

Skin Cancer

Tarek El-Hajjaoui

2023-01-30

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/SkinCancer/skincancer.csv'
Skincancer = read.table(file_path, header=TRUE, sep=";", dec=".")
```

```
summary(Skincancer)
```

```
##      State      Lat      Mort      Ocean
## Length:49      Min.   :28.00      Min.   : 86.0      Min.   :0.000
## Class :character 1st Qu.:36.00      1st Qu.:128.0      1st Qu.:0.000
## Mode  :character Median :39.50      Median :147.0      Median :0.000
##              Mean  :39.53      Mean  :152.9      Mean  :0.449
##              3rd Qu.:43.00      3rd Qu.:178.0      3rd Qu.:1.000
##              Max.   :47.50      Max.   :229.0      Max.   :1.000
##
##      Long
## Min.   : 69.00
## 1st Qu.: 78.50
## Median : 89.50
## Mean   : 90.94
## 3rd Qu.:100.00
## Max.   :121.00
```

Creating a Linear Model where latitude (X) is predictive of mortality rate (Y) due to skin cancer.

```
model <- lm(Mort ~ Lat, data=Skincancer)
summary(model)

##
## Call:
## lm(formula = Mort ~ Lat, data = Skincancer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.972 -13.185   0.972  12.006  43.938
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 389.1894    23.8123   16.34 < 2e-16 ***
## Lat        -5.9776     0.5984   -9.99 3.31e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.12 on 47 degrees of freedom
## Multiple R-squared:  0.6798, Adjusted R-squared:  0.673
## F-statistic: 99.8 on 1 and 47 DF, p-value: 3.309e-13
```

Find and interpret a 99% confidence interval for the mean mortality rate when Lat=40.

```
predict(model, list(Lat=40), level = 0.99, interval= "c")
```

```
##          fit      lwr      upr  
## 1 150.0839 142.7148 157.453
```

Find and interpret a 99% prediction interval for the individual mortality rate when Lat=40.

```
predict(model, list(Lat=40), level = 0.99, interval= "p")
```

```
##          fit      lwr      upr  
## 1 150.0839  98.24214 201.9257
```

Plotting 99% confidence and prediction intervals:

```
predictions <- predict(model, level = 0.99, interval="prediction")  
new_df <- cbind(Skincancer, predictions)  
  
ggplot(new_df, aes(x=Lat, y=Mort))+  
  geom_point() +  
  geom_line(aes(y=lwr), color = "red", linetype = "dashed")+  
  geom_line(aes(y=upr), color = "red", linetype = "dashed")+  
  geom_smooth(method=lm, se=TRUE)
```

