EDS 241: Assignment 1

Alexandra Yousefivand

01/21/2022

```
# load raw data
data_raw <- read.csv("CES4.csv")

# clean data
data <- data_raw %>%
          janitor::clean_names() %>%
          select(i_census_tract,
                 total_population,
                 california_county,
                 low_birth_weight,
                 pm2_5,
                 poverty)
```

1 (a) What is the average concentration of PM2.5 across all census tracts in California?

```
avg_pm2_5 <- mean(data$pm2_5, na.rm = TRUE)

print(paste((round(avg_pm2_5, 2)), "micrograms per cubic meter"))
```

```
## [1] "10.15 micrograms per cubic meter"
```

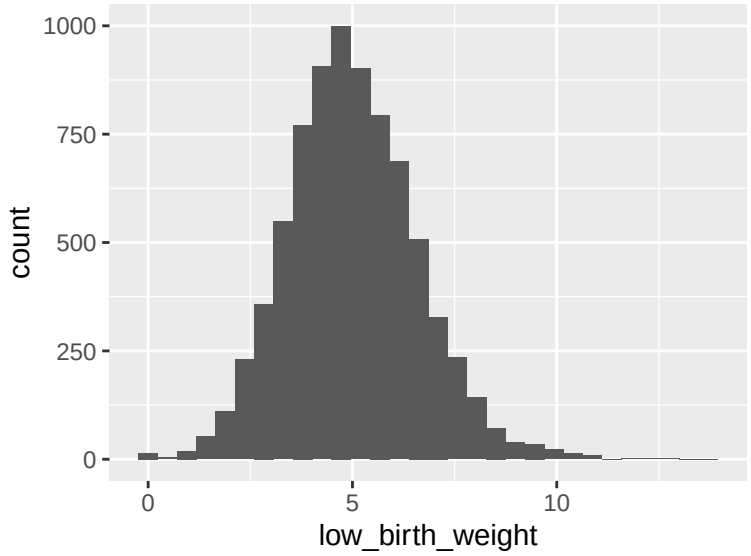
2 (b) What county has the highest level of poverty in California?

```
# using the average of all census tracts within that particular county
high_poverty <- data %>% group_by(california_county) %>%
                         summarize(mean_poverty = mean(poverty, na.rm = TRUE))

county_high_poverty <- high_poverty %>% filter(mean_poverty == max(mean_poverty))

county_high_poverty[[1]]
```

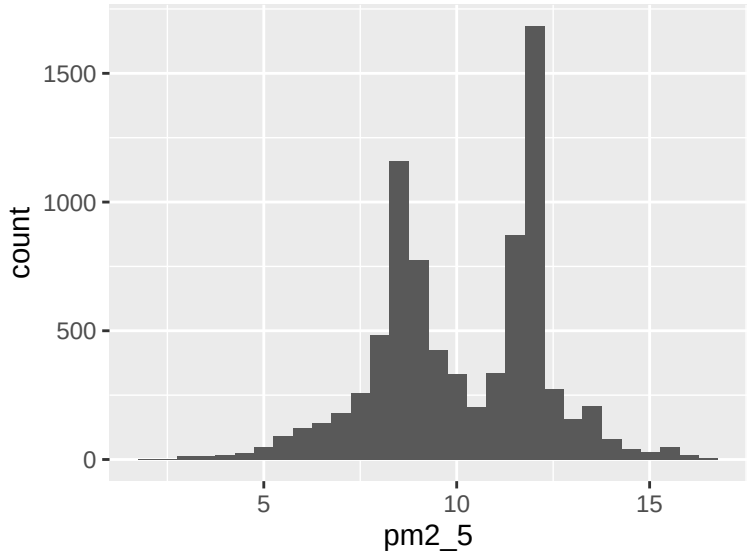
```
## [1] "Tulare "
```

3 (c) Make a histogram depicting the distribution of percent low birth weight and PM2.5.

```
ggplot(data = data, mapping = aes(x = low_birth_weight)) +
  geom_histogram()
```




```
ggplot(data = data, mapping = aes(x = pm2_5)) +
  geom_histogram()
```



4 (d) Estimate a OLS regression of LowBirthWeight on PM25. Report the estimated slope coefficient and its heteroskedasticity-robust standard error. Interpret the estimated slope coefficient. Is the effect of PM25 on LowBirthWeight statistically significant at the 5%?

```
model_robust <- lm_robust(formula = low_birth_weight ~ pm2_5, data = data)
summary(model_robust)
```

```
##
## 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: < 2.2e-16
```

Slope coefficient = 0.1179

Heteroskedasticity-robust standard error = 0.0084024

Slope coefficient interpretation: For every microgram per cubic meter increase in concentration of PM2.5, low birth rate increases by 0.1179%.

Yes; the effect of PM2.5 on LowBirthWeight is statistically significant at the 5% level, since the p-value $3.2561061 \times 10^{-44}$ « 0.05.

5 (f) Add the variable Poverty as an explanatory variable to the regression in (d). Interpret the estimated coefficient on Poverty. What happens to the estimated coefficient on PM25,compared to the regression in (d). Explain.

```
model_robust_poverty <- lm_robust(formula = low_birth_weight ~ pm2_5 + poverty, data = data)
summary(model_robust_poverty)
```

```
##
## 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: < 2.2e-16
```

For every 1% increase in population living below two times the federal poverty level, there is a 0.027% increase in LowBirthWeight.

```
pm2_5 <- as.numeric(model_robust$coefficients[2])
pm2_5_poverty <- as.numeric(model_robust_poverty$coefficients[2])
difference <- pm2_5 - pm2_5_poverty
difference
```

```
## [1] 0.0588228
```

The estimated coefficient on PM2.5 decreased by 0.0588228 because more of the model is explained by the poverty explanatory term (omitted variable bias).

6 (g) From the regression in (f), test the null hypothesis that the effect of PM2.5 is equal to the effect of Poverty.

```
model_hyp_test <- linearHypothesis(model = model_robust_poverty,
                                   "pm2_5 = poverty",
                                   white.adjust = "hc2")

p <- model_hyp_test$`Pr(>Chisq)`[2]
p
```

```
## [1] 0.0002426369
```

Because the p-value of $2.4263693 \times 10^{-4} \ll 0.05$, we reject the null hypothesis and accept the alternative: the effect of PM2.5 is NOT equal to the effect of Poverty.