

AE 05: Multiple linear regression

Houses in Levittown

Sep 21, 2022

```
library(tidyverse)
library(tidymodels)
library(knitr)
library(GGally)
```

The data set contains the sales price and characteristics of 85 homes in Levittown, NY that sold between June 2010 and May 2011. Levittown was built right after WWII and was the first planned suburban community built using mass production techniques.

```
levittown <- read_csv("data/homeprices.csv")
```

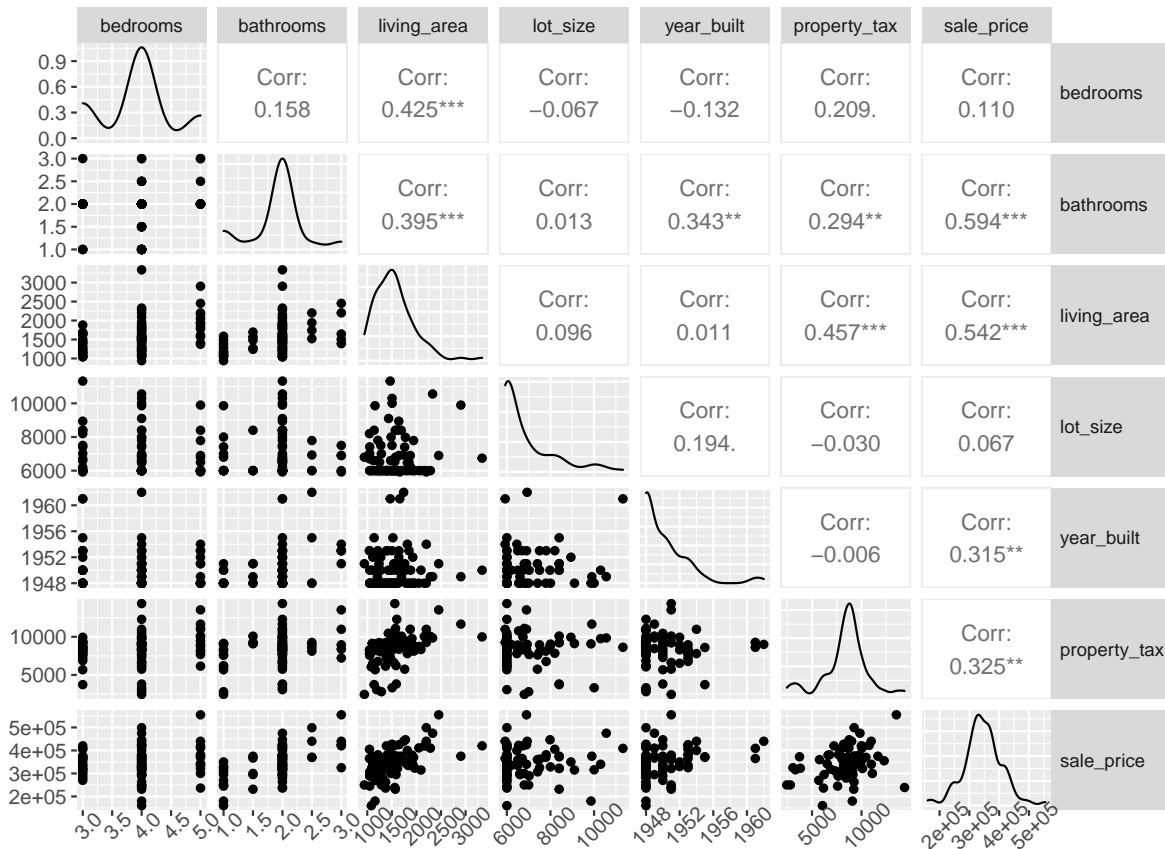
The variables used in this analysis are

- **bedrooms**: Number of bedrooms
- **bathrooms**: Number of bathrooms
- **living_area**: Total living area of the house (in square feet)
- **lot_size**: Total area of the lot (in square feet)
- **year_built**: Year the house was built
- **property_tax**: Annual property taxes (in USD)
- **sale_price**: Sales price (in USD)

The goal of the analysis is to use the characteristics of a house to understand variability in the sales price.

Exploratory data analysis

```
ggpairs(levittown) +  
  theme(  
    axis.text.y = element_text(size = 10),  
    axis.text.x = element_text(angle = 45, size = 10),  
    strip.text.y = element_text(angle = 0, hjust = 0)  
  )
```



Linear model

Fit a linear model of housing prices versus the house characteristics in Levittown. Neatly display model using 3 digits.

```
# fit model

priceModel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(sale_price ~ bedrooms + bathrooms + living_area + lot_size +
      year_built + property_tax, data = levittown)

# display model with 3 digits
priceModel %>%
  tidy() %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-7148818.957	3820093.694	-1.871	0.065
bedrooms	-12291.011	9346.727	-1.315	0.192
bathrooms	51699.236	13094.170	3.948	0.000
living_area	65.903	15.979	4.124	0.000
lot_size	-0.897	4.194	-0.214	0.831
year_built	3760.898	1962.504	1.916	0.059
property_tax	1.476	2.832	0.521	0.604

Interpretation

- Interpret the coefficient of **bedrooms** in the context of the data.

Holding all other predictor variables constant, for each additional bedroom in a house, we expect the sales price of a house in Levittown, NY, to decrease by \$12,291.01 on average.

- The intercept is the estimated sales price for what subset of houses? Be specific.

The intercept is the estimated sales price for the subset of houses with 0 bedrooms, 0 bathrooms, 0 square feet of living area, 0 square feet of lot size, and 0 dollars in annual property taxes that were built in the year 0.

Prediction

What is the predicted sale price for a house in Levittown, NY with 4 bedrooms, 2 bathrooms, 1,000 square feet of living area, 6,000 square foot lot size, built in 1947 with \$7,403 in property taxes?

- Report the predicted value and appropriate interval.

```
# create tibble for new observation
test <- tibble(bedrooms = 4, bathrooms = 2, living_area = 1000, lot_size = 6000,
               year_built = 1947, property_tax = 7403)
# prediction + interval
predict(priceModel, new_data = test)
```

```
# A tibble: 1 x 1
  .pred
  <dbl>
1 299332.
```

```
predict(priceModel, new_data = test, type = "pred_int", level = 0.95) %>%
  kable()
```

<u>.pred_lower</u>	<u>.pred_upper</u>
201784.3	396879

- Interpret the interval in the context of the data.

We are 95% confident that the predicted sale price for a house in Levittown, NY that has 4 bedrooms, 2 bathrooms, 1,000 square feet of living area, 6,000 square foot lot size, built in 1947 with /\$7,403 in annual property taxes is between \$201784.3 and \$396879.

! Important

To submit the AE:

- Render the document to produce the PDF with all of your work from today's class.
- Push all your work to your `ae-05-` repo on GitHub. (You do not submit AEs on Gradescope).