HW6

12351235

2025-02-19

```r
#HW6

library(estimatr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
#Question 1

load("mroz.RData")

mroz <- data

#a
lpm_model <- lm_robust(inlf ~ educ + exper + expersq, data = mroz)
summary(lpm_model)
```

```
##
## Call:
## lm_robust(formula = inlf ~ educ + exper + expersq, data = mroz)
##
## Standard error type:  HC2
##
## Coefficients:
##               Estimate Std. Error t value  Pr(>|t|)  CI Lower   CI Upper  DF
## (Intercept) -0.1616014  0.0870177  -1.857 6.369e-02 -0.332429  0.0092263 749
## educ         0.0321910  0.0070564   4.562 5.919e-06  0.018338  0.0460436 749
## exper        0.0465279  0.0056923   8.174 1.273e-15  0.035353  0.0577027 749
## expersq     -0.0008996  0.0001805  -4.985 7.717e-07 -0.001254 -0.0005453 749
##
## Multiple R-squared:  0.169 , Adjusted R-squared:  0.1657
## F-statistic: 63.88 on 3 and 749 DF,  p-value: < 2.2e-16
```

```r
#delta_p_inlf = -0.16 + .032 * delta_educ

#b
mroz$inflhat_lpm <- fitted(lpm_model)
summary(mroz$inflhat_lpm)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.0006465  0.4205597  0.5954567  0.5683931  0.7315301  0.9841544
```

The range is -.00065 to .98415. This is a problem because the value going below 0 is nonsensical.

```r
#c
probit_model <- glm(inlf ~ educ + exper + expersq, mroz, family = binomial(link = "probit"))
summary(probit_model)
```

```
##
## Call:
## glm(formula = inlf ~ educ + exper + expersq, family = binomial(link = "probit"),
##     data = mroz)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.9254937  0.2886193  -6.671 2.53e-11 ***
## educ         0.0971238  0.0221693   4.381 1.18e-05 ***
## exper        0.1271342  0.0177861   7.148 8.81e-13 ***
## expersq     -0.0023927  0.0005763  -4.152 3.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1029.75  on 752  degrees of freedom
## Residual deviance:  894.33  on 749  degrees of freedom
## AIC: 902.33
##
## Number of Fisher Scoring iterations: 4
```

```r
mroz$inflhat_probit <- fitted(probit_model)
summary(mroz$inflhat_probit)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.07495 0.41844 0.59941 0.56871 0.74410 0.92065
```

The range is .075 to .921, which is not a problem because it is between 0 and 1.

#d

They are roughly the same statistical significance but their signs are different. This is what I expected because they are measuring the same effect but the magnitude of the probit model coefficient is not easily interpretable, while the coefficient for the LPM model is.

```r
#e
beta_educ <- coef(probit_model)["educ"]
linear_preds <- probit_model$linear.predictors

marginal_effects_educ <- dnorm(linear_preds) * beta_educ
APE_educ <- mean(marginal_effects_educ)
APE_educ
```

```
## [1] 0.03278275
```

```r
#f
betas <- coef(probit_model)
X <- model.matrix(probit_model)[,-1]
xbar <- colMeans(X)

linear_pred_avg <- betas[1] + sum(betas[-1] * xbar)

density_at_avg <- dnorm(linear_pred_avg)

PEA_educ <- density_at_avg * betas["educ"]
PEA_educ
```

```
## (Intercept)
##  0.03802895
```

#g

They are all roughly same except for the coefficient in the probit regression. This is what I would expect because LPM, APE, and PEA all are measuring very similar effects, while the magnitude of the coefficient in the probit model is not really interpretable.

```r
#Question 3

load("kielmc.RData")

kielmc <- data

#a
only_1981 <- kielmc |>
  filter(year == 1981)

simple_model_81 <- lm_robust(lrprice ~ nearinc, only_1981)
summary(simple_model_81)
```

```
##
## Call:
## lm_robust(formula = lrprice ~ nearinc, data = only_1981)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper  DF
```

```
## (Intercept)  11.4785    0.03170 362.102 6.145e-210  11.4158   11.541 140
## nearinc      -0.4026    0.07159  -5.623  9.829e-08  -0.5441   -0.261 140
##
## Multiple R-squared:  0.2172 ,    Adjusted R-squared:  0.2116
## F-statistic: 31.62 on 1 and 140 DF,  p-value: 9.829e-08
```

The results show that the incinerator had an adverse impact on home prices. However, this model does not contain any control variable to make causality a more reasonable assumption.

```
#b
model_w_controls_81 <- lm_robust(lprice ~ nearinc + age + agesq + rooms + baths + lintst + larea + llan
summary(model_w_controls_81)
```

```
##
## Call:
## lm_robust(formula = lprice ~ nearinc + age + agesq + rooms +
##     baths + lintst + larea + lland, data = only_1981)
##
## Standard error type:  HC2
##
## Coefficients:
##               Estimate Std. Error t value  Pr(>|t|)   CI Lower   CI Upper  DF
## (Intercept)  8.674e+00  0.7347586 11.8054 1.906e-22  7.221e+00  1.013e+01 133
## nearinc     -1.364e-01  0.0818480 -1.6665 9.797e-02 -2.983e-01  2.549e-02 133
## age         -8.317e-03  0.0026353 -3.1562 1.978e-03 -1.353e-02 -3.105e-03 133
## agesq        4.145e-05  0.0000227  1.8265 7.002e-02 -3.438e-06  8.634e-05 133
## rooms        1.760e-02  0.0259984  0.6771 4.995e-01 -3.382e-02  6.903e-02 133
## baths        1.315e-01  0.0470726  2.7943 5.971e-03  3.843e-02  2.246e-01 133
## lintst      -8.171e-02  0.0614364 -1.3300 1.858e-01 -2.032e-01  3.981e-02 133
## larea        3.385e-01  0.0955037  3.5446 5.434e-04  1.496e-01  5.274e-01 133
## lland        8.070e-02  0.0401501  2.0099 4.647e-02  1.282e-03  1.601e-01 133
##
## Multiple R-squared:  0.7619 ,    Adjusted R-squared:  0.7475
## F-statistic: 74.72 on 8 and 133 DF,  p-value: < 2.2e-16
```

The coefficient on nearinc is smaller and the standard error is larger. This goes a long way towards resolving the issues found in a) but we still can't be sure there aren't other variables we need to control for to ensure causality.

```
#c
only_1978 <- kielmc |>
  filter(year == 1978)

simple_model_78 <- lm_robust(lrprice ~ nearinc, only_1978)
summary(simple_model_78)
```

```
##
## Call:
## lm_robust(formula = lrprice ~ nearinc, data = only_1978)
##
## Standard error type:  HC2
##
```

```
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)  11.2854    0.02510 449.679 1.336e-272  11.2359  11.3350 177
## nearinc      -0.3399    0.06245  -5.443  1.729e-07  -0.4632  -0.2167 177
##
## Multiple R-squared:  0.1855 ,    Adjusted R-squared:  0.1809
## F-statistic: 29.62 on 1 and 177 DF,  p-value: 1.729e-07
```

The coefficient for nearinc is almost as large as it was when using the 1981 data. This tells us that there is something else about the areas where the incincerators were placed that is hurting home prices and that is what our analysis in a) is picking up because the incinerators weren't even built in 1978.

```
#d
kielmc_81ind <- kielmc |>
  mutate(y81 = ifelse(year == 1981, 1, 0))

pooled_model <- lm_robust(lrprice ~ nearinc + y81 + y81 * nearinc, kielmc_81ind)
summary(pooled_model)
```

```
##
## Call:
## lm_robust(formula = lrprice ~ nearinc + y81 + y81 * nearinc,
##     data = kielmc_81ind)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error  t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept) 11.28542    0.02510 449.6786 0.000e+00  11.2360  11.3348 317
## nearinc     -0.33992    0.06245  -5.4428 1.052e-07  -0.4628  -0.2170 317
## y81          0.19309    0.04043   4.7758 2.740e-06   0.1135   0.2726 317
## nearinc:y81 -0.06265    0.09501  -0.6594 5.101e-01  -0.2496   0.1243 317
##
## Multiple R-squared:  0.246 , Adjusted R-squared:  0.2388
## F-statistic: 27.55 on 3 and 317 DF,  p-value: 7.423e-16
```

This model tells us that for home prices in 1981, being near the incinerator on average reduced the price by $(.4628 + .2496 = .7124)\%$. But, about 2/3 of the effect on home prices comes regardless of whether they were measured before or after the incinerator was built. So, while the incinerator does have some impact on housing prices, there is something else area having a larger impact.

```
#e
pooled_model_control <- lm_robust(lprice ~ nearinc + y81 + y81 * nearinc + age + agesq + rooms + baths +
summary(pooled_model_control)
```

```
##
## Call:
## lm_robust(formula = lprice ~ nearinc + y81 + y81 * nearinc +
##     age + agesq + rooms + baths + lintst + larea + lland, data = kielmc_81ind)
##
## Standard error type:  HC2
##
```

```
## Coefficients:
##                Estimate Std. Error t value  Pr(>|t|)   CI Lower  CI Upper  DF
## (Intercept)  7.652e+00  5.375e-01  14.235 9.948e-36  6.594e+00  8.709460 310
## nearinc      3.223e-02  6.485e-02   0.497 6.195e-01 -9.537e-02  0.159829 310
## y81          4.260e-01  2.720e-02  15.660 3.969e-41  3.725e-01  0.479497 310
## age         -8.359e-03  1.657e-03  -5.045 7.721e-07 -1.162e-02 -0.005099 310
## agesq        3.763e-05  1.137e-05   3.310 1.042e-03  1.526e-05  0.000060 310
## rooms        4.733e-02  1.776e-02   2.665 8.095e-03  1.239e-02  0.082279 310
## baths        9.428e-02  2.846e-02   3.312 1.036e-03  3.827e-02  0.150285 310
## lintst      -6.145e-02  3.569e-02  -1.722 8.616e-02 -1.317e-01  0.008786 310
## larea        3.508e-01  6.295e-02   5.572 5.461e-08  2.269e-01  0.474632 310
## lland        9.985e-02  3.410e-02   2.928 3.663e-03  3.275e-02  0.166939 310
## nearinc:y81 -1.315e-01  6.082e-02  -2.162 3.135e-02 -2.512e-01 -0.011844 310
##
## Multiple R-squared:  0.7904 ,    Adjusted R-squared:  0.7837
## F-statistic:   135 on 10 and 310 DF,  p-value: < 2.2e-16
```

#Yes, it makes sense to include these control variables because they are all plausibly have meaningful impacts on home price and could vary depending on whether the home is close to the incinerator or not.

```
#f
nearinc_data <- kielmc_81ind |>
  filter(nearinc == 1)

summary(nearinc_data)
```

```
##       year           age             agesq              nbh
##  Min.   :1978   Min.   :  0.00   Min.   :    0.0   Min.   :0.000
##  1st Qu.:1978   1st Qu.: 11.75   1st Qu.:  138.2   1st Qu.:0.750
##  Median :1978   Median : 25.50   Median :  650.5   Median :4.000
##  Mean   :1979   Mean   : 34.85   Mean   : 2347.3   Mean   :2.625
##  3rd Qu.:1981   3rd Qu.: 51.75   3rd Qu.: 2679.8   3rd Qu.:4.000
##  Max.   :1981   Max.   :189.00   Max.   :35721.0   Max.   :4.000
##       cbd            intst           lintst          price
##  Min.   : 1000   Min.   : 1000   Min.   :6.908   Min.   : 31000
##  1st Qu.: 3000   1st Qu.: 4000   1st Qu.:8.294   1st Qu.: 48000
##  Median : 4000   Median : 6000   Median :8.700   Median : 62000
##  Mean   : 5458   Mean   : 6521   Mean   :8.567   Mean   : 75465
##  3rd Qu.: 6250   3rd Qu.: 7000   3rd Qu.:8.854   3rd Qu.: 83750
##  Max.   :14000   Max.   :17000   Max.   :9.741   Max.   :300000
##      rooms            area           land            baths
##  Min.   :4.000   Min.   : 750   Min.   :  1710   Min.   :1.000
##  1st Qu.:5.000   1st Qu.:1394   1st Qu.:  8253   1st Qu.:1.000
##  Median :6.000   Median :1651   Median : 12500   Median :2.000
##  Mean   :6.083   Mean   :1888   Mean   : 22392   Mean   :1.844
##  3rd Qu.:7.000   3rd Qu.:2228   3rd Qu.: 21290   3rd Qu.:2.000
##  Max.   :9.000   Max.   :5136   Max.   :282704   Max.   :4.000
##       dist           ldist            wind            lprice
##  Min.   : 5000   Min.   :8.517   Min.   : 3.000   Min.   :10.34
##  1st Qu.: 8500   1st Qu.:9.048   1st Qu.: 3.000   1st Qu.:10.78
##  Median :10850   Median :9.292   Median : 5.000   Median :11.03
##  Mean   :10539   Mean   :9.223   Mean   : 5.177   Mean   :11.11
##  3rd Qu.:12325   3rd Qu.:9.419   3rd Qu.: 5.000   3rd Qu.:11.34
```

```
##  Max.   :15600   Max.   :9.655   Max.   :11.000   Max.   :12.61
##       y81             larea           lland           y81ldist
##  Min.   :0.0000   Min.   :6.620   Min.   : 7.444   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:7.240   1st Qu.: 9.018   1st Qu.:0.000
##  Median :0.0000   Median :7.409   Median : 9.433   Median :0.000
##  Mean   :0.4167   Mean   :7.461   Mean   : 9.552   Mean   :3.839
##  3rd Qu.:1.0000   3rd Qu.:7.709   3rd Qu.: 9.966   3rd Qu.:9.205
##  Max.   :1.0000   Max.   :8.544   Max.   :12.552   Max.   :9.616
##     lintstsq        nearinc     y81nrinc          rprice          lrprice
##  Min.   :47.72   Min.   :1   Min.   :0.0000   Min.   : 31000   Min.   :10.34
##  1st Qu.:68.79   1st Qu.:1   1st Qu.:0.0000   1st Qu.: 47464   1st Qu.:10.77
##  Median :75.68   Median :1   Median :0.0000   Median : 54500   Median :10.91
##  Mean   :73.88   Mean   :1   Mean   :0.4167   Mean   : 66579   Mean   :11.00
##  3rd Qu.:78.39   3rd Qu.:1   3rd Qu.:1.0000   3rd Qu.: 70675   3rd Qu.:11.17
##  Max.   :94.89   Max.   :1   Max.   :1.0000   Max.   :300000   Max.   :12.61
```

```
farinc_data <- kielmc_81ind |>
  filter(nearinc == 0)

summary(farinc_data)
```

```
##       year           age              agesq              nbh
##  Min.   :1978   Min.   :  0.00   Min.   :    0.0   Min.   :0.000
##  1st Qu.:1978   1st Qu.:  0.00   1st Qu.:    0.0   1st Qu.:0.000
##  Median :1978   Median :  1.00   Median :    1.0   Median :2.000
##  Mean   :1979   Mean   : 10.82   Mean   :  969.5   Mean   :2.031
##  3rd Qu.:1981   3rd Qu.:  8.00   3rd Qu.:   64.0   3rd Qu.:5.000
##  Max.   :1981   Max.   :188.00   Max.   :35344.0   Max.   :6.000
##       cbd            intst           lintst           price
##  Min.   : 9000   Min.   : 7000   Min.   : 8.854   Min.   : 26000
##  1st Qu.:14000   1st Qu.:14000   1st Qu.: 9.547   1st Qu.: 76900
##  Median :21000   Median :22000   Median : 9.999   Median : 94376
##  Mean   :20244   Mean   :20676   Mean   : 9.870   Mean   :104905
##  3rd Qu.:24000   3rd Qu.:26000   3rd Qu.:10.166   3rd Qu.:129900
##  Max.   :35000   Max.   :34000   Max.   :10.434   Max.   :234552
##      rooms            area           land            baths            dist
##  Min.   : 4.0   Min.   : 735   Min.   :  7500   Min.   :1.000   Min.   :16000
##  1st Qu.: 6.0   1st Qu.:1836   1st Qu.: 37026   1st Qu.:2.000   1st Qu.:19000
##  Median : 7.0   Median :2240   Median : 44001   Median :3.000   Median :25500
##  Mean   : 6.8   Mean   :2200   Mean   : 46985   Mean   :2.551   Mean   :25058
##  3rd Qu.: 7.0   3rd Qu.:2580   3rd Qu.: 47480   3rd Qu.:3.000   3rd Qu.:28600
##  Max.   :10.0   Max.   :4056   Max.   :544500   Max.   :4.000   Max.   :40000
##      ldist             wind            lprice           y81
##  Min.   : 9.680   Min.   : 3.000   Min.   :10.17   Min.   :0.0000
##  1st Qu.: 9.852   1st Qu.: 7.000   1st Qu.:11.25   1st Qu.:0.0000
##  Median :10.146   Median : 7.000   Median :11.46   Median :0.0000
##  Mean   :10.100   Mean   : 7.747   Mean   :11.49   Mean   :0.4533
##  3rd Qu.:10.261   3rd Qu.:11.000   3rd Qu.:11.77   3rd Qu.:1.0000
##  Max.   :10.597   Max.   :11.000   Max.   :12.37   Max.   :1.0000
##      larea           lland           y81ldist         lintstsq         nearinc
##  Min.   :6.600   Min.   : 8.923   Min.   : 0.000   Min.   : 78.39   Min.   :0
##  1st Qu.:7.515   1st Qu.:10.519   1st Qu.: 0.000   1st Qu.: 91.14   1st Qu.:0
##  Median :7.714   Median :10.692   Median : 0.000   Median : 99.98   Median :0
##  Mean   :7.655   Mean   :10.622   Mean   : 4.557   Mean   : 97.57   Mean   :0
```

```
##  3rd Qu.:7.856   3rd Qu.:10.768   3rd Qu.:10.043   3rd Qu.:103.35   3rd Qu.:0
##  Max.   :8.308   Max.   :13.208   Max.   :10.569   Max.   :108.87   Max.   :0
##     y81nrinc     rprice          lrprice
##  Min.   :0   Min.   : 26000   Min.   :10.17
##  1st Qu.:0   1st Qu.: 73000   1st Qu.:11.20
##  Median :0   Median : 89500   Median :11.40
##  Mean   :0   Mean   : 91035   Mean   :11.37
##  3rd Qu.:0   3rd Qu.:107527   3rd Qu.:11.59
##  Max.   :0   Max.   :180147   Max.   :12.10
```

This person does have a point as the houses near the incinerator appear to have very different characteristics than the houses away form the incinerator. To provide an unbiased estimate we'd need to know that the prices of the homes near and not near to the incinerator would have had a similar trend. Data on home prices in the years prior to 1978 would tell us if our assumption is reasonable.

#g

This would reduce the internal validity of the research design because the closing of the schools would also negatively effect the price of homes in the area, so we'd be unable to determine if the school closing or the incinerator would be responsible for the change in price.

Question 2

a)

The coefficient estimates are consistent and unbiased, so they get closer to the value of the true parameter as sample size increases. The standard errors shrink as the sample size increases.

b)

See handwritten answer

c)

The estimate would be less precise because we would have less variation in income to inform our estimate.

d)

It would improve the precision of the estimates because education is likely to have a strong relationship with whether or not someone voted.

HW6

2. b) $se(\hat{B}_j) = \frac{\sqrt{(\Sigma \hat{u}_i^2)/(n-k-1)}}{\sqrt{SST_j(1-R_j^2)}}$