

# Comparing Regression Predictions

eCornell

7/28/2021

This R Markdown file demonstrates how to use the results from simple and multiple linear regression analyses to predict a response value. Then, it demonstrates how to use the predict function with multiple linear regression. The example here makes predictions about prestige scores based on data from the Prestige data set.

## Step 1: Load the data and define colors.

```
knitr::opts_chunk$set(echo = TRUE)
# The Prestige data set is available in the carData library
library(carData)

# Load the Prestige data set
data(Prestige)
# Exclude any observations that do not have an entry in the type column
Prestige = Prestige[!is.na(Prestige$type),]
# Add log(income) as a variable to the Prestige data frame
Prestige$logincome <- log(Prestige$income)

#eCornell Hex Codes:
crimson = '#b31b1b' #Crimson
lightGray = '#cecece' #lightGray
darkGray = '#606366' #darkGray
skyBlue = '#92b2c4' #skyblue
gold = '#fbb040' #gold
ecBlack = '#393f47' #ecBlack
```

## Step 2: Use simple linear regression to make predictions.

Calculate the regression line based only on education, then use it to predict the prestige score of a job with eight years of education.

```
fit.edu <- lm(prestige ~ education, data = Prestige)
fit.edu

##
## Call:
## lm(formula = prestige ~ education, data = Prestige)
##
## Coefficients:
## (Intercept)      education
##      -10.841         5.388
```

```
predict.edu.8 = -10.841 + 5.388*8
```

### Step 3: Use multiple linear regression to make predictions.

The following regression uses both education and log(income) as predictor variables. They are separated in the formula by the + sign.

```
fit.edu.inc <- lm(prestige ~ education + logincome, data = Prestige)
fit.edu.inc
```

```
##
## Call:
## lm(formula = prestige ~ education + logincome, data = Prestige)
##
## Coefficients:
## (Intercept)      education      logincome
##      -101.188         4.038         12.056
```

### Step 4: Use the predict() function to make predictions.

You can use the predict() function to make predictions based on many predictor variables at once.

```
# Write out your prediction rule:
predict.edu.8.loginc.10 = -101.188 + 4.038*8 + 12.056*10

# Create a data frame that contains values of predictor variables for which to predict the values of re
new.dat = data.frame(education = c(8,8,8), logincome = c(8, 9, 10))
new.dat

##      education logincome
## 1           8          8
## 2           8          9
## 3           8         10

# Use the predict() function with the data frame you created
# The regression object is the first argument, the data frame for which to calculate predictions is the
predict(fit.edu.inc, newdata = new.dat)

##           1           2           3
## 27.56324 39.61964 51.67604
```