

## Lab 3

[Vaishak Balachandra]

**Q.N. 1)** The gross domestic product (GDP) of the United States in trillions of dollars from 1950-2013 are provided in the link below

[http://media.pearsoncmg.com/aw/aw\\_sharpe\\_business\\_3/datasets/txt/GDP\\_2013.txt](http://media.pearsoncmg.com/aw/aw_sharpe_business_3/datasets/txt/GDP_2013.txt)

- Display the data using a scatterplot.
- Fit a simple linear regression model
- Add the fitted line to the scatter plot.
- Determine the coefficient of determination.
- Analyze the residual plots. Is your model questionable?
- Perform the Box-Cox transformation to improve the model.

```
> ##### LAB 3
> 
> #####
> 
> # Q1
> Q1 <- read.table("https://media.pearsoncmg.com/aw/aw_sharpe_business_3/datasets/txt/GDP_2013.txt",
+ header = T, sep = "\t")
> head(Q1,5)
  Year GDP...T.
1 2013   13.75
2 2012   13.67
3 2011   13.44
4 2010   13.18
5 2009   12.87
> dim(Q1)
[1] 64 2
> names(Q1) = c("Year", "GDP")
> head(Q1,5)
  Year  GDP
1 2013 13.75
2 2012 13.67
3 2011 13.44
4 2010 13.18
5 2009 12.87
> attach(Q1)
> 
> # a
> plot(Year, GDP, main = "Scatterplot: GDP against Year", pch = 17, cex = 1.2, col.main = "orange",
+ col.lab = "purple", col = "green")
> 
> # b
> model = lm(GDP~Year)
> model

Call:
lm(formula = GDP ~ Year)

Coefficients:
(Intercept)      Year
   -387.8433     0.1993

> cat("Fitted Model:
+ GDP = -387.8433 + 0.1993* Year")
Fitted Model:
```

```

GDP = -387.8433 + 0.1993* Year
>
> # c
> abline(model, lwd = 2, col = "red")
>
> # d
> summary(model)

Call:
lm(formula = GDP ~ Year)

Residuals:
    Min       1Q   Median       3Q      Max
-1.23604 -0.63427 -0.07458  0.51860  1.35054

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.878e+02  9.217e+00  -42.08  <2e-16 ***
Year         1.993e-01  4.651e-03   42.84  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6874 on 62 degrees of freedom
Multiple R-squared:  0.9673, Adjusted R-squared:  0.9668
F-statistic: 1835 on 1 and 62 DF, p-value: < 2.2e-16

> cat("R-squared value: 96.73%")
R-squared value: 96.73%
>
>
>
> # e
> plot(model,1)
> cat("Residual plot is not soo good!! -> as it has a clear structure(parabolic). Thus, needs some
transformation!!")
Residual plot is not soo good!! -> as it has a clear structure(parabolic). Thus, needs some transf
ormation!!
>
> # f
> install.packages("MASS")
> library(MASS)
> b = boxcox(model)
> b = boxcox(model, lambda = seq(-1,1))
> y1 = GDP^0.25
> new_model = lm(y1~Year)
> summary(new_model)

Call:
lm(formula = y1 ~ Year)

Residuals:
    Min       1Q   Median       3Q      Max
-0.039953 -0.010020 -0.000066  0.010955  0.028132

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.266e+01  2.015e-01  -112.5  <2e-16 ***
Year         1.224e-02  1.017e-04   120.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01503 on 62 degrees of freedom
Multiple R-squared:  0.9957, Adjusted R-squared:  0.9957
F-statistic: 1.448e+04 on 1 and 62 DF, p-value: < 2.2e-16

> new_model

Call:
lm(formula = y1 ~ Year)

```

Coefficients:

(Intercept)	Year
-22.66422	0.01224

```
> cat("New Fitted Model:  
+ GDP^0.2 = -22.66422 + 0.01224*Year")
```

```
New Fitted Model:  
GDP^0.2 = -22.66422 + 0.01224*Year
```

```
> cat("R-squared value: 99.57%")
```

```
R-squared value: 99.57%
```

```
> plot(new_model,1)
```

```
> cat("Residual plot has no clear structure, hence a better model!!")
```

Residual plot has no clear structure, hence a better model!!



