EDS241: Assignment 2

Peter Menzies

02/06/2022

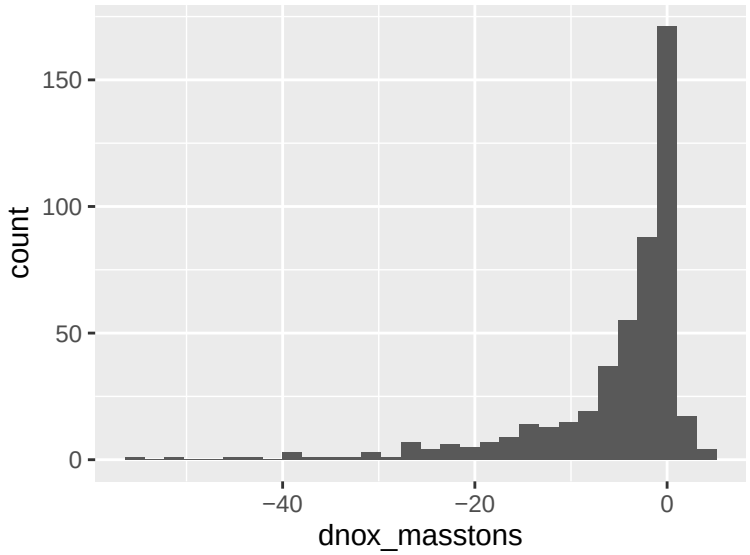
Reading in data

```
df <- read_excel(here("data", "NBP.xls")) %>%
  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 pct_black for counties above the median is 19.31375

(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	-8.62	0.00	-4.45	-2.80	483.00	dnox_masstons
2	nbp	-3.92	0.80	-4.93	0.00	-5.48	-2.36	483.00	dnox_masstons

The intercept represents the estimated decrease in tons of NOx between 2000 and 2008 in counties without nbp in effect.

The coefficient on nbp represents the estimated difference in change in tons of NOx between counties with nbp in effect and without nbp in effect between 2000 and 2008.

(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()
```

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 2.4180753 tons of NOx between 2000 and 2008.

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	-2.42	0.44	-5.47	0.00	-3.29	-1.55	481.00	dnox_masstons
2	nbp	-7.14	1.26	-5.68	0.00	-9.61	-4.67	481.00	dnox_masstons
3	D	-2.59	0.85	-3.03	0.00	-4.26	-0.91	481.00	dnox_masstons
4	nbp:D	6.37	1.61	3.95	0.00	3.20	9.54	481.00	dnox_masstons

The coefficient on `nbp` represents the estimated difference in change in tons of NOx in counties with `nbp` in effect and without `nbp` in effect that both have `pct_black` lower than the median.

The coefficient on `D` represents the estimated difference in change in tons of NOx in counties with `pct_black` higher than median and `pct_black` lower than median both without `nbp` in effect.

The coefficient on the interaction between `nbp` and `D` represents the estimated difference in change in tons of NOx associated with implementation of `nbp` in counties with `pct_black` higher than median and counties with `pct_black` lower than 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 -5.006106 tons of NOx. The 95% confidence interval for this prediction ranges from -6.4400648 to -3.5721472.