Linear Regression Analysis- Utah Population Projection

Written by Savahnna L. Cunningham

Date: September 15, 2018

```
# Load the Data
UT_data<- read_csv("~/Desktop/MSDA Portfolio/2. DS Tools & Techniques/P2. R/ut_census.csv")

View(UT_data)
```

Introduction

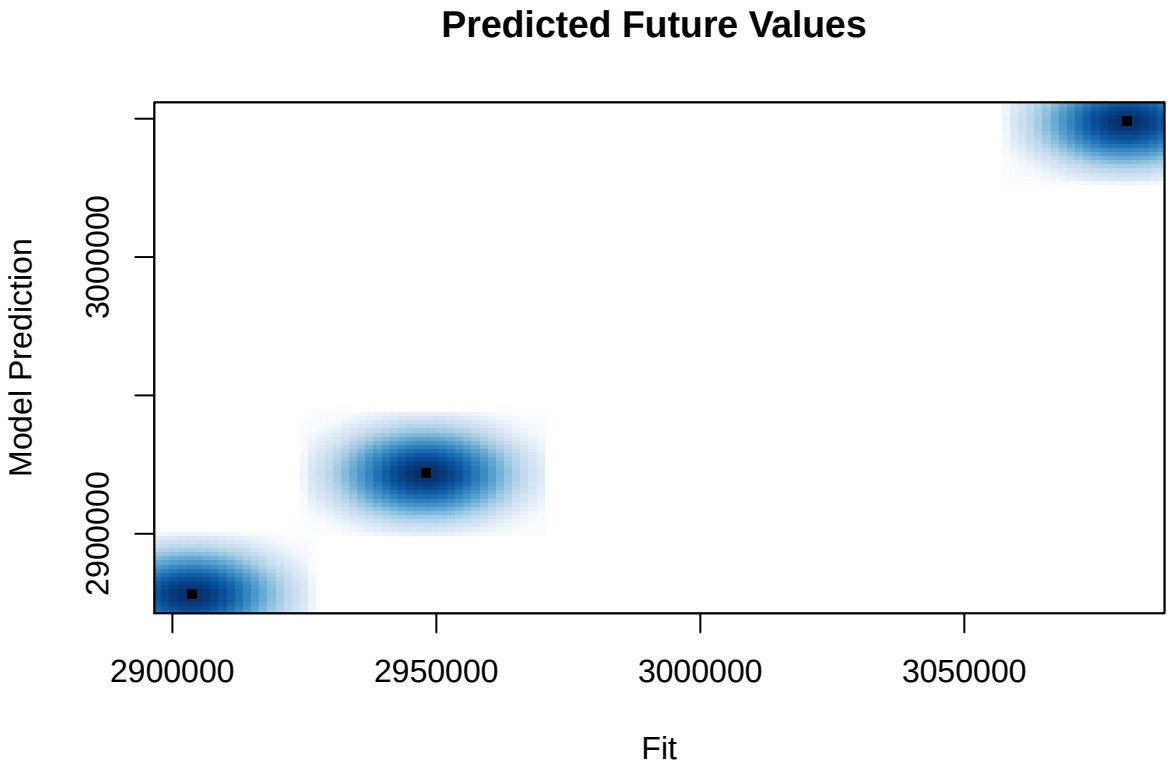
Conduct a linear regession analysis using the U.S. Census Bureau data for the state of Utah to predict the size of the population in 2020.

```
# Split the dataset into the Training set and Test set
set.seed(123)
split = sample.split(UT_data$Pop_Est, SplitRatio = 2/3)
training_set = subset(UT_data, split == TRUE)
test_set = subset(UT_data, split == FALSE)

# Fit the Simple Linear Regression to the Training set
regressor = lm(formula = Pop_Est ~ Year,
               data = training_set)
```

```
#Predict the Test set results
y_pred = predict.lm(regressor, newdata = test_set,interval = "prediction",level = 0.95)

#Visualize the result
smoothScatter(y_pred,pch = ".", cex = 5,
                     col = "black",colramp =
                     colorRampPalette(c("white", blues9)),
                     xlab = "Fit",
                     ylab = "Model Prediction",
                     main="Predicted Future Values ")
```



The visualization represents the 95% prediction interval with data points representing the models predicted values. As you can see, the model did very well predicting the population estimate, as all data points are within the prediction interval.

```
summary(regressor)
```

```
##
## Call:
## lm(formula = Pop_Est ~ Year, data = training_set)
##
## Residuals:
##       1       2       3       4       5
##  4420.4   308.1 -5182.2 -7334.0  7787.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -86236535    2847644  -30.28 7.91e-05 ***
## Year            44282       1415   31.30 7.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7324 on 3 degrees of freedom
## Multiple R-squared:  0.9969, Adjusted R-squared:  0.9959
## F-statistic: 979.7 on 1 and 3 DF,  p-value: 7.165e-05
```

```
#Confidence Interval for the Utah population estimate for the year 2020
y_pred <- lm(regressor, data = test_set)
new_data <-data.frame(Year=2020)
utah_predict<- predict(y_pred,new_data,interval="predict",level=.95)
utah_predict
```

```
##       fit     lwr     upr
## 1 3254141 3002358 3505925
```

Using a 95% confidence interval, the results of the linear regression model indicate the estimated population for the state of Utah will be between 3,002,358 and 3,505,925 people.

```
#Utah state population estimate for the year 2020
utah_predict<- predict(y_pred,new_data)
utah_predict
```

```
##       1
## 3254141
```

The Linear Regression model result for the 2020 Utah state population is 3,254,141 people.

Population prediction for the state of Utah in 5 years time.

```
#Confidence Interval for the Utah population estimate for the year 2023
y_pred <- lm(regressor, data = test_set)
new_data <-data.frame(Year=2023)
utah_predict<- predict(y_pred,new_data,interval="predict",level=.95)
utah_predict
```

```
##       fit     lwr     upr
## 1 3408259 3049965 3766552
```

Using a 95% confidence interval, the results of the linear regression model indicate the estimated population for the state of Utah will be between 3,049,965 and 3,766,552 people.

```
##       1
## 3408259
```

Predicted Utah state population for the year 2023 is 3,408,259 people.