40 39 50 24 31 43 41
```

```
print("Treatment 2 Patients - $2000")
```

```
## [1] "Treatment 2 Patients - $2000"
```

```
(j2)
```

```
##  [1] 15  8 38 45 13 28  2 23 21  4 18 29 44 37
```

```
print("Treatment 3 Patients - $1000")
```

```
## [1] "Treatment 3 Patients - $1000"
```

```
(j3)
```

```
##  [1]  7 32 25 10  9 11 16 22 33 35 20  3 17 46 26 19  5  1 47 48 30  6 49
## [24] 14 34 12 42
```