Homework Three Submission 1

Safia Read

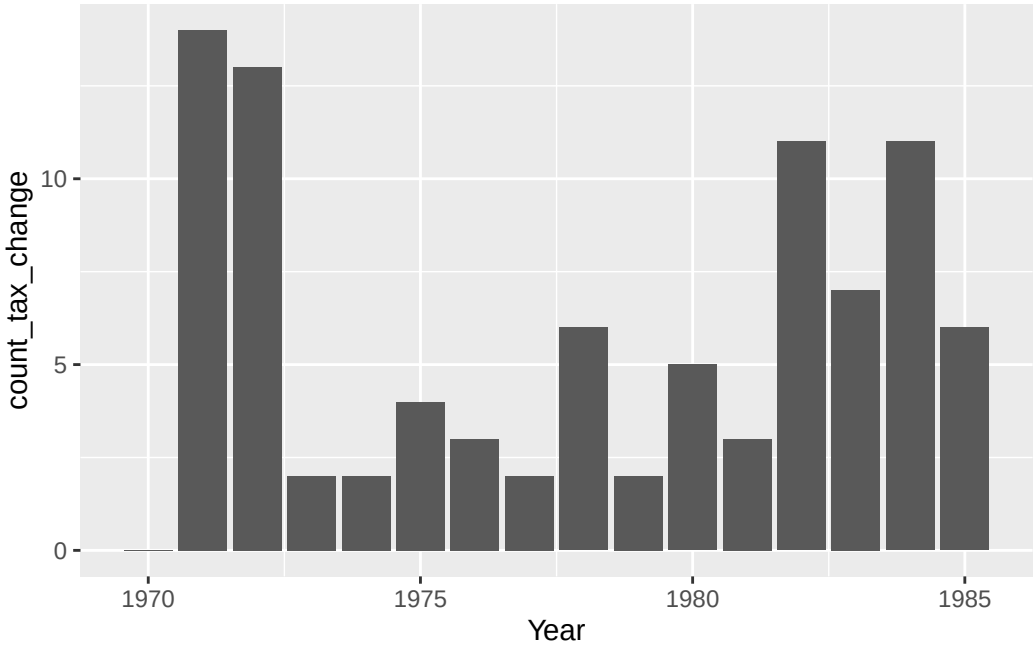
This is my first submission of the third homework for Econ 470.

Link to Github

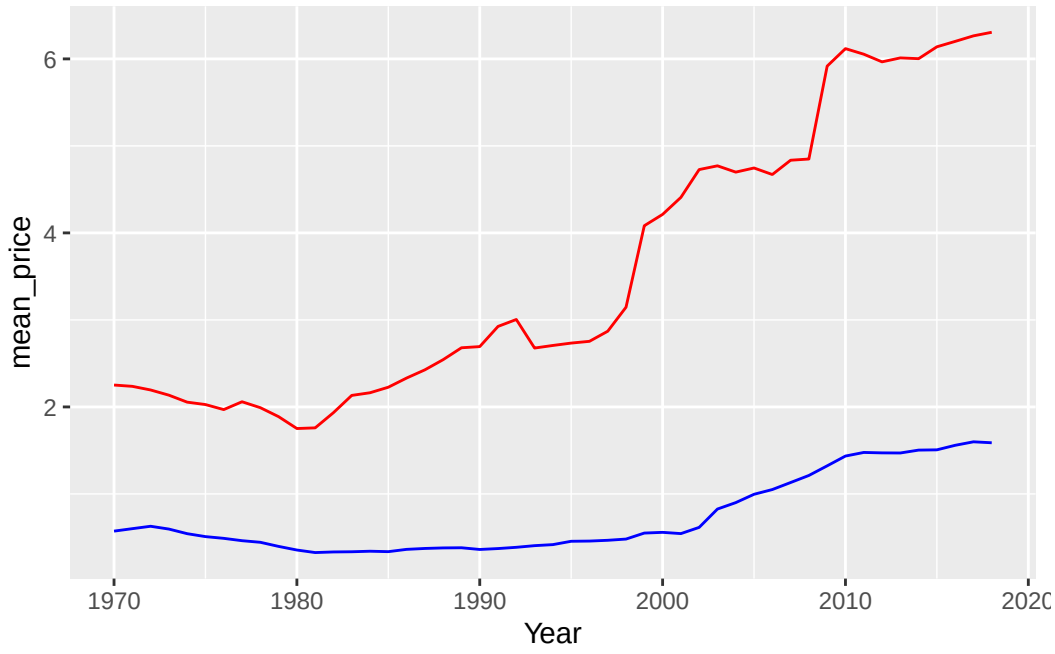
https://github.com/safiaread/homework-3

Summarize the data

1. Present a bar graph showing the proportion of states with a change in their cigarette tax in each year from 1970 to 1985.

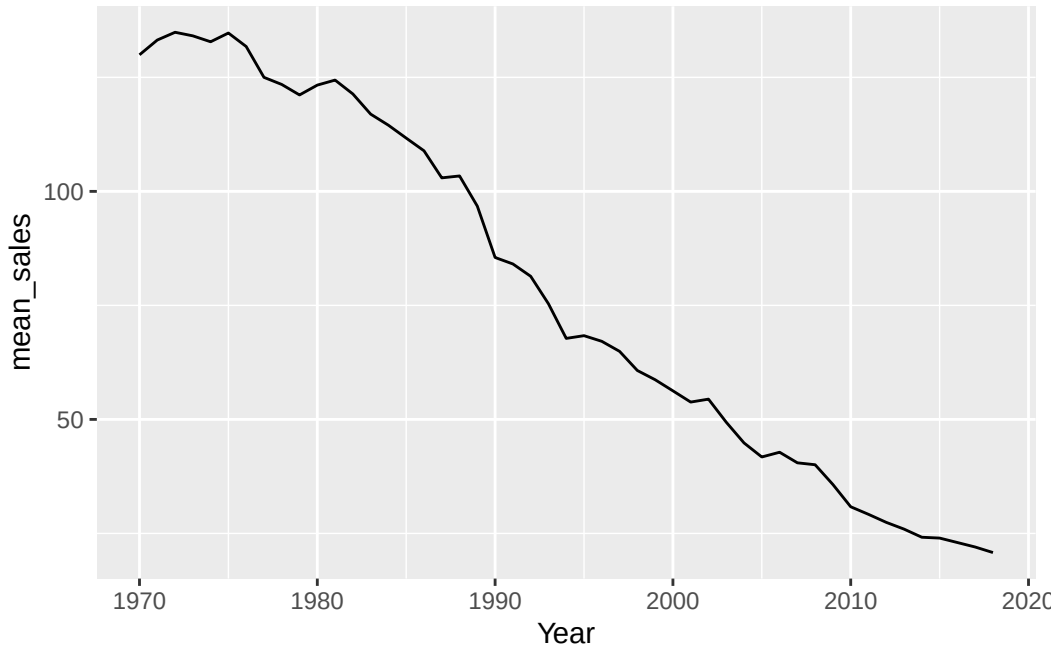


2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.



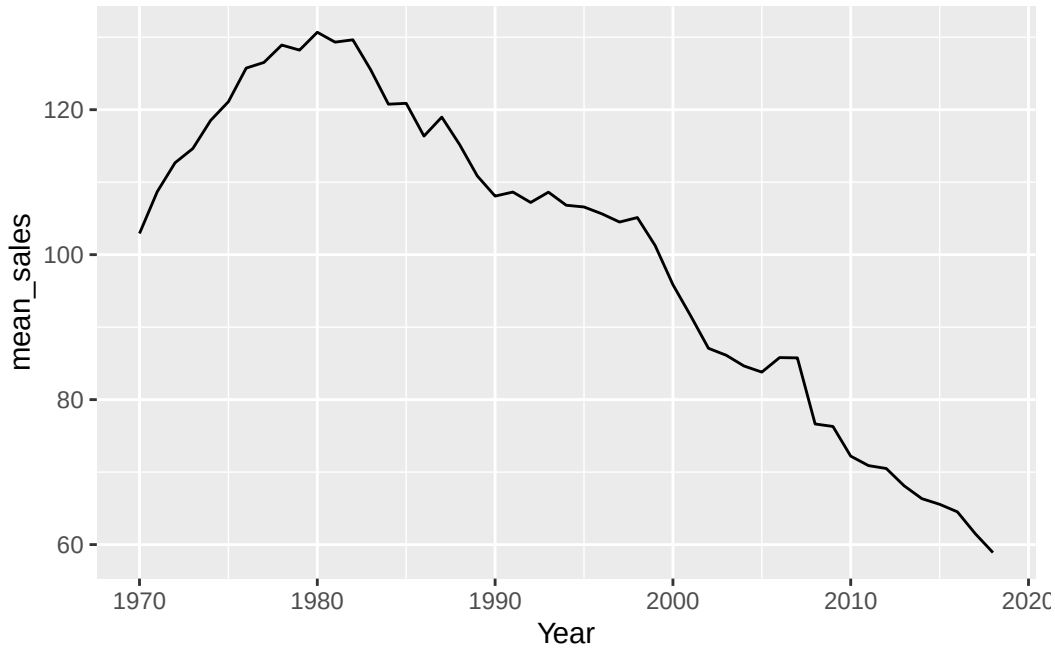
3. Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

```
# A tibble: 5 x 2
# Groups:   state [5]
  state                price_change
  <chr>                       <dbl>
1 District of Columbia         7.09
2 New York                     6.99
3 Rhode Island                 6.40
4 Hawaii                       6.35
5 Massachusetts                6.35
```



4. Identify the 5 states with the lowest increases in cigarette prices over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

```
# A tibble: 5 x 2
# Groups:   state [5]
  state        price_change
  <chr>               <dbl>
1 Missouri             2.36
2 Tennessee            2.37
3 North Dakota         2.49
4 Alabama              2.55
5 Georgia              2.66
```



5. Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.

In both groups, average packs sold decreased. However in the states with the lowest price increases, there are still more sales per capita on average.

Estimate ATEs

Now let's work on estimating a demand curve for cigarettes. Specifically, we're going to estimate the price elasticity of demand for cigarettes. When explaining your findings, try to limit your discussion just to a couple of sentences.

6. Focusing only on the time period from 1970 to 1990, regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

The regression coefficient is -0.8094, which suggests sales and price have an inverse relationship. As price increases we can expect sales to decrease.

```
Call:
lm(formula = ln_sales ~ ln_price, data = q6)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-0.68335 -0.08598 -0.00284  0.08778  0.83516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.42738    0.02975   182.4   <2e-16 ***
ln_price    -0.80944    0.03837   -21.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1894 on 1069 degrees of freedom
Multiple R-squared:  0.294, Adjusted R-squared:  0.2933
F-statistic: 445.1 on 1 and 1069 DF,  p-value: < 2.2e-16
```

7. Again limiting to 1970 to 1990, regress log sales on log prices using the total (federal and state) cigarette tax (in dollars) as an instrument for log prices. Interpret your results and compare your estimates to those without an instrument. Are they different? If so, why?

Yes, the effect is bigger than without the instrument (-1.06252 compared to -0.80944). This is because using tax as an instrument for price can help reduce the effect of confounders, better revealing the true effect.

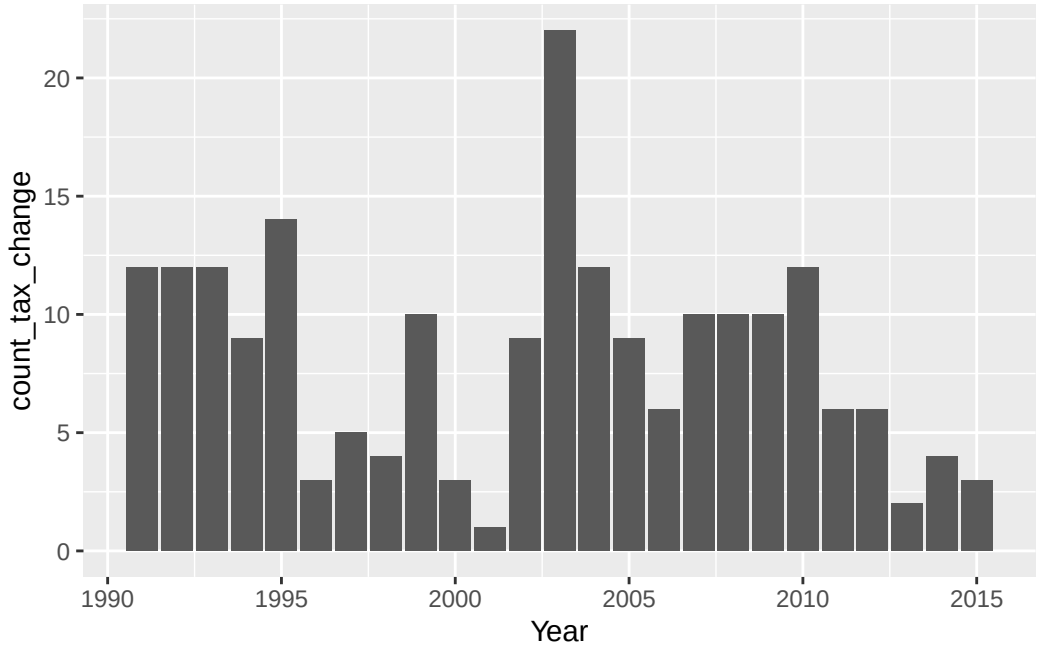
```
TSLS estimation, Dep. Var.: ln_sales, Endo.: ln_price, Instr.: ln_tax_2012
Second stage: Dep. Var.: ln_sales
Observations: 1,071
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   5.61989   0.064208  87.5262 < 2.2e-16 ***
fit_ln_price -1.06252   0.084053 -12.6411 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.193027   Adj. R2: 0.264556
F-test (1st stage), ln_price: stat = 296.0, p < 2.2e-16 , on 1 and 1,069 DoF.
                  Wu-Hausman: stat =  12.2, p = 5.046e-4, on 1 and 1,068 DoF.
```

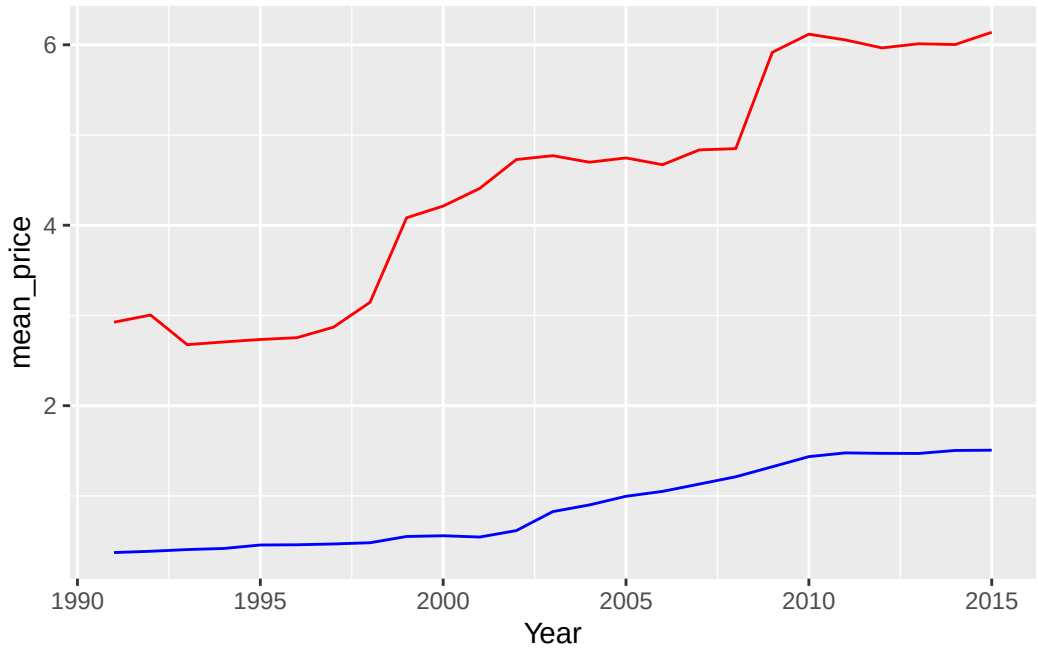
8. Show the first stage and reduced-form results from the instrument.

```
Call:
lm(formula = ln_sales ~ pricehat, data = q6)

Residuals:
     Min       1Q   Median       3Q      Max
```

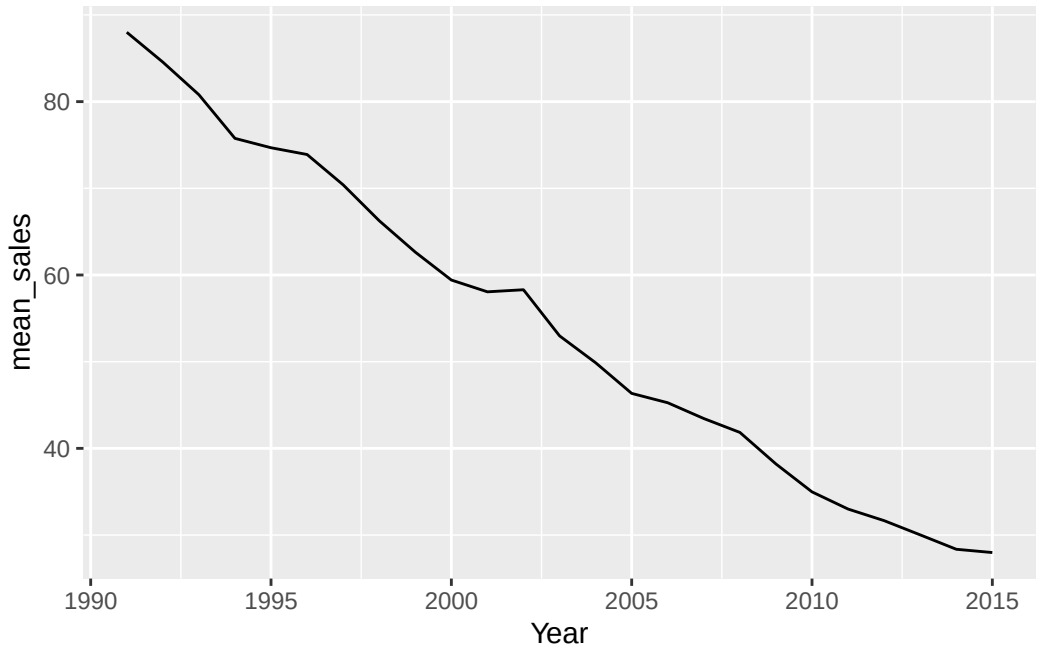
```
-0.86712 -0.10042  0.01055  0.09232  0.93952

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.61989    0.07067   79.52   <2e-16 ***
pricehat    -1.06252    0.09252  -11.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2127 on 1069 degrees of freedom
Multiple R-squared:  0.1098,    Adjusted R-squared:  0.109
F-statistic: 131.9 on 1 and 1069 DF,  p-value: < 2.2e-16
```

9. Repeat questions 1-3 focusing on the period from 1991 to 2015.

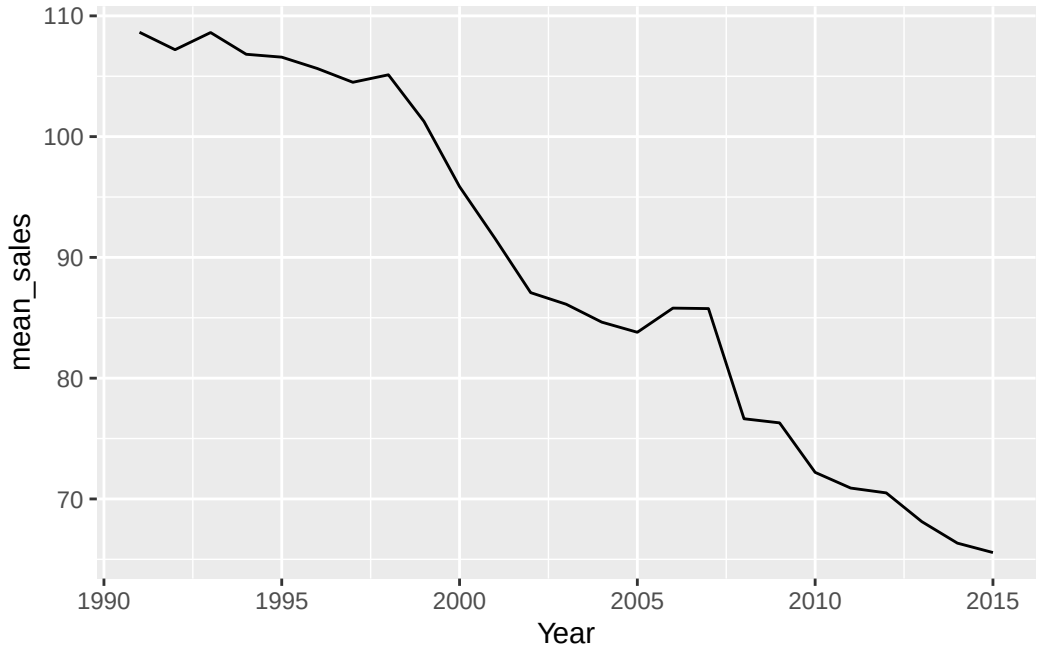





```
# A tibble: 5 x 2
# Groups:   state [5]
  state         price_change
  <chr>                <dbl>
1 New York              6.49
2 Massachusetts         5.69
3 Alaska                5.24
4 Hawaii                5.20
5 Rhode Island          5.15
```




```
# A tibble: 5 x 2
# Groups:   state [5]
  state        price_change
  <chr>               <dbl>
1 North Dakota         1.72
2 Missouri             1.81
3 Georgia              1.98
4 California           2.03
5 Tennessee            2.07
```



I did the above but based on context I think I was actually supposed to repeat steps 6-8, so that is below.

The regression coefficient is -0.99681, which suggests sales and price have an inverse relationship. As price increases we can expect sales to decrease.

```
Call:
lm(formula = ln_sales ~ ln_price, data = q7)

Residuals:
     Min       1Q   Median       3Q      Max
-0.92230 -0.17004  0.00664  0.17869  1.10282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.65996    0.03638  155.56   <2e-16 ***
ln_price    -0.99681    0.02469  -40.37   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.296 on 1273 degrees of freedom
Multiple R-squared:  0.5614,    Adjusted R-squared:  0.5611
F-statistic:  1630 on 1 and 1273 DF,  p-value: < 2.2e-16
```

Yes, the effect is bigger than without the instrument (-1.28634 compared to -0.99681). This is because using tax as an instrument for price can help reduce the effect of confounders, better revealing the true effect.

```
TSLS estimation, Dep. Var.: ln_sales, Endo.: ln_price, Instr.: ln_tax_2012
Second stage: Dep. Var.: ln_sales
Observations: 1,275
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   6.07536   0.050098 121.2703 < 2.2e-16 ***
fit_ln_price -1.28634   0.034383 -37.4118 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.311338   Adj. R2: 0.513699
F-test (1st stage), ln_price: stat = 1,697.4, p < 2.2e-16, on 1 and 1,273 DoF.
                  Wu-Hausman: stat =   214.0, p < 2.2e-16, on 1 and 1,272 DoF.



Call:
lm(formula = ln_sales ~ pricehat, data = q7)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01081 -0.17323  0.02462  0.19135  0.99218

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.07536    0.04905  123.87   <2e-16 ***
pricehat    -1.28634    0.03366  -38.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.305 on 1273 degrees of freedom
Multiple R-squared:  0.5343,    Adjusted R-squared:  0.5339
F-statistic:  1460 on 1 and 1273 DF,  p-value: < 2.2e-16
```

10. Compare your elasticity estimates from 1970-1990 versus those from 1991-2015. Are they different? If so, why?

The elasticity estimates for 1991-2015 are less than 1970-1990, showing the price of cigarettes are more inelastic during that period. This could be because there was more change in price and taxes during the later period, so the inelasticity of the good is better determined.