EDS241: Assignment 1

Connor Flynn

01/21/2022

```
data <- read.xlsx(here('data/CES4.xlsx'))
```

```
data <- data %>%
  select(Census.Tract, Total.Population, California.County, Low.Birth.Weight, PM2.5, Poverty)
```

```
PM2.5_Concentration_California_Mean <- mean(data$PM2.5)
PM2.5_Concentration_California_Mean
```

```
## [1] 10.1527
```

A). The average concentration of PM2.5 acrosss all census tracts in California is 10.1527 micro grams /cubic meter.

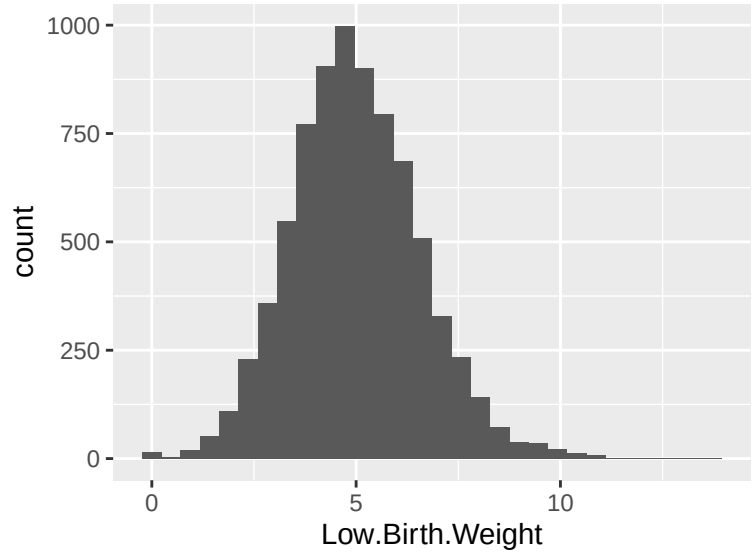
```
Poverty_California_County <- data %>%
  group_by(California.County) %>%
  summarise(Poverty_Mean_County = mean(Poverty, na.rm = TRUE))
```

B). Based on mean poverty at the county level, Tulare has the highest level of poverty in California.

C).

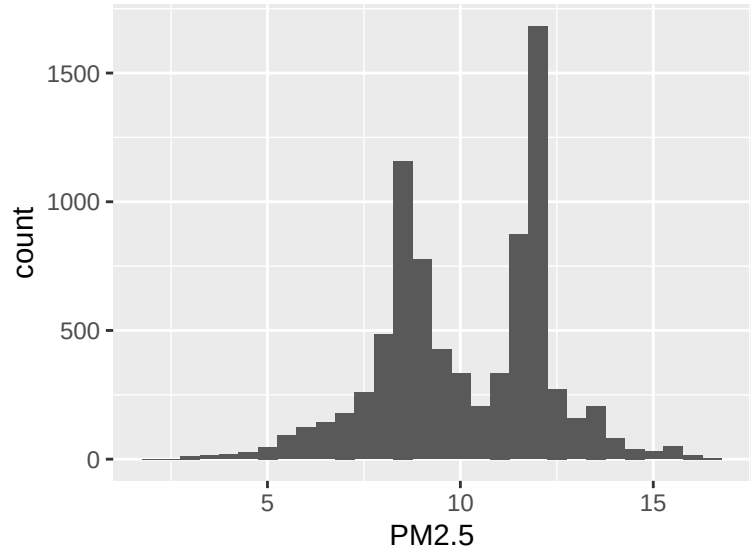
```
Low.Birth.Weight_hist <- ggplot(data = data, aes(x = Low.Birth.Weight)) +
  geom_histogram()

Low.Birth.Weight_hist
```



```r
PM2.5_hist <- ggplot(data = data, aes(x = PM2.5)) +
  geom_histogram()

PM2.5_hist
```




```r
model_1 <- lm_robust(Low.Birth.Weight ~ PM2.5, data = data)
summary(model_1)
```

```
##
## Call:
## lm_robust(formula = Low.Birth.Weight ~ PM2.5, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)   3.8010   0.088583   42.91 0.000e+00   3.6273   3.9746 7806
## PM2.5         0.1179   0.008402   14.04 3.256e-44   0.1015   0.1344 7806
##
## Multiple R-squared:  0.02499 ,   Adjusted R-squared:  0.02486
## F-statistic:   197 on 1 and 7806 DF,  p-value: < 0.00000000000000022
```

D). Estimated Slope Coefficient = 0.1179 Heteroskedasticity-robust standard error = 0.008402 For every one unit increase in PM2.5 concentration, there is a 0.1179 increase in percentage of low birth weights Based on the p value, the effect of PM2.5 on LowBirthWeight is statistically significant at the 95% Confidence level

```r
model_3 <- lm_robust(Low.Birth.Weight ~ PM2.5 + Poverty, data = data)
summary(model_3)
```

```
##
## Call:
## lm_robust(formula = Low.Birth.Weight ~ PM2.5 + Poverty, data = data)
##
## Standard error type:  HC2
```

```
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  3.54374   0.084733  41.823  0.000e+00  3.37764  3.70984 7802
## PM2.5        0.05911   0.008293   7.127  1.116e-12  0.04285  0.07536 7802
## Poverty      0.02744   0.001002  27.374 1.287e-157  0.02547  0.02940 7802
##
## Multiple R-squared:  0.1169 ,    Adjusted R-squared:  0.1167
## F-statistic: 494.8 on 2 and 7802 DF,  p-value: < 0.00000000000000022
```

F). For every one unit increase in poverty, there is a 0.02 increase in percentage of low birth weights. The estimated coefficient on PM2.5 decreases compared to the regression excluding poverty as an explanatory variable (D).
This makes sense because we've added an explanatory variable that effects the percentage of low birth weights, therefore decreasing the significance of PM2.5 in the regression.

```
linearHypothesis(model_3,c("PM2.5 = Poverty"), white.adjust = "hc2")
```

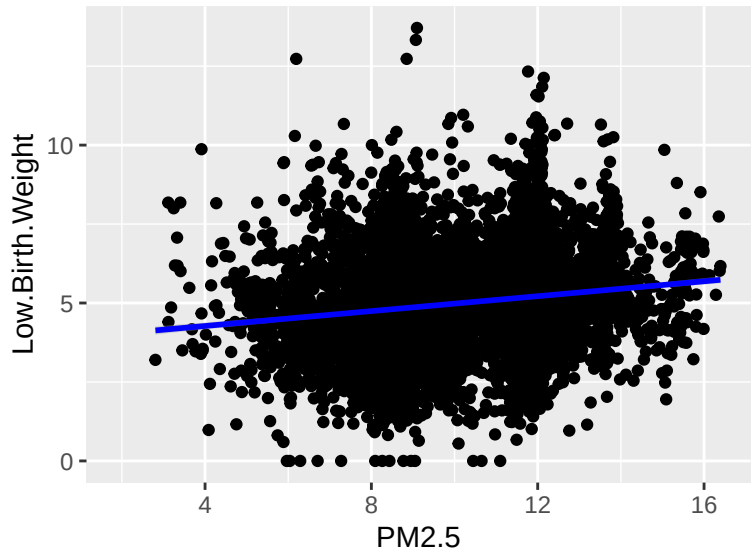
```
## Linear hypothesis test
##
## Hypothesis:
## PM2.5 - Poverty = 0
##
## Model 1: restricted model
## Model 2: Low.Birth.Weight ~ PM2.5 + Poverty
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1   7803
## 2   7802  1 13.468  0.0002426 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

G). Based on the linear hypothesis test above, we can reject the null hypothesis that the effect of PM2.5 is equal to the effect of Poverty in regression F.

Exploratory Visualization & Analysis

```
birth_weight_PM2.5_plot = ggplot(data = data, aes(x = PM2.5,
                                                  y = Low.Birth.Weight)) +
  geom_point()+
  geom_smooth(method  = "lm", col = "blue")

birth_weight_PM2.5_plot
```

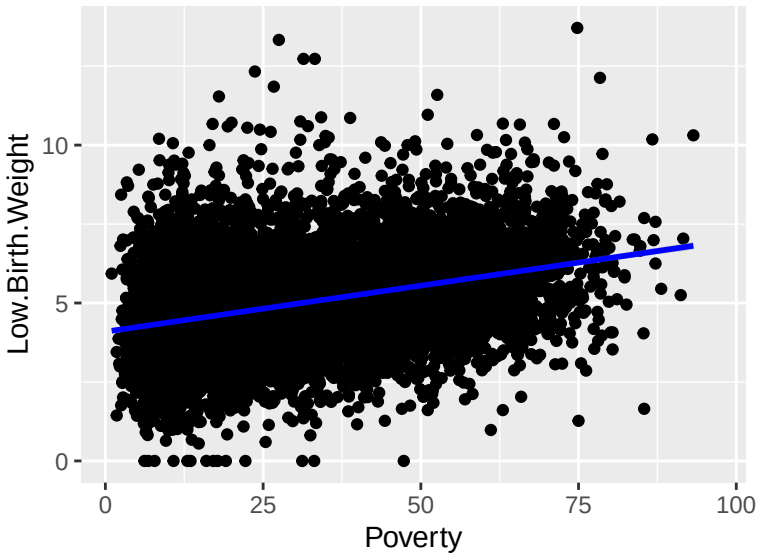



```
model_5 <- lm_robust(Low.Birth.Weight ~ Poverty, data = data)
summary(model_5)
```

```
##
## Call:
## lm_robust(formula = Low.Birth.Weight ~ Poverty, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  4.09145   0.034870  117.34  0.000e+00   4.0231  4.15981 7803
## Poverty      0.02919   0.000964   30.28 1.576e-190   0.0273  0.03108 7803
##
## Multiple R-squared:  0.111 , Adjusted R-squared:  0.1109
## F-statistic: 916.9 on 1 and 7803 DF,  p-value: < 0.00000000000000022
```

```
birth_weight_Poverty_plot = ggplot(data = data, aes(x = Poverty,
                                                    y = Low.Birth.Weight)) +
  geom_point()+
  geom_smooth(method  = "lm", col = "blue")

birth_weight_Poverty_plot
```


Low.Birth.Weight
10
5
0
0
25
50
75
100
Poverty