Preston Robertson
Summer 2022: Data Analysis
Homework VI
Due: TBA
Submit through Canvas

Instructions: Provided solutions to these questions using this template. Include graphics with your solutions. Put all code an appendix to this homework. Use the *verbatim* command to leave code unchanged.

1. Use *set.seed(1,2,3)* to simulate IID $\{\epsilon\}_{t=1}^{30}$ where $\epsilon \sim N(0, \sigma^2 = 25)$.

a) Plot $\{(t, Y_t)_{t=1}^{30}\}$ where Y_t follows the model

$$Y_i = \beta_0 + \beta_1 t + \epsilon_i \tag{1}$$

where $\beta_0 = 1$ and $\beta_1 = 2$.

Solution:
Figure (1) depicts the requested plot.

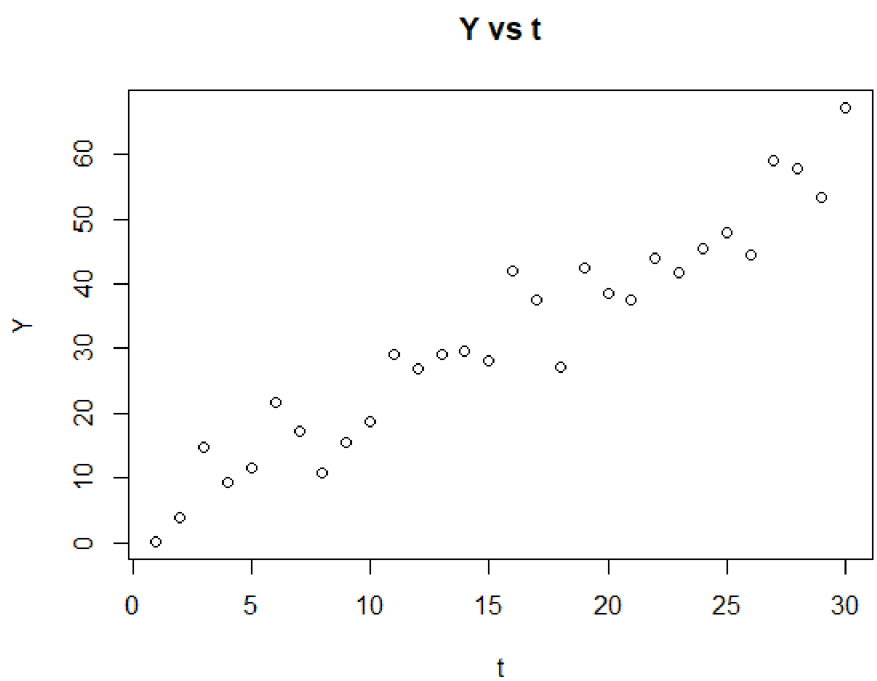

Figure 1: Linear plot with some errors.

b) Write out a design matrix $\mathbf{X}$ for the regression model in Equation (1).
Solution:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ & \\ 1 & N \end{bmatrix}$$

c) Obtain an estimate (β_0, β_1) by solving $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$.
Solution:
Our least squares estimates are

$$\hat{\boldsymbol{\beta}} = \begin{matrix} 2.671154 \\ 1.876889 \end{matrix}$$

d) Compute $\hat{\sigma}^2 = \frac{\mathbf{Y'(I-X(X'X)^{-1}X')Y}}{(30-2)}$.

Solution:

Tedious matrix computation yield $\hat{\sigma}^2 = 23.75$, which is a reasonable estimate.

e) Compute $Var(\hat{\boldsymbol{\beta}})$.
Solution:
From lecture we see that

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= \sigma^2 \left(\mathbf{X'X}\right)^{-1} \\ &= 25 \times \begin{bmatrix} 0.140229885 & -0.0068965517 \\ -0.006896552 & 0.0004449388 \end{bmatrix}. \end{aligned}$$

It is important to differentiate between the variance and standard errors. We define the standard error as $\hat{\sigma}^2$ times the covariance matrix. We would write

$$\begin{aligned} SE(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\boldsymbol{X'X})^{-1} \\ &= \hat{\sigma}^2 \times \begin{bmatrix} 0.140229885 & -0.0068965517 \\ -0.006869552 & 0.0004449388 \end{bmatrix} \\ &= 23.7 \times \begin{bmatrix} 0.140229885 & -0.0068965517 \\ -0.006869552 & 0.0004449388 \end{bmatrix} \end{aligned}$$

f) Test the hypothesis $H_0 : \beta_1 = 1$ vs. $H_a : \beta_1 > 1$.
Solution:

$$SE(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 3.3234483 & -.16344828 \\ -.16344828 & 0.01054505 \end{bmatrix}$$

Therefore we have a test statistic,

$$t = \frac{\hat{\beta}_1 - 1}{SE(\hat{\beta}_1)} = 83.165.$$

Which we have rejected at the 95% value.

g) Use *set.seed(1,2,3)* to simulate IID $\{\epsilon\}_{t=1}^{30}$ where $\epsilon \sim N(0,1)$. Obtain $\hat{\boldsymbol{\beta}}$. Repeat this process n = 2500 times. make a scatter plot of all 250 pairs of $\hat{\boldsymbol{\beta}}$. Estimate $Cov(\hat{\boldsymbol{\beta}}$.

Solution:

Below, Figure (2 depicts the results of the simulation study.

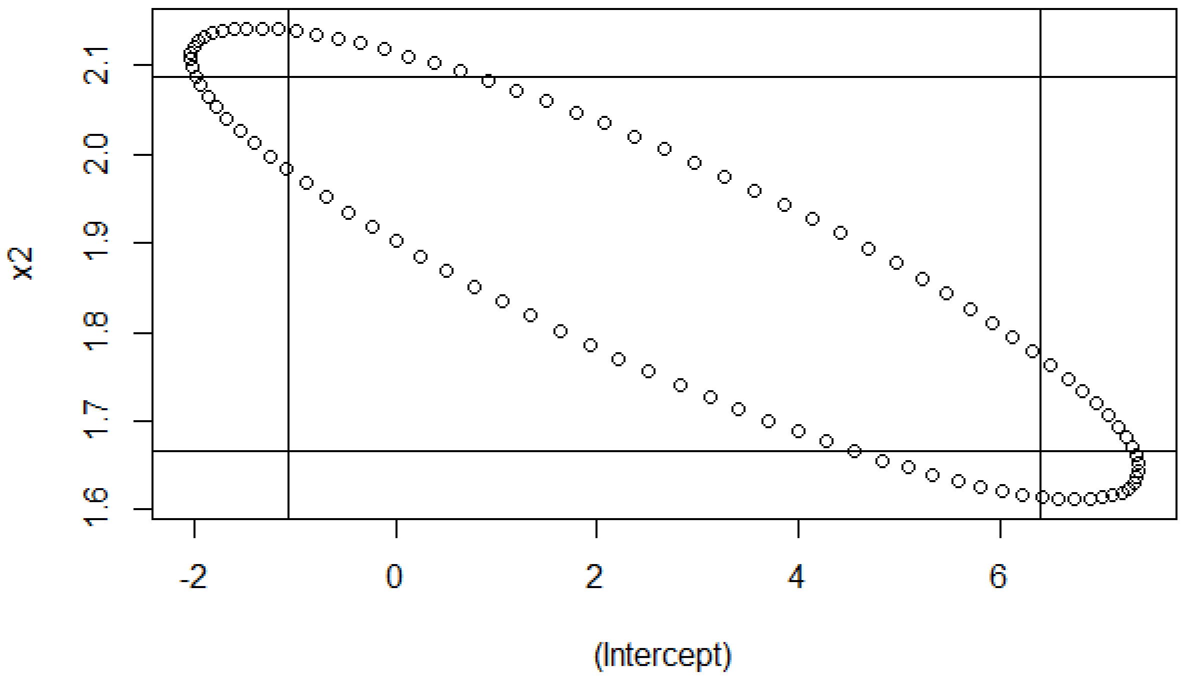

Figure 2: A simple linear regression with parameters $beta_0 = 1$ and β_1 was simulated and the parameters where estimated. This graphic depicts those estimated pairs.

h) Plot a 95% confidence ellipse about $\hat{\boldsymbol{\beta}}$.

Solution:
Here, the R package "ellipse" was used to plot the 95% confidence ellipse for the last estimated pair of $\hat{\beta}_0$, $\hat{\beta}_1$.

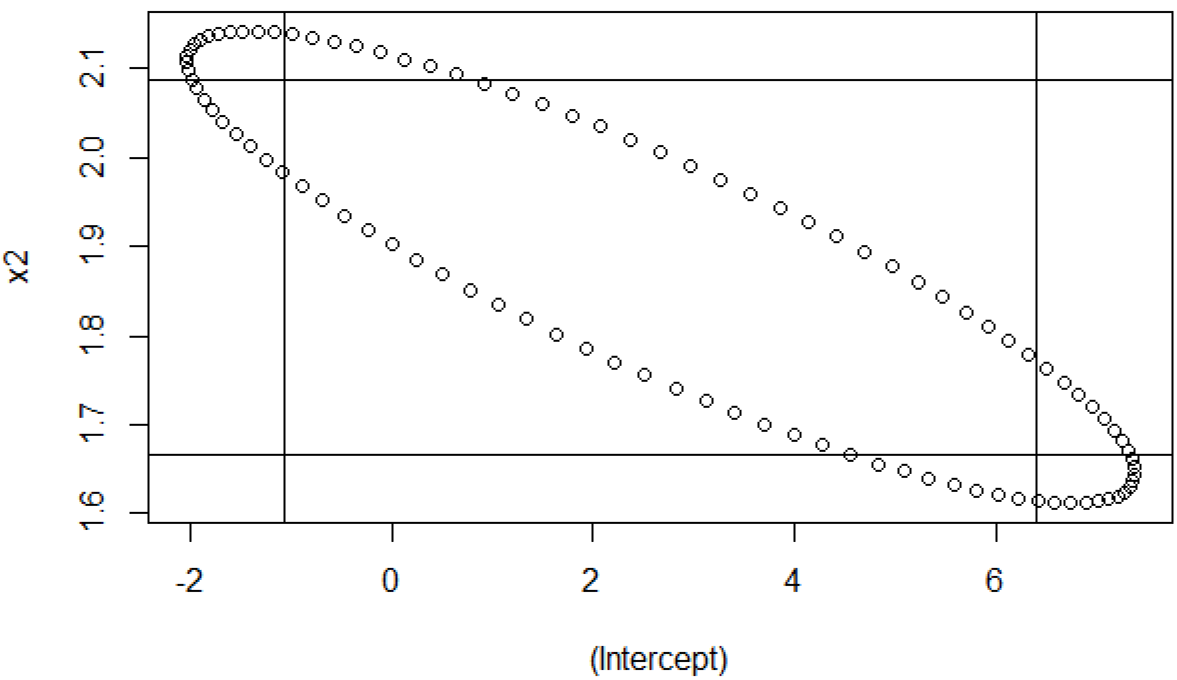

Figure 3: A single 95% confidence ellipse. The vertical and horizontal lines depict the uni-variate 95% confidence intervals for each parameter.

2. Use the Plotly package to plot a bivariate normal distribution with mean $\boldsymbol{\mu} = [0, 1]'$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 3 & 0.2 \\ 0.2 & 1 \end{pmatrix}.$$

Solution:

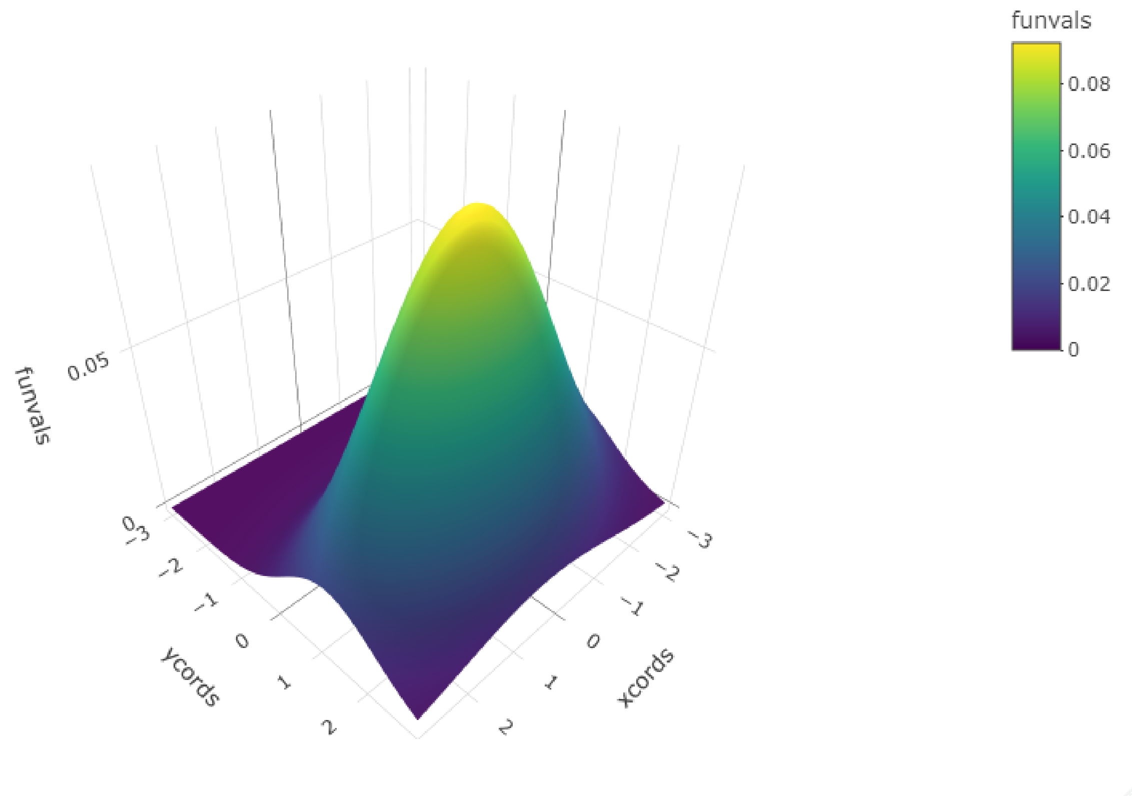

Figure 4: Caption

3. The Mauna Loa monthly CO_2 data series is freely available from the NOAA. You may access this data from the hyperlink below.

Mauna Loa

Fit data to the model

$$Y_t = \mu + \alpha_1 t + \alpha_2 t^2 + \gamma \cos\left(\frac{2\pi(t-\tau)}{12}\right) + \epsilon_t$$

where Y_t are the monthly average CO_2 observations. Assume $\{\epsilon_t\}$ are IID $N(0, \sigma^2)$ Let $t = 1$ being January 1959, $t = 2$ February 1959,..., with ending date December 2020. Use the methods discussed in class to obtain estimates for the model parameters and $\hat{\sigma}^2$. Plot Y_t and $\hat{Y}_t$ verses time on the same graph.

Solution:

For this analysis, we found each component of the equation individually then bound them together once found. This was done by in order first find the t and t^2 values. Then used R to find the final part of the linear regression for the oscillation.

```
    ---
title: "DA HW 6"
author: "Preston Robertson"
date: "6/22/2022"
output: pdf_document
---

```{r}
# Loading Libraries
library(plotly)
library(mvtnorm)
library(matrixcalc)
library(pracma)
library(data.table)
library(ellipse)
library(climate)

```


```{r}
# Introduction to Plotly
```
```

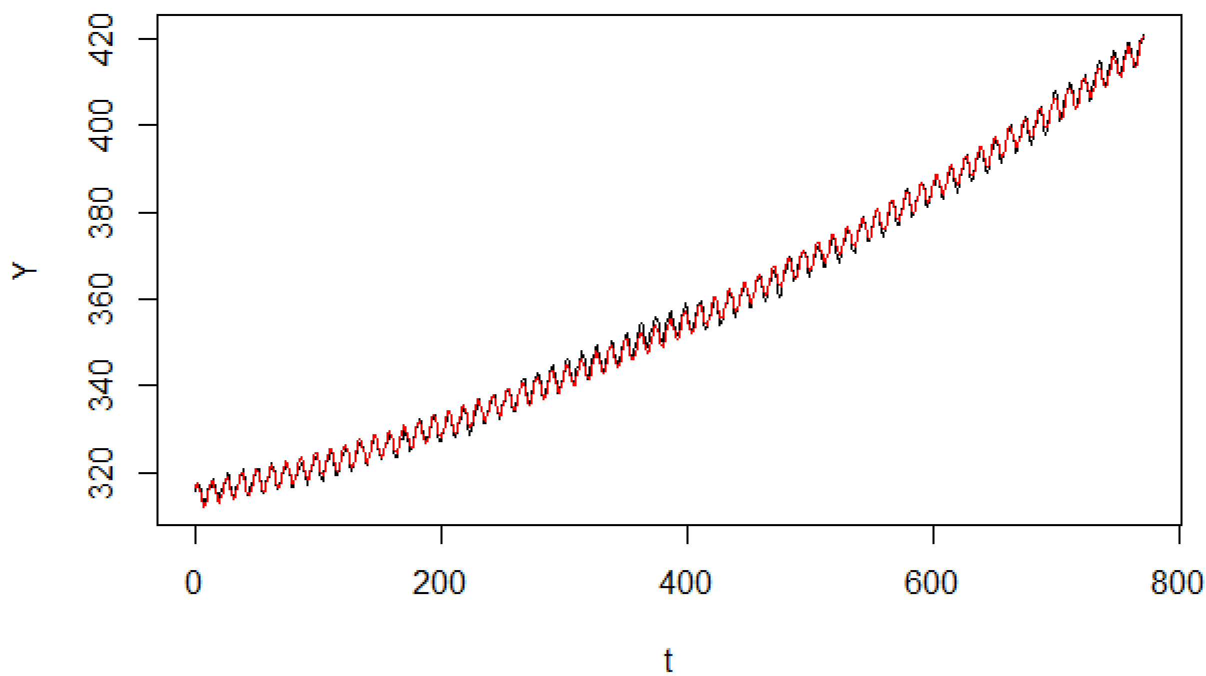

Figure 5: Plotting the regression formula using R.

```
fig <- plot_ly(z = ~volcano)
fig <- fig%>% add_surface()
fig


‘‘‘






‘‘‘{r}

xcords <- seq(-3,3, by = 0.1)
ycords <- seq(-3,3,by = 0.1)

lng = length(xcords)



funvals = matrix(0, lng, lng)



mu = c(0,0)
sig <- matrix(c(1,-0.5,-0.5,1), ncol = 2)
sig



for (i in c(1:lng)){
  for (j in c(1:lng)){
    funvals[i,j] = dmvnorm(c(xcords[j], ycords[i]), mean = mu, sigma = sig)
  }
}

fig <- plot_ly(type = 'surface', x = ~xcords, y = ~ycords, z = ~funvals)
fig

‘‘‘

Note that the ‘echo = FALSE‘ parameter was added to the code chunk to prevent printing of t
```

````
```{r}

is.positive.definite(sig, tol = 1e-8)


sig <- matrix(c(1,2,0,3), ncol = 2)
is.positive.definite(sig, tol = 1e-8)

```
````



````
```{r}

# Question 2

xcords <- seq(-3,3, by = 0.1)
ycords <- seq(-3,3,by = 0.1)

lng = length(xcords)



funvals = matrix(0, lng, lng)



mu = c(0,1)
sig <- matrix(c(3,0.2,0.2,1), ncol = 2)
sig



for (i in c(1:lng)){
  for (j in c(1:lng)){
    funvals[i,j] = dmvnorm(c(xcords[j], ycords[i]), mean = mu, sigma = sig)
  }
}

fig <- plot_ly(type = 'surface', x = ~xcords, y = ~ycords, z = ~funvals)
fig
```
````

```
‘‘‘{r}

# Question 1 part A

set.seed(123)

eps <- rnorm(30,0,5)
t <- c(1:30)
Y = rep(0,30)

beta0 = 1
beta1 = 2


for(i in t){
  Y[i] = beta0 + beta1*i + eps[i]
}


plot(t,Y, main = ’Y vs t’, xlab = ’t’, ylab = ’Y’)

‘‘‘
```

Missed Stuff

```
‘‘‘{r}

# Question 1b


‘‘‘
```

```
‘‘‘{r}

# Question 1C and 1D

x1<- rep(1,30)
x2<- c(1:30)
X<- cbind(x1,x2)
```

````
betahat = solve(t(x) %*% X) %*% t(x) %*% Y
betahat


I = eye(30)
sigmahatsqrd = t(Y)%*%(I - X%*%solve(t(X)%*%X)%*%t(x)%*%Y/(30-2))
view(sigmahatsqrd)

```


```{r}
# Question 1E




# Question 1F




```


```{r}
# Question 1G

set.seed(123)
n=2500
index = c(1:n)
beta1h = rep(0,n)
beta2h = rep(0,n)

for (i in index){
  eps = rnorm(30,0,5)
  betan = c(1,2)
  Y = X %*% betan + eps
  betahat = solve(t(x) %*% X ) %*% t(X) %*% Y
  beta1h[i] = betahat[1]
  beta2h[i] = betahat[2]
}

betahatot = cbind(beta1h, beta2h)
plot(betahatot)
````

```
‘‘‘


‘‘‘{r}


# Question 1G ???

lmod1 = lm(Y~x2)
summary(lmod1)
confint(lmod1)
plot(ellipse(lmod1, type = "l"))
abline(v = -1.063572)
abline(v = 6.405880)
abline(h = 1.666617)
abline(h = 2.087361)


‘‘‘




‘‘‘{r}
# Question 3
Mauna = meteo_noaa_co2()
head(Mauna)

Y = Mauna$co2_avg
size(Y)
t = c(1:771)
plot(t,Y, type = "l")
plot(Mauna$yy_d, Mauna$co2_avg, type='l') # What the one above is supposed to be
‘‘‘


‘‘‘{r}


lmodM = lm(Y ~ t)
lmodM
plot(lmodM)
plot(lmodM$residuals)


‘‘‘
```

````
```{r}

ts = rep(0, 744)
for (i in t){
  ts[i] = i^2
}

lmodMq = lm(Y~t+ts)
plot(t, lmodMq$residuals, type = 'l')
```


```{r}

T = 12
C = rep(0,744)

for (i in t){
  C[i] = cos(2*pi*i/T)
}

S = rep(0, 744)

for (i in t) {
  S[i] = sin(2*pi*i/T)
}

plot(t,C)
plot(t,S)

lmodS = lm(Y~t+ts+C+S)
plot(t, lmodS$residuals, type = "l")


```

```{r}

ones = rep(1, 744)

X = cbind(ones, t, ts, S, C)

#View(X)
````
````

```
betahat = solve (t(X) %*% X) %*% t(X) %*% Y

yhat = X %*% betahat
yhat

plot(t, Y, type = "l")
lines(t, yhat, col = "red")
```