EDS241: Assignment 2

Peter Menzies

02/01/2022

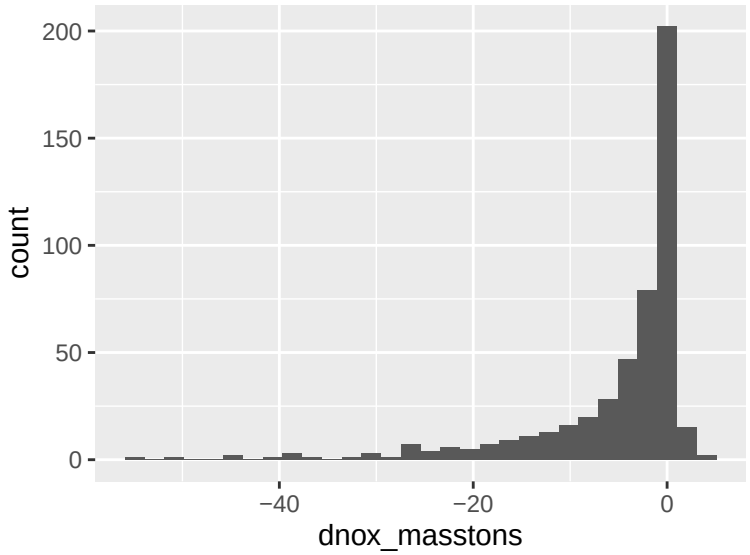
Reading in data

```
df <- read_csv(here("data", "NBP.csv")) %>%
  clean_names()
```

(a) `dnox_masstons` distribution

```
dnox_dist <- ggplot(df, aes(x = dnox_masstons)) +
  geom_histogram()

dnox_dist
```



(b) Creating an indicator (D = 1) representing pct_black above the sample median

```
df_D <- df %>%
  mutate("D" = case_when(
    pct_black > median(pct_black) ~ 1,
    pct_black <= median(pct_black) ~ 0
  ))

df_D1 <- df_D %>%
  filter(D == 1)

avg_pct_D1 <- mean(df_D1$pct_black)
```

The average of pct_black for counties above the median is 19.9090909

(c) Regression of dnox_masstons on nbp

```
mdl_nbp <- lm_robust(dnox_masstons ~ nbp, df)
```

```
mdl_nbp %>%
  tidy %>%
  xtable()
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	-3.62	0.42	-8.59	0.00	-4.44	-2.79	483.00	dnox_masstons
2	nbp	-3.91	0.80	-4.91	0.00	-5.47	-2.34	483.00	dnox_masstons

```
intercept <- mdl_nbp$coefficients[['(Intercept)']]
nbp <- mdl_nbp$coefficients[['nbp']]
```

The intercept tells us that counties *without* NBP in effect saw an estimated decrease of 3.6153846 tons of NOx between 2000 and 2008. The coefficient on nbp tells us that counties *with* NBP in effect saw an additional estimated decrease of 3.9082003 tons of NOx between 2000 and 2008 (that is—in addition to the decrease represented by the intercept).

(d) Linear regression of dnox_masstons on nbp, D, and nbp x D

```
mdl_nbp_D <- lm_robust(dnox_masstons ~ nbp + D + nbp * D, df_D)
```

```
mdl_nbp_D %>%
  tidy %>%
  xtable()
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	-2.60	0.47	-5.55	0.00	-3.52	-1.68	481.00	dnox_masstons
2	nbp	-6.33	1.22	-5.21	0.00	-8.72	-3.94	481.00	dnox_masstons
3	D	-2.21	0.86	-2.58	0.01	-3.90	-0.53	481.00	dnox_masstons
4	nbp:D	5.04	1.59	3.16	0.00	1.91	8.16	481.00	dnox_masstons

```
intercept <- mdl_nbp_D$coefficients[['(Intercept)']]
nbp <- mdl_nbp_D$coefficients[['nbp']]
D <- mdl_nbp_D$coefficients[['D']]
nbp_x_D <- mdl_nbp_D$coefficients[['nbp:D']]
```

The intercept tells us that counties *without* NBP in effect *and* with a `pct_black` less than or equal to the median saw an estimated decrease of 2.6013514 tons of NOx between 2000 and 2008.

The coefficient on `nbp` tells us that a county *with* NBP in effect *and* with a `pct_black` less than or equal to the median saw an estimated decrease of 6.3326109 tons of NOx in the total change in NOx between 2000 and 2008.

The coefficient on `D` tells us that a county *without* NBP in effect *and* a `pct_black` greater than the median saw an estimated decrease of 2.2146486 tons of NOx in the total change in NOx between 2000 and 2008.

The coefficient on the interaction between `nbp` and `D` represents an increase of 5.0354034 tons of NOx in the total change in NOx (in addition to the estimated change associated with the other two coefficients) between 2000 and 2008 in counties *with* NBP in effect *and* `pct_black` greater than the median.

(e) Predicted `dnox_masstons` in a county without NBP in effect and where `pct_black` is above the sample median

```
x_vals <- tribble(~nbp, ~D,
                  0,   1)

pred_dnox <- predict(mdl_nbp_D,
                     newdata = x_vals,
                     se.fit = TRUE,
                     interval = 'confidence')
```

The predicted `dnox_masstons` in a county without NBP in effect and where `pct_black` is above the sample median is -4.816 tons of NOx. The 95% confidence interval for this prediction ranges from -6.2299914 to -3.4020086.