EDS241: Assignment 1

Jake Eisaguirre

01/19/2022

Read in, inspect data, and wrangle

```
data <- read_excel(here("data","CES4.xlsx"), sheet = 1)

clean_data <- data %>%
  as.data.frame() %>%
  select(`Census Tract`, `Total Population`, `Low Birth Weight`, PM2.5, Poverty, `California County`) %
  na.omit() %>%
  clean_names() %>%
  filter(!str_detect(low_birth_weight, "NA"))
```

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

```
print(mean(clean_data$pm2_5))
```

```
## [1] 10.19529
```

The average concentration of PM2.5 across all census tracts in California is 10.15 micrograms per cubic meters

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

```
pov_table <- clean_data[which(clean_data$poverty == max(clean_data$poverty)), ]
```

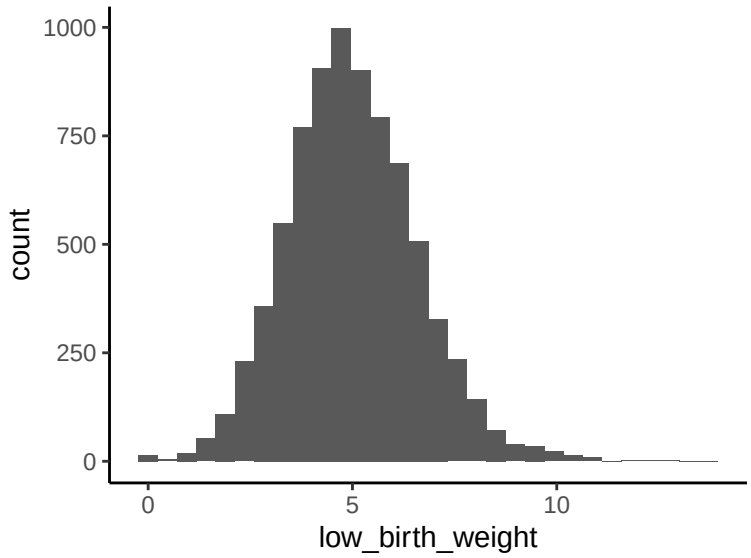
```
kableExtra::kable(pov_table)
```

	census_tract	total_population	low_birth_weight	pm2_5	poverty	california_county
354	6037206300	6103	10.31	12.39953	93.2	Los Angeles

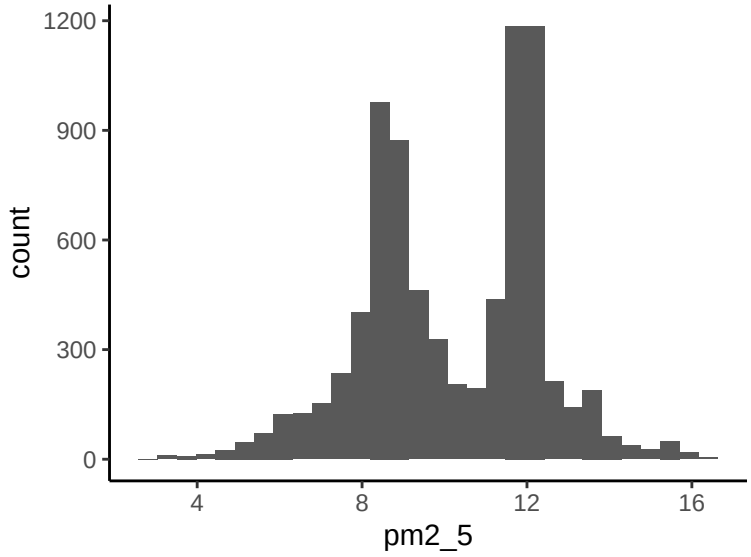
The county that has the highest level of poverty in California is Los Angeles County/ Censes tract 6037206300

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

```
ggplot(data = clean_data, aes(x = as.numeric(low_birth_weight))) +
  geom_histogram() +
  theme_classic() +
  xlab("low_birth_weight")
```




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



(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_1 <- lm_robust(low_birth_weight ~ pm2_5, data = clean_data)
```

```
broom::tidy(model_1) %>%
  kableExtra::kable()
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	3.7995702	0.0885779	42.89522	0	3.6259337	3.9732067	7803	low_birth_weight
pm2_5	0.1181619	0.0084014	14.06456	0	0.1016929	0.1346309	7803	low_birth_weight

The estimated slope coefficient is 0.1182 and its heteroskedasticity-robust standard error is 0.008401.

Slope interpretation: For every one unit increase in PM2.5 we see an 0.1182 percent increase in census tract birth weights less than 2500g.

Yes the effect of PM2.5 on low birth weights is statistically significant at the 5% level since the p-values is extremely small (2.179e-44).

(e) Suppose a new air quality policy is expected to reduce PM2.5 concentration by 2 micrograms per cubic meters. Predict the new average value of LowBirthWeight and derive its 95% confidence interval. Interpret the 95% confidence interval.

```
lbw <- as.numeric(clean_data$low_birth_weight)

new_average <- (0.1182*(mean(lbw)-2) + 3.7996)
print(new_average)
```

```
## [1] 4.154704
```

```
CI_pos <- mean(lbw) + 1.96 * (sd(lbw)/sqrt(7805))
CI_neg <- mean(lbw) - 1.96 * (sd(lbw)/sqrt(7805))
print(CI_pos)
```

```
## [1] 5.039494
```

```
print(CI_neg)
```

```
## [1] 4.969036
```

The new average value of birth weights less than 2500g is 4.154% with a decrease of 2 micrograms per cubic meters of PM2.5. We 95% confident that the population of baby birth weights less than 2500g is between 4.969% and 5.039%.

(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_2 <- lm_robust(low_birth_weight ~ pm2_5 + poverty, data = clean_data)
```

```
broom::tidy(model_2) %>%
  kableExtra::kable()
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	3.5437420	0.0847329	41.82252	0	3.3776428	3.7098411	7802	low_birth_weight
pm2_5	0.0591077	0.0082932	7.12723	0	0.0428508	0.0753647	7802	low_birth_weight
poverty	0.0274353	0.0010022	27.37448	0	0.0254707	0.0293999	7802	low_birth_weight

Poverty coefficient Interpretation: When PM2.5 is held fixed, a one unit increase in poverty will result in a 0.02744 percent increase in census tract birth weights less than 2500g.

The coefficient of PM2.5 in this model compared to the previous model seems to have decreased by 50%. This happens most likely due to the two predictor variables being some what correlated or having co-linearity. The model is unsure what variable is explaining the variance.

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

```
lh <- linearHypothesis(model_2, c("pm2_5=0", "poverty=0"), white.adjust = "hc2")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Wed, Jan 19, 2022 - 09:16:48

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Res.Df	2	7,803.000	1.414	7,802	7,802.5	7,803.5	7,804
Df	1	2.000		2.000	2.000	2.000	2.000
Chisq	1	989.606		989.606	989.606	989.606	989.606
Pr(>Chisq)	1	0.000		0.000	0.000	0.000	0.000

We reject the null hypothesis that the effect of PM2.5 is equal to the effect of Poverty