

Assignment 4

Miguel Perez Luna

9/24/2020

Contents

Linear Regression	2
Model Fit	4
Statistically Insignificant Variables	4
Values of Statistically Significant Individual Model Coefficients	5

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Before running my linear regression, I manipulated two of my categorical variables.

In my Means of Transportation to Work variable, I mutated the “Car, Truck, or Van” value to “1Car, Truck, or Van” so that all other values get compared to whether or not someone takes a car, truck, or van to work.

In my Educational Attainment variable, I mutated the “Regular high school diploma” value to “1High School Diploma” so that all other vlaues get compared to whether or not someone’s highest educational attainemnt is a high school diploma.

```
transpo_data <- read.csv("transpo_data.csv") %>%
  mutate(meansTW = case_when(
    JWTR_label=="Bicycle"~"Bicycle",
    JWTR_label=="Bus or trolley bus"~"Bus or\ntrrolley bus",
    JWTR_label=="Car, truck, or van"~"1Car, truck,\nor van",
    JWTR_label=="Motorcycle"~"Motorcycle",
    JWTR_label=="Other method"~"Other",
    JWTR_label=="Railroad"~"Railroad",
    JWTR_label=="Streetcar or trolley car (carro publico in Puerto Rico)"~"Streetcar or\ntrrolley car",
    JWTR_label=="Subway or elevated"~"Subway or\nelevated",
```

```

JWTR_label=="Taxicab"~"Taxicab",
JWTR_label=="Walked"~"Walk")) %>%
mutate(edu = case_when(
  SCHL_label=="Some college, but less than 1 year"~"Less 1 year of college",
  SCHL_label=="Regular high school diploma"~"High school diploma",
  SCHL_label=="Bachelor's degree"~"Bachelor's degree",
  SCHL_label=="1 or more years of college credit, no degree"~"More 1 yr of college,\nno degree",
  SCHL_label=="Master's degree"~"Master's degree",
  SCHL_label=="GED or alternative credential"~"GED",
  SCHL_label=="Doctorate degree"~"Doctorate degree",
  SCHL_label=="Associate's degree"~"Associate's degree",
  SCHL_label=="Grade 8"~"Grade 8",
  SCHL_label=="Grade 7"~"Grade 7",
  SCHL_label=="Grade 10"~"Grade 10",
  SCHL_label=="12th grade - no diploma"~"Grade 12 -\nno diploma",
  SCHL_label=="Grade 6"~"Grade 6",
  SCHL_label=="Grade 11"~"Grade 11",
  SCHL_label=="Professional degree beyond a bachelor's degree"~"Professional degree",
  SCHL_label=="Grade 9"~"Grade 9",
  SCHL_label=="No schooling completed"~"No schooling",
  SCHL_label=="Grade 5"~"Grade 5",
  SCHL_label=="Grade 4"~"Grade 4",
  SCHL_label=="Nursery school, preschool"~"Preschool",
  SCHL_label=="Grade 3"~"Grade 3",
  SCHL_label=="Grade 1"~"Grade 1",
  SCHL_label=="Kindergarten"~"Kindergarten",
  SCHL_label=="Grade 2"~"Grade 2"))

```

Linear Regression

After mutating those few variables, I ran the linear regression to predict Income based on variation in Sex, Means of Transportation to Work, Travel Time to Work, Number of Accessible Vehicles, Rent, and Age.

I included options(scipen = 999) so that all values would be reported in standard notation.

```

model <- lm(PINCP ~ SEX_label + meansTW + edu + JWMNP + vehicle + GRNTP + AGE,
            data = transpo_data)
options(scipen = 999)
summary(model)

```

```

##
## Call:
## lm(formula = PINCP ~ SEX_label + meansTW + edu + JWMNP + vehicle +
##     GRNTP + AGE, data = transpo_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -106817  -17346   -3709   11007  469745
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept) -13892.1871   2217.9958  -6.263
## SEX_labelMale 12660.7552    883.5243  14.330

```

## meansTWBicycle	-12587.9144	3254.9841	-3.867
## meansTWBus or\ntrolley bus	-8378.8701	2599.6721	-3.223
## meansTWMotorcycle	-358.6898	9079.4633	-0.040
## meansTWOther	2407.8624	4677.2547	0.515
## meansTWRailroad	4135.6233	6445.9454	0.642
## meansTWStreetcar or\ntrolley car	3528.8252	11489.2225	0.307
## meansTWSubway or\nelevated	-5633.2840	6253.3617	-0.901
## meansTWTaxicab	-2357.8065	8343.2471	-0.283
## meansTWWalk	-3823.7025	2043.6837	-1.871
## eduAssociate's degree	5033.8675	1816.2136	2.772
## eduBachelor's degree	18153.2293	1376.4953	13.188
## eduDoctorate degree	25053.5591	3828.7967	6.543
## eduGED	2293.3529	2321.6023	0.988
## eduGrade 1	13149.4021	25631.7847	0.513
## eduGrade 10	-7968.1836	3946.6425	-2.019
## eduGrade 11	-6731.8987	3143.6728	-2.141
## eduGrade 12 -\nno diploma	2566.6930	3584.7370	0.716
## eduGrade 2	-25435.4778	25640.2934	-0.992
## eduGrade 3	-2292.4248	20947.5795	-0.109
## eduGrade 4	-6182.5758	20964.2283	-0.295
## eduGrade 5	-1099.9128	14845.7144	-0.074
## eduGrade 6	-6435.2441	5141.6076	-1.252
## eduGrade 7	-10891.5747	9423.6588	-1.156
## eduGrade 8	-4502.4662	7471.4426	-0.603
## eduGrade 9	-6764.9026	4955.2926	-1.365
## eduKindergarten	-34218.5166	25640.8474	-1.335
## eduLess 1 year of college	4572.3733	1977.4870	2.312
## eduMaster's degree	25879.3543	1879.1949	13.772
## eduMore 1 yr of college,\nno degree	2966.3574	1499.8275	1.978
## eduNo schooling	-5635.8698	5237.6577	-1.076
## eduPreschool	-13198.9113	36245.7783	-0.364
## eduProfessional degree	62647.4429	3668.0107	17.079
## JWMNP	42.6007	20.5903	2.069
## vehicle	-2917.3641	449.7630	-6.486
## GRNTP	15.2303	0.8123	18.751
## AGEP	652.8961	35.1781	18.560
##	Pr(> t)		
## (Intercept)	0.000000003994	***	
## SEX_labelMale	< 0.0000000000000002	***	
## meansTWBicycle	0.000111	***	
## meansTWBus or\ntrolley bus	0.001274	**	
## meansTWMotorcycle	0.968488		
## meansTWOther	0.606708		
## meansTWRailroad	0.521164		
## meansTWStreetcar or\ntrolley car	0.758744		
## meansTWSubway or\nelevated	0.367704		
## meansTWTaxicab	0.777492		
## meansTWWalk	0.061390	.	
## eduAssociate's degree	0.005593	**	
## eduBachelor's degree	< 0.0000000000000002	***	
## eduDoctorate degree	0.0000000000644	***	
## eduGED	0.323270		
## eduGrade 1	0.607960		
## eduGrade 10	0.043528	*	

```

## eduGrade 11                                0.032276 *
## eduGrade 12 -\nno diploma                   0.474012
## eduGrade 2                                  0.321227
## eduGrade 3                                  0.912860
## eduGrade 4                                  0.768071
## eduGrade 5                                  0.940941
## eduGrade 6                                  0.210758
## eduGrade 7                                  0.247816
## eduGrade 8                                  0.546779
## eduGrade 9                                  0.172239
## eduKindergarten                           0.182074
## eduLess 1 year of college                   0.020795 *
## eduMaster's degree                         < 0.0000000000000002 ***
## eduMore 1 yr of college,\nno degree         0.047991 *
## eduNo schooling                            0.281952
## eduPreschool                               0.715757
## eduProfessional degree                     < 0.0000000000000002 ***
## JWMNP                                       0.038586 *
## vehicle                                    0.00000000000940 ***
## GRNTP                                       < 0.0000000000000002 ***
## AGEP                                        < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36210 on 6876 degrees of freedom
## Multiple R-squared:  0.2159, Adjusted R-squared:  0.2117
## F-statistic: 51.18 on 37 and 6876 DF, p-value: < 0.00000000000000022

```

Model Fit

The multiple R-squared value for this model is 0.216, which means that this model predicts about 22% of the variation in Income in this dataset.

Statistically Insignificant Variables

Several of the variables in this model do not reach statistical significance at the 95-percent confidence interval because they have p-values that are greater than 0.05.

For Means of Transportation to Work, the statistically insignificant variables are:

- Motorcycle
- Railroad
- Streetcar or Trolley Car
- Subway
- Taxicab
- Walking
- Other

For Educational Attainment, the statistically insignificant variables are:

- No Schooling

- Preschool
- Kindergarten
- Grade 1
- Grade 2
- Grade 3
- Grade 4
- Grade 5
- Grade 6
- Grade 7
- Grade 8
- Grade 9
- Grade 12, no high school diploma
- GED or alternative high school diploma

Values of Statistically Significant Individual Model Coefficients

All of the following variables had individual model coefficients that reached statistical significance at the 95-percent confidence interval because they have p-values that are less than 0.05.

Sex

The coefficient estimate for Male is 12660.7552. Since this variable is compared to Female, this tells us that on average and controlling for everything else in the model, males will earn \$12,660.76 more than females.

Means of Transportation to Work

For means of transportation to work, everything is compared to Car, Truck, or Van.

The coefficient for Bicycle is -12587.9144. This means that on average and controlling for everything else in the model, individuals who ride a bicycle to work earn \$12,587.91 less than those who travel by a Car, Truck, or Van.

The coefficient for Bus or Trolley Bus is -8378.8701. This means that on average and controlling for everything else in the model, individuals who ride a bicycle to work earn \$8,378.87 less than those who travel by a Car, Truck, or Van.

Educational Attainment

For educational attainment, everything is compared to earning just a high school diploma.

The coefficient for Grade 10 is -7968.1836. This means that on average and controlling for everything else in the model, individuals who only complete school through 10th grade earn \$7,968.18 less than those who earn a high school diploma.

The coefficient for Grade 11 is -6731.8987. This means that on average and controlling for everything else in the model, individuals who only complete school through 11th grade earn \$6,731.90 less than those who earn a high school diploma.

The coefficient for Less than 1 Year of College is 4572.3733. This means that on average and controlling for everything else in the model, individuals who attend college for less than one year earn \$4,572.37 more than those who earn just a high school diploma.

The coefficient for At Least 1 Year of College but No Degree is 2966.3574. This means that on average and controlling for everything else in the model, individuals who attend college for at least one year but never

earn a degree earn \$2,966.36 more than those who earn just a high school diploma. It's interesting, to me why this number would be lower than those who attend college for less than a year. The sample size of those who attended college for at least a year is larger than those who attended less than one year, so perhaps that reflects a more accurate variation in eventual income.

The coefficient for Associate's Degree is 5033.8675. This means that on average and controlling for everything else in the model, individuals who earn an Associate's degree earn \$5,033.87 more than those who earn just a high school diploma.

The coefficient for Bachelor's Degree is 18153.2293. This means that on average and controlling for everything else in the model, individuals who earn a bachelor's degree earn \$18,153.23 more than those who earn just a high school diploma.

The coefficient for Master's Degree is 25879.3543. This means that on average and controlling for everything else in the model, individuals who earn a master's degree earn \$25,879.35 more than those who earn just a high school diploma. This bodes well for my future!

The coefficient for Professional Degree (beyond a bachelor's) is 62647.4429. This means that on average and controlling for everything else in the model, individuals who earn a professional degree earn \$62,647.44 more than those who earn just a high school diploma.

The coefficient for Doctorate Degree is 25053.5591. This means that on average and controlling for everything else in the model, individuals who earn a Doctorate degree earn \$25,053.56 more than those who earn just a high school diploma.

Travel Time to Work

The coefficient for Travel Time to Work is 42.6007. This means that on average and controlling for everything else in the model, individuals earn \$42.60 more for every additional minute they travel to work. This is interesting and I wonder how this can be interpreted in a few different ways given the context of Colorado. There are few cities with dense urban cores with high concentrations of jobs in the city. Rather jobs are fairly spread out. In such cities, though, wealthier people might choose to live further from the center in suburbs, which tend to be more expensive. That could explain why a longer travel time is related to higher income. However, this coefficient does suggest that people with lower incomes tend to live closer to their jobs, which I don't think is typically thought of to be the trend in big cities.

Number of Vehicles Accessible

The coefficient for Number of Vehicles Accessible is -2917.3641. This means that on average and controlling for everything else in the model, individuals earn \$2,917.36 less for every additional vehicle they have access to/own. This is surprising to me, as I would have thought that having more vehicles would mean that you could afford them.

Rent

The coefficient for Rent is 15.2303. This means that on average and controlling for everything else in the model, individuals earn \$15.23 more for every additional dollar they spend on rent.

Age

The coefficient for Age is 652.8961. This means that on average and controlling for everything else in the model, individuals earn \$652.90 more for every additional year of their life.