|  |
| --- |
| UNIVERSITY OF SOUTH FLORIDA |
| Project |
| Ganti Uday, |
|  |

|  |
| --- |
|  |

Project

Ganti Uday

2022-11-12

**Data Source:**

### <https://www.kaggle.com/datasets/syedasimalishah/credit-card-limit-prediction?resource=download>

## About Dataset

A credit card is a payment card issued to users (cardholders) to enable the cardholder to pay a merchant for goods and services based on the cardholder's accrued debt (i.e., promise to the card issuer to pay them for the amounts plus the other agreed charges). The card issuer (usually a bank or credit union) creates a revolving account and grants a line of credit to the cardholder, from which the cardholder can borrow money for payment to a merchant or as a cash advance. There are two credit card groups: consumer credit cards and business credit cards. Most cards are plastic, but some are metal cards (stainless steel, gold, palladium, titanium), and a few gemstone-encrusted metal cards.

**Excel file taken:**

## credit.csv

**Variables Used:**

Income – [Continuous Independent Variable] Customers Income Amount

Limit – [Dependent Target Variable] Credit Limit

Married - [Binary Independent Variable] Marital Status

Balance - [Continuous Independent Variable] Current Balance in bank account

### Reading and cleaning Data

rm(list=ls())  
library(rio)  
library(car)

## Loading required package: carData

library(corrplot)

## corrplot 0.92 loaded

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':  
##   
## recode

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

set.seed(123)  
data = import("credit.csv")  
colnames(data)=tolower(make.names(colnames(data)))  
attach(data)  
head(data)

## income limit rating cards age education gender student married ethnicity  
## 1 14.891 3606 283 2 34 11 Male No Yes Caucasian  
## 2 106.025 6645 483 3 82 15 Female Yes Yes Asian  
## 3 104.593 7075 514 4 71 11 Male No No Asian  
## 4 148.924 9504 681 3 36 11 Female No No Asian  
## 5 55.882 4897 357 2 68 16 Male No Yes Caucasian  
## 6 80.180 8047 569 4 77 10 Male No No Caucasian  
## balance  
## 1 333  
## 2 903  
## 3 580  
## 4 964  
## 5 331  
## 6 1151

data = subset(data, select=-c(ethnicity,gender,student,education,cards,rating,age))  
data$married [data$married == "Yes"] = 1  
data$married [data$married == "No"] = 0  
data$married <- as.integer(data$married)  
str(data)

## 'data.frame': 400 obs. of 4 variables:  
## $ income : num 14.9 106 104.6 148.9 55.9 ...  
## $ limit : int 3606 6645 7075 9504 4897 8047 3388 7114 3300 6819 ...  
## $ married: int 1 1 0 0 1 0 0 0 0 1 ...  
## $ balance: int 333 903 580 964 331 1151 203 872 279 1350 ...

data %>% count(married)

## married n  
## 1 0 155  
## 2 1 245

head(data)

## income limit married balance  
## 1 14.891 3606 1 333  
## 2 106.025 6645 1 903  
## 3 104.593 7075 0 580  
## 4 148.924 9504 0 964  
## 5 55.882 4897 1 331  
## 6 80.180 8047 0 1151

cor(data)

## income limit married balance  
## income 1.00000000 0.79208834 0.03565236 0.46365646  
## limit 0.79208834 1.00000000 0.03115483 0.86169727  
## married 0.03565236 0.03115483 1.00000000 -0.00567349  
## balance 0.46365646 0.86169727 -0.00567349 1.00000000

### 

### Creating Sample Data

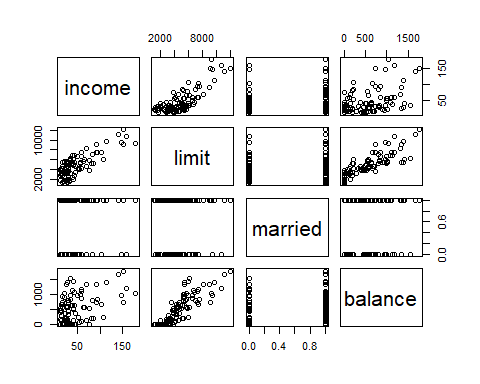
sample\_data = data[sample(1:nrow(data),100),]  
sample\_data %>% count(married)

## married n  
## 1 0 38  
## 2 1 62

attach(sample\_data)

## The following objects are masked from data:  
##   
## balance, income, limit, married

plot(sample\_data)

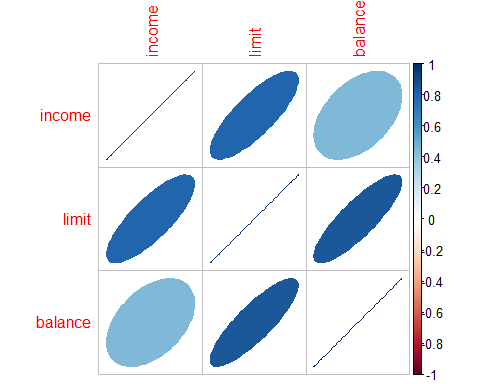


### Correlation between variables investigated

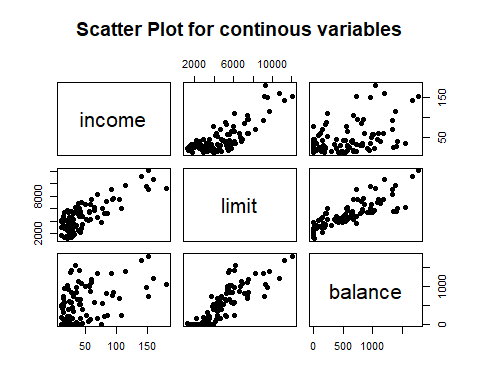
cont=subset(sample\_data, select=c("income","limit","balance"))  
  
corelated=round(cor(cont),3)  
  
round(cor(corelated),3)

## income limit balance  
## income 1.000 -0.074 -0.918  
## limit -0.074 1.000 0.464  
## balance -0.918 0.464 1.000

corrplot(corelated, method="ellipse")



var = cont[,c("income","limit","balance")]  
plot(var, col='black', pch=19,   
 main = 'Scatter Plot for continous variables')



### Linear Regression

regout1=lm(limit~income,data = sample\_data)  
summary(regout1)

##   
## Call:  
## lm(formula = limit ~ income, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2558.64 -1021.04 89.31 935.04 2350.17   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2638.702 214.824 12.28 <2e-16 \*\*\*  
## income 48.528 3.687 13.16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1357 on 98 degrees of freedom  
## Multiple R-squared: 0.6386, Adjusted R-squared: 0.6349   
## F-statistic: 173.2 on 1 and 98 DF, p-value: < 2.2e-16

coefficients(regout1)

## (Intercept) income   
## 2638.70196 48.52757

Inference:

When we run the linear regression model for limit (dependent) with income as a independent variable we see a P value less than 5% and has significance in the model. We also notice a R-square value of 63%. There is no difference between r-square and R- adjected which means there is no over fitting problem. This model can be better. As the R-square value is 63% which means there is still 37% of unexplained relationship in the dataset.Note: standard deviation is observed as 1357.

Regression Equation:

Limit = 48.52757\*income+2638.70196

regout2=lm(limit~married,data = sample\_data)  
summary(regout2)

##   
## Call:  
## lm(formula = limit ~ married, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3722.9 -1552.1 -321.0 914.6 7110.1   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4626.1 365.2 12.667 <2e-16 \*\*\*  
## married 329.8 463.8 0.711 0.479   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2251 on 98 degrees of freedom  
## Multiple R-squared: 0.005132, Adjusted R-squared: -0.00502   
## F-statistic: 0.5055 on 1 and 98 DF, p-value: 0.4788

coefficients(regout2)

## (Intercept) married   
## 4626.1053 329.7657

Inference:

When we run the linear regression model for limit (dependent) with married as a independent variable we see a P value less than 5%. But, we also notice a R-square value of 0.51% which means that this variable has no significance on the credit limit.Though, our p value is less than 5% R-sqaure value shows this model can be improved with other variables to get the best fit.

Regression Equation:

Limit = 329.7657 \*married+4626.1053

regout3=lm(limit~balance,data = sample\_data)  
summary(regout3)

##   
## Call:  
## lm(formula = limit ~ balance, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2885.9 -715.3 -89.0 496.2 3503.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2639.1818 183.4019 14.39 <2e-16 \*\*\*  
## balance 4.1166 0.2607 15.79 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1199 on 98 degrees of freedom  
## Multiple R-squared: 0.7178, Adjusted R-squared: 0.715   
## F-statistic: 249.3 on 1 and 98 DF, p-value: < 2.2e-16

coefficients(regout3)

## (Intercept) balance   
## 2639.181824 4.116578

Inference:

When we run the linear regression model for limit (dependent) with balance as a independent variable we see a P value less than 5% and has significance in the model. We also notice a R-square value of 71%. There is no difference between r-square and R- adjected which means there is no over fitting problem. This model can be better. As the R-square value is 71% which means there is still 29% of unexplained relationship in the dataset.Note: standard deviation is observed as 1199.

Regression Equation:

Limit = 4.116578\*balance+2639.181824

# Multiple regression with 2 independent variables

regout4=lm(limit~income+married,data = sample\_data)  
summary(regout4)

##   
## Call:  
## lm(formula = limit ~ income + married, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2567.48 -1013.12 81.17 925.91 2363.70   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2626.097 269.191 9.756 4.52e-16 \*\*\*  
## income 48.503 3.719 13.041 < 2e-16 \*\*\*  
## married 22.108 281.951 0.078 0.938   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1364 on 97 degrees of freedom  
## Multiple R-squared: 0.6387, Adjusted R-squared: 0.6312   
## F-statistic: 85.72 on 2 and 97 DF, p-value: < 2.2e-16

coefficients(regout4)

## (Intercept) income married   
## 2626.09739 48.50317 22.10766

Inference:

After understanding which variables are significant we run a multi-regression model for limit (dependent) with income and married as a independent variable. we notice a P value less than 5% for income (significance variable in the model) and married with bigger value indicating it has no affect on the credit limit. We also notice a R-square value of 63%. There is no difference between r-square and R- adjected which means there is no over fitting problem. This model has a scope for better improvement, as there is still 37% of unexplained relationship.

Regression Equation:

Limit = 48.50317\*income+22.10766\*married+2626.09739

regout5=lm(limit~balance+married,data = sample\_data)  
summary(regout5)

##   
## Call:  
## lm(formula = limit ~ balance + married, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2934.0 -694.2 -76.1 504.4 3450.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2554.1361 235.6388 10.839 <2e-16 \*\*\*  
## balance 4.1093 0.2619 15.690 <2e-16 \*\*\*  
## married 143.3910 248.1367 0.578 0.565   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1203 on 97 degrees of freedom  
## Multiple R-squared: 0.7188, Adjusted R-squared: 0.713   
## F-statistic: 124 on 2 and 97 DF, p-value: < 2.2e-16

coefficients(regout5)

## (Intercept) balance married   
## 2554.136147 4.109333 143.391043

regout6=lm(limit~balance+income,data = sample\_data)  
summary(regout6)

##   
## Call:  
## lm(formula = limit ~ balance + income, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1471.04 54.73 229.92 322.21 554.02   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1805.0492 96.6634 18.67 <2e-16 \*\*\*  
## balance 2.9759 0.1363 21.83 <2e-16 \*\*\*  
## income 31.9119 1.7037 18.73 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 560.9 on 97 degrees of freedom  
## Multiple R-squared: 0.9389, Adjusted R-squared: 0.9376   
## F-statistic: 745.1 on 2 and 97 DF, p-value: < 2.2e-16

coefficients(regout6)

## (Intercept) balance income   
## 1805.049228 2.975856 31.911902

Regression Equation:

Limit = 31.911902\*income+2.975856\*balance+1805.049228

# Main Multiple Regression with 3 independent variables

regout7 = lm(limit ~ income + married + balance, data=sample\_data)  
summary(regout7)

##   
## Call:  
## lm(formula = limit ~ income + married + balance, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1475.73 55.99 232.98 319.29 556.79   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1809.395 117.463 15.404 <2e-16 \*\*\*  
## income 31.920 1.717 18.594 <2e-16 \*\*\*  
## married -7.675 116.563 -0.066 0.948   
## balance 2.976 0.137 21.717 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 563.8 on 96 degrees of freedom  
## Multiple R-squared: 0.9389, Adjusted R-squared: 0.937   
## F-statistic: 491.6 on 3 and 96 DF, p-value: < 2.2e-16

coefficients(regout7)

## (Intercept) income married balance   
## 1809.395245 31.919780 -7.674806 2.975962

Inference:

After running regression model on two variables we got to know tht there is still room for improvement. so we ran a model with three three independent variables ( income, married and balance) we do notice that income and balance of a person is significant in the model with p-values less than 5%. R-square value is 93% which is a good fit. Overall income and balance of a person overrule the rest of the variables. we also notice this model is comparatively tigher with standard deviation of 563.

But we still want to make more models to see its behaviour.

Regression Equation:

Limit = 31.919780\*income+2.975962 \*balance-7.674806\*married+1809.395245

# X1 \* X2

sample\_data$incbal <- sample\_data$income\*sample\_data$balance  
  
regout8 = lm(limit ~ income + balance + incbal, data=sample\_data)  
summary(regout8)

##   
## Call:  
## lm(formula = limit ~ income + balance + incbal, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1475.08 41.01 235.14 321.13 538.93   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.764e+03 1.494e+02 11.808 < 2e-16 \*\*\*  
## income 3.296e+01 3.385e+00 9.736 5.49e-16 \*\*\*  
## balance 3.034e+00 2.113e-01 14.359 < 2e-16 \*\*\*  
## incbal -1.182e-03 3.291e-03 -0.359 0.72   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 563.4 on 96 degrees of freedom  
## Multiple R-squared: 0.939, Adjusted R-squared: 0.9371   
## F-statistic: 492.3 on 3 and 96 DF, p-value: < 2.2e-16

coefficients(regout8)

## (Intercept) income balance incbal   
## 1.764258e+03 3.296112e+01 3.033641e+00 -1.182135e-03

Inference:

After adding variable incbal( multiplication of income and balance) we notice this variable is not significant (p-value> 5%) still gives a R-sqaure value of 93% and a standard deviation of 563.4.

Let us run few more models.

Regression Equation:

Limit = 3.296112e+01\*income+2.975962 \*balance+3. -1.182135e-03\*incbal+1809.395245

sample\_data$income2 <- sample\_data$income\*sample\_data$income  
sample\_data$balance2 <- sample\_data$balance\*sample\_data$balance  
  
# variables income + income \*\* 2   
regout9=lm(limit~income+income2,data = sample\_data)  
summary(regout9)

##   
## Call:  
## lm(formula = limit ~ income + income2, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2555.60 -1031.80 55.93 965.82 2333.60   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2564.60494 361.45389 7.095 2.12e-10 \*\*\*  
## income 51.68668 12.90414 4.005 0.000121 \*\*\*  
## income2 -0.02021 0.07907 -0.256 0.798820   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1363 on 97 degrees of freedom  
## Multiple R-squared: 0.6389, Adjusted R-squared: 0.6314   
## F-statistic: 85.8 on 2 and 97 DF, p-value: < 2.2e-16

coefficients(regout9)

## (Intercept) income income2   
## 2564.60494090 51.68667869 -0.02020954

# variables balance +balance \*\*2  
  
regout10=lm(limit~balance+balance2,data = sample\_data)  
summary(regout10)

##   
## Call:  
## lm(formula = limit ~ balance + balance2, data = sample\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2822.2 -699.2 -64.5 509.6 3453.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.590e+03 2.212e+02 11.709 < 2e-16 \*\*\*  
## balance 4.412e+00 7.743e-01 5.699 1.3e-07 \*\*\*  
## balance2 -2.176e-04 5.364e-04 -0.406 0.686   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1204 on 97 degrees of freedom  
## Multiple R-squared: 0.7183, Adjusted R-squared: 0.7125   
## F-statistic: 123.7 on 2 and 97 DF, p-value: < 2.2e-16

coefficients(regout10)

## (Intercept) balance balance2   
## 2.589515e+03 4.412191e+00 -2.176327e-04

Inference :

After running a model with income+ income\*\*2 as independent variables we still see income beening significant and R -square value of 63% which is less than our previous models. Same with balance +balance square we see significance in balance but the variable balance square has no importance in the model.But the R-square value is 71% which is a good model fit. But we have already got a good model fit in the previous model.

Regression Equation:

Limit = 2.589515e +51.68667869 \*income+-0.02020954\*income2

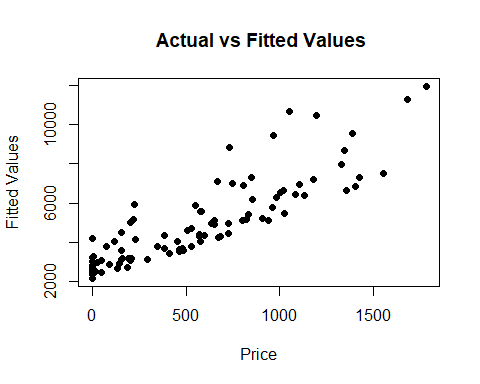
Limit = 2.589515e+4.412191e+00\*balance+-2.176327e-04\*balance2

##### Model——regout7 = lm(limit ~ income + married + balance, data=sample\_data)

is the best fit for the given data as it shows good R-sqare value of 93% invloves less variables and has less standard deviation compared to thde rest of the models

# Linearity

plot(sample\_data$balance, regout7$fitted.values, pch=19,   
 main = " Actual vs Fitted Values",xlab = 'Price', ylab = 'Fitted Values')

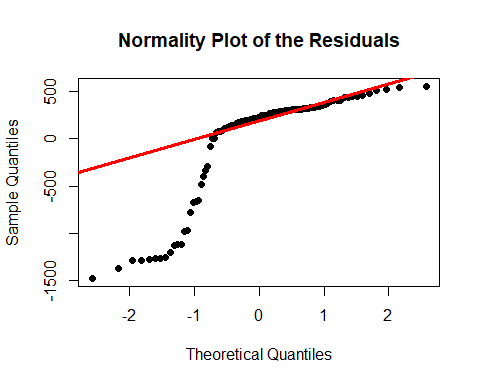


Inference :

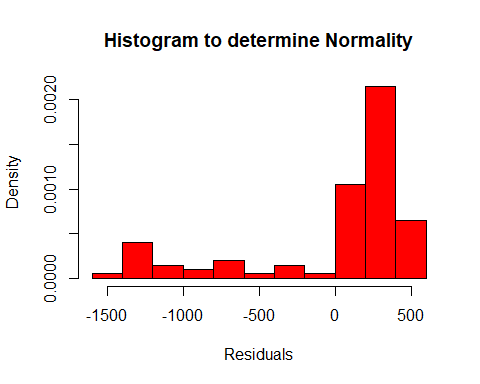
The data is linear in nature. There is a constant increase fitted values with increase in price.

# Normality

qqnorm(regout7$residuals,pch=19,main="Normality Plot of the Residuals")  
qqline(regout7$residuals,col='red',lwd=3)



hist(regout7$residuals, col='red', probability = TRUE,main = 'Histogram to determine Normality',xlab = 'Residuals')

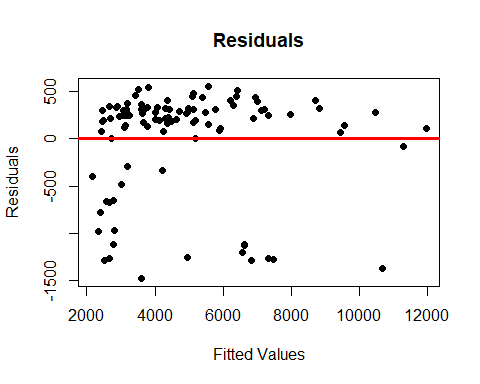


Inference :

While the data might not appear normally distributed initially, the data is left skewed for the most part.

# Equality of Variance

#Equality of Variance  
plot(regout7$fitted.values,regout7$residuals,pch=19,  
 main="Residuals",xlab = 'Fitted Values', ylab = 'Residuals')  
abline(0,0,col="red",lwd=3)



Inference :

Most of the observations fall inside the first 2 standard deviations.

# Prediction 1

predict(regout7, newdata= sample\_data, interval="predict")

## fit lwr upr  
## 179 4061.346 2926.966 5195.727  
## 14 6413.055 5274.962 7551.148  
## 195 2772.273 1636.287 3908.260  
## 306 2582.478 1446.454 3718.503  
## 118 8708.740 7563.093 9854.388  
## 299 2473.039 1332.690 3613.389  
## 229 4481.377 3336.264 5626.489  
## 244 6205.758 5075.249 7336.267  
## 399 3005.990 1869.554 4142.425  
## 374 6555.349 5421.563 7689.136  
## 153 3595.731 2458.076 4733.385  
## 90 7209.311 6063.721 8354.901  
## 91 5490.622 4346.742 6634.502  
## 256 4013.917 2871.587 5156.246  
## 197 7000.500 5863.864 8137.136  
## 388 2336.997 1196.445 3477.549  
## 348 10471.280 9280.830 11661.731  
## 137 3777.423 2634.747 4920.100  
## 355 3416.001 2282.643 4549.358  
## 328 6307.110 5166.997 7447.223  
## 26 4248.341 3112.584 5384.097  
## 7 3083.703 1947.184 4220.222  
## 383 6625.272 5462.082 7788.463  
## 254 5177.227 4033.728 6320.727  
## 211 2867.844 1734.137 4001.551  
## 78 2909.182 1771.838 4046.526  
## 81 3157.439 2025.240 4289.637  
## 43 4356.979 3228.196 5485.763  
## 359 3795.516 2665.853 4925.178  
## 332 3613.909 2480.357 4747.461  
## 143 7116.555 5965.766 8267.345  
## 32 2725.511 1589.549 3861.473  
## 109 5924.153 4767.167 7081.139  
## 263 3523.509 2389.352 4657.667  
## 330 7326.004 6188.984 8463.025  
## 23 2443.404 1307.122 3579.686  
## 309 4313.418 3173.507 5453.328  
## 135 5393.116 4260.208 6526.024  
## 364 5570.214 4440.903 6699.526  
## 224 4385.124 3254.823 5515.424  
## 166 4308.239 3177.486 5438.991  
## 217 2458.136 1318.635 3597.637  
## 290 3599.130 2460.931 4737.330  
## 69 5136.831 4002.406 6271.256  
## 72 6960.483 5824.049 8096.917  
## 76 2941.919 1806.693 4077.145  
## 63 2788.758 1648.167 3929.349  
## 141 7313.863 6158.237 8469.488  
## 210 8830.239 7653.070 10007.409  
## 347 5188.821 4049.757 6327.885  
## 390 6887.622 5746.339 8028.905  
## 294 11282.628 10107.633 12457.623  
## 277 3168.910 2032.651 4305.168  
## 41 3073.790 1934.227 4213.352  
## 389 6378.314 5231.154 7525.474  
## 316 3641.757 2508.890 4774.625  
## 223 7486.462 6314.530 8658.394  
## 16 2442.957 1306.674 3579.240  
## 116 4618.036 3484.349 5751.722  
## 94 5123.878 3977.124 6270.631  
## 262 10684.138 9483.724 11884.553  
## 235 7980.178 6836.108 9124.248  
## 86 11957.275 10771.742 13142.809  
## 39 3766.125 2627.705 4904.545  
## 159 5097.050 3968.321 6225.779  
## 240 3179.024 2046.880 4311.167  
## 209 4132.788 2995.128 5270.447  
## 397 3664.433 2527.006 4801.860  
## 34 2802.948 1666.934 3938.962  
## 4 9431.844 8250.866 10612.822  
## 13 4982.062 3840.396 6123.728  
## 357 5772.865 4631.016 6914.713  
## 243 2517.034 1381.224 3652.843  
## 308 4208.663 3061.307 5356.020  
## 278 4707.649 3579.449 5835.848  
## 89 4021.474 2887.798 5155.150  
## 25 2152.278 1011.155 3293.401  
## 291 3116.623 1984.300 4248.946  
## 286 2407.143 1270.762 3543.524  
## 336 3155.155 2023.212 4287.097  
## 121 2671.247 1535.286 3807.209  
## 110 3118.445 1985.508 4251.383  
## 158 6632.881 5498.703 7767.058  
## 64 2656.289 1518.017 3794.560  
## 199 2674.695 1538.734 3810.655  
## 67 9565.752 8412.647 10718.858  
## 151 5571.418 4442.157 6700.679  
## 85 3199.182 2056.789 4341.576  
## 165 5883.412 4750.973 7015.851  
## 136 3252.771 2110.450 4395.093  
## 51 4908.667 3779.194 6038.140  
## 74 4363.672 3227.558 5499.786  
## 178 3658.823 2528.143 4789.502  
## 236 2705.383 1570.928 3839.837  
## 98 3095.047 1962.630 4227.465  
## 214 4963.425 3829.064 6097.785  
## 127 6823.515 5663.088 7983.942  
## 212 5130.961 3992.506 6269.416  
## 174 4431.507 3294.916 5568.098  
## 273 4946.672 3809.880 6083.463

A prediction interval reflects the uncertainty around a single value, while a confidence interval reflects the uncertainty around the mean prediction values. Thus, a prediction interval will be generally much wider than a confidence interval for the same value.

Which one should we use? The answer to this question depends on the context and the purpose of the analysis. Generally, we are interested in specific individual predictions, so a prediction interval would be more appropriate. Using a confidence interval when you should be using a prediction interval will greatly underestimate the uncertainty in a given predicted value