

# multiple linear regression

Rohan Krishnan

2024-07-26

## Linear Regression (and other linear methods)

### Loading Libraries/Functions

```
#Load train test split function
source("../scripts/useful-functions.R")
#Load tidyverse
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2     3.4.4      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## 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

library(stargazer, verbose = FALSE)

##
## Please cite as:
##
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

### Data + Splitting

```
#Load data
df <- read.csv("../data/modelling/modelling.csv")
df <- df %>% select(-date)

#Create training and testing sets
train_test_split(df, propTrain = 0.70, propTest = 0.30)
```

## Create LR model + Summarize

```
#Create multiple linear regression
model <- lm(clcsHPI ~ . , train)

#Look at coefficients
summary(model)
```

Call: lm(formula = clcsHPI ~ ., data = train)

Residuals: Min 1Q Median 3Q Max -8.5284 -3.3087 0.2935 2.8541 9.7541

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.481e+02 6.394e+02 0.232 0.817198

urbanCPI 1.491e-01 2.475e-01 0.602 0.547822

fedFunds 3.998e+00 4.701e-01 8.504 1.60e-14 **buildPermits -5.556e-04 3.156e-03 -0.176 0.860510**

**constructionPI 4.127e-01 4.960e-02 8.320 4.66e-14** delRate -2.281e+00 4.719e-01 -4.834 3.24e-06

**houseSub 1.053e+00 4.192e-01 2.511 0.013076**

**income -2.255e-03 1.017e-03 -2.217 0.028076 \***

**mortRate 1.441e-01 1.004e+00 0.143 0.886109**

**constructionUn 1.836e-02 4.612e-03 3.980 0.000106** totalHouse -2.659e-03 1.122e-03 -2.369 0.019073

totalConstSpend -2.290e-01 2.987e-01 -0.767 0.444416

urbanPop 7.607e-01 9.482e+00 0.080 0.936157

unemploymentRate 2.086e+00 5.839e-01 3.573 0.000473 \* **imputeGDP 7.883e-03 2.442e-03 3.229**

**0.001525** — Signif. codes: 0 ‘**0.001**’ ‘0.01’ ‘0.05’ ‘0.1’ ‘1’

Residual standard error: 4.385 on 152 degrees of freedom Multiple R-squared: 0.9899, Adjusted R-squared: 0.989 F-statistic: 1062 on 14 and 152 DF, p-value: < 2.2e-16

```
#Generate nice table
stargazer(model, type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Mon, Jul 29, 2024 - 23:19:04

Table 1:

	<i>Dependent variable:</i>
	clcsHPI
urbanCPI	0.149 (0.247)
fedFunds	3.998*** (0.470)
buildPermits	-0.001 (0.003)
constructionPI	0.413*** (0.050)
delRate	-2.281*** (0.472)
houseSub	1.053** (0.419)
income	-0.002** (0.001)
mortRate	0.144 (1.004)
constructionUn	0.018*** (0.005)
totalHouse	-0.003** (0.001)
totalConstSpend	-0.229 (0.299)
urbanPop	0.761 (9.482)
unemploymentRate	2.086*** (0.584)
imputeGDP	0.008*** (0.002)
Constant	148.063 (639.441)
Observations	167
R <sup>2</sup>	0.990
Adjusted R <sup>2</sup>	0.989
Residual Std. Error	4.385 (df = 152)
F Statistic	1,062.194*** (df = 14; 152)
<i>Note:</i> *p $\geq$ 0.1; **p<0.05; ***p<0.01	