EDS241: Take Home Final

Peter Menzies

03/12/2022

```
#simple function to make formatted table
tablr <- function(obj) {

    obj %>%
    tidy() %>%
    xtable()

}
```

Reading in data

```
df <- read_csv(here("data", "KM_EDS241.csv")) %>%
  clean_names() %>%
  mutate(nearinc = as.factor(nearinc))
```

(a) OLS regression of real house values on the indicator for being located near the incinerator in 1981.

```
df_81 <- df %>%
  filter(year == "1981")
```

```
ols_a <- lm_robust(rprice ~ nearinc, df_81)
```

```
tablr(ols_a)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	101307.51	2944.81	34.40	0.00	95485.47	107129.56	140.00	rprice
2	nearinc1	-30688.27	6243.17	-4.92	0.00	-43031.35	-18345.20	140.00	rprice

The estimated "penalty" in value for houses near the incinerator based on the previous OLS regression is -3.0688274×10^4..

This estimate does *not* correspond to the causal effect of being near the incinerator on housing values. The regression does not include the other observed determinants of housing value included in the dataset (`age`, `rooms`, `area`, `land`—which are in fact significantly correlated with both `rprice` and `nearinc`), nor does it control for unobserved determinants of housing value—thus the estimator is subject to omitted variable bias and we cannot infer causality.

(b) Provide evidence that the location choice of the incinerator was not "random", but rather selected on the basis of house values and characteristics.

```
df_78 <- df %>%
  filter(year == 1978)
```

```
ols_b1 <- lm_robust(rprice ~ nearinc, df_78)
ols_b2 <- lm_robust(age ~ nearinc, df_78)
ols_b3 <- lm_robust(rooms ~ nearinc, df_78)
```

```
# rprice ~ nearinc
tablr(ols_b1)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	82517.23	1878.28	43.93	0.00	78810.53	86223.93	177.00	rprice
2	nearinc1	-18824.37	6010.01	-3.13	0.00	-30684.88	-6963.86	177.00	rprice

```
# age ~ nearinc
tablr(ols_b2)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	12.75	3.23	3.95	0.00	6.38	19.12	177.00	age
2	nearinc1	27.04	5.76	4.69	0.00	15.67	38.40	177.00	age

```
# rooms ~ nearinc
tablr(ols_b3)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	6.83	0.07	95.08	0.00	6.69	6.97	177.00	rooms
2	nearinc1	-0.79	0.16	-4.99	0.00	-1.11	-0.48	177.00	rooms