# Homework 8

Arun Mahadevan Sathia Narayanan (as235872)

2025-04-11

***GitHub Link***:
To GitHub

***GitHub Link (Text Format)***:
https://github.com/arunmsn/SDS315/tree/main/Week_13/Homework_8

# Problem 1 - Regression Warm Up

## 1 - Part A

```
## [1] "Expected creatinine clearance rate for a 55 year old: 113.72 mL/minute"
```

The way I determined that the expected creatinine clearance rate for a 55 year old should be around 113.7 is by performing a linear regression on the given data, relating the `creatclear` column with the `age` column, and retrieving the coefficient values from the model. I then put the coefficient values, 147.8129158 for the intercept and -0.6198159 for the slope, multiplying the slope by 55 and adding the intercept (in y = mx + b format) to get the expected creatinine clearance rate.

## 1 - Part B

```
## [1] "Creatinine Clearance Rate change with Age: -0.62 mL/minute/year"
```

The way I determined how the clearance rate changes with age is by simply getting the slope, as determined in the Part A.

## 1 - Part C

```
## [1] "Difference for 40 year old: 11.98 mL/minute"
```

```
## [1] "Difference for 60 year old: 1.38 mL/minute"
```

The one with the higher difference in reality versus expected values is the 40 year old, suggesting the 40 year old is relatively healthier than the 60 year old. The way I calculated the values is by using the prediction model's intercept and correlation coefficient (similar to Part A), and predicted the creatinine clearance rate for the 40 year old and the 60 year old. Using the given values of 135 and 112 for the 40 and 60 year old, respectively, I was able to get the difference between the actual and expected value for each age. The difference for the 40 year old turned out to be 11.98 and the difference for the 60 year old turned out to be 1.38, which indicates the 40 year old is healthier.

# Problem 2 - Modeling Disease Growth

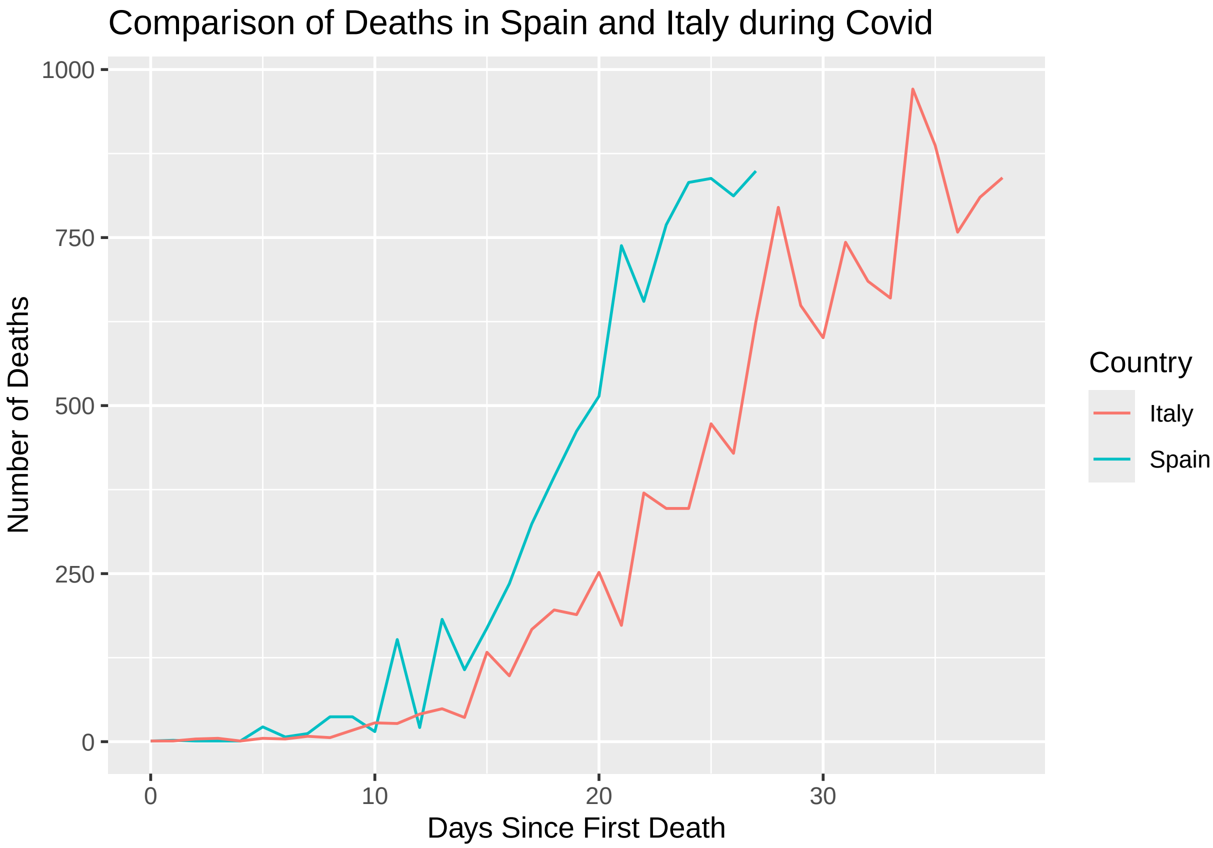
## 2 - Part 1

```
## [1] "Italy Data"
## [1] "The 95% confidence interval for the growth rate is:"
## [1] "Lower: 0.16 deaths/day"
## [1] "Upper: 0.209 deaths/day"
## [1] "The 95% confidence interval for the doubling time (using rule of 70) is:"
## [1] "Lower: 3.4 days"
## [1] "Upper: 4.4 days"
```

## 2 - Part 2

```
## [1] "Spain Data"
## [1] "The 95% confidence interval for the growth rate is:"
## [1] "Lower: 0.236 deaths/day"
## [1] "Upper: 0.318 deaths/day"
## [1] "The 95% confidence interval for the doubling time (using rule of 70) is:"
## [1] "Lower: 2.2 days"
## [1] "Upper: 3 days"
```

## 2 - Part 3



Comparison of Deaths in Spain and Italy during Covid
1000
750
500
250
0
Number of Deaths
0
10
20
30
Days Since First Death
Country
Italy
Spain

# Problem 3 - Price Elasticity of Demand

```
## [1] "Milk Price Elasticity of Demand"
## [1] "The 95% confidence interval for the Price Elasticity of Demand for Milk is:"
## [1] "Lower: -1.776 %"
## [1] "Upper: -1.458 %"
## [1] "Estimate: -1.619 %"
```

The estimated value, using power law, for the Price Elasticity of Demand (PED) for Milk is -1.619, and the 95% confidence interval for the PED for Milk is -1.775 to -1.458. Since PED is a ratio of percentage change in quantity and percentage change in price, and the absolute values of the PED values are all greater than 1, this means that the product, which in this case is milk, is *elastic*. This means a 1% increase in price leads to more than a 1% decrease in sales, which can be observed from the -1.775% and the -1.458% both being a greater loss than -1%. I estimated the elasticity by using a power law model (linear regression with both axes on log scale), setting the x to be log(price) and the y to be log(sales). Getting the correlation coefficient (as similarly done in the previous problems) gave the PED, which was -1.619.