MATH 4753 Laboratory 14: Simple Linear Regression

In this lab we will investigate more detail concerning the mathematical underpinning of *Simple* Linear Regression.

Why is it called "Simple"? The reason is because we use only one independent variable (one x). When more than one x is used we call this *multiple* regression.

As I will show you in class:

$$SS_{xx} = \sum_{i=1}^{n} (x - \bar{x})^2$$

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Tasks

All output should be made through RMD. Please upload the following files:

- HTML
- RMD

All plots should be made through RMD and knitted into suitable formats.

You are expected to adjust the functions as needed to answer the questions within the tasks below.

- Task 1
 - o Make a folder LAB14
 - o Download the file "lab14.r"
 - o Place this file with the others in LAB14.
 - Start Rstudio
 - o Open "lab14.r" from within Rstudio.
 - Go to the "session" menu within Rstudio and "set working directory" to where the source files are located.
 - o Issue the function getwd() and copy the output here.
 - o Create your own R file and record the R code you used to complete the lab.
- Task 2
 - Make a function (mylsq) that will calculate estimates of the slope and intercept under least squares regression. The function will operate on two vectors of the same length, x and y, where x is the independent variable and y the dependent variable. It is partially made below. **Hint:** Use the formulae above.

mylsq=function(x,y) {

```
ssxx=sum((x-mean(x))^2 )
ssxy=sum() ## fill in the missing portion
blhat=ssxy/ssxx
b0hat= ## fill in the missing portion
return(list(b0hat=b0hat,b1hat= )) #fill in the missing portion
}
```

- o Suppose x=1:20 and set.seed(29); y=4+6*x + rnorm(20,0,5)
- Use mylsq() to calculate the least squares estimates of parameters β_0 , and β_1 .
- O Plot the points and the least squares line, with a heading and appropriate x and y labels. Also make the line have lwd=2 and be blue in colour. Hint: You can use abline()
- Check your calculations using $slr=lm(y\sim x)$; summary(slr).

• Task 3

O Now make a function that will predict the average y value from a given xnew. The function is mypred() will take three arguments, the x value, b0hat and b1hat.

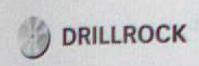
```
mypred=function(x,b0,b1){
ym=b0+ ## fill in the gap
ym
```

- O Use the same data in Task 2 and predict a new mean y value (\hat{y}) when xnew=15.5
- o Plot this point (xnew,ym) with the previous data and least squares line. **Hint:** use points(), cex=3,col="Green",pch=19
- O Use the functions you have created so far and answer 10.12 page 498 in MS 6th edition
 - a. Find the least-squares line for the data.
 - b. Interpret β_0 and β_1 in the words of the problem.
 - c. Predict the sweetness index if the amount of pectin in the orange juice is 300 ppm.

OJUICE

Run	Sweetness Index	Pectin (ppm)	
1	5.2	220 227	
2	5.5		
3	6.0	259	

- a)
- **■** b)
- c)
- Use the functions you have created so far and answer 10.80 page 553 in MS 6th edition



Depth at Which Drilling Begins x, feet	Time to Drill 5 Feet y, minutes	
0	4.90	
25	7.41	
50	6.19	

- a. Construct a scattergram for the data.
- b. Find the least-squares prediction equation.
- c. Graph the least-squares line on the scattergram.
- d. Interpret the values of β_0 and β_1 .
 - a)
 - b)
 - c`
 - d)
- Task 4
 - On page 501 MS proves that the least squares estimator $\hat{\beta}_1$ is an unbiased estimator of β_1 .
 - On page 503 MS shows that an unbiased estimator of σ^2 is

$$s^2 = \frac{SSR}{n-2}$$
, where $SSR = \sum (y_i - \hat{y}_i)^2$

• Complete the following function that calculates s^2

```
mysq=function(x,y) {
n=length(x)
ssxx=sum((x-mean(x))^2)
ssxy=sum() ## fill in the missing portion
b1hat=ssxy/ssxx
b0hat= ## fill
yhat=b0hat+ ## fill
ssr=sum((y-##)^2) # fill
sq= ## fill
return(list(ssr=ssr,sq=sq))
}
```

• Using x and y from Task 2 estimate σ^2 . How close did you get?

- O Now answer MS page 506 10.25 below
 - a. Plot the data points on a scattergram.
 - b. Fit a simple linear model relating carbon content in a pilot test, y, to the carbon content in a lab furnace, x. Interpret the estimates of the model parameters.
 - c. Compute SSE and s^2 .
- d. Compute s and interpret its value.



CARBON

Carbon Content (%)		Carbon Content (%)	
Pilot Plant	Lab Furnace	Pilot Plant	Lab Furnace
1.7	1.6	3.4	4.3
3.1	2.4	3.2	3.6

- **a**)
- **■** b)
- **c**)
- d)