

# HW 1

## Exercises

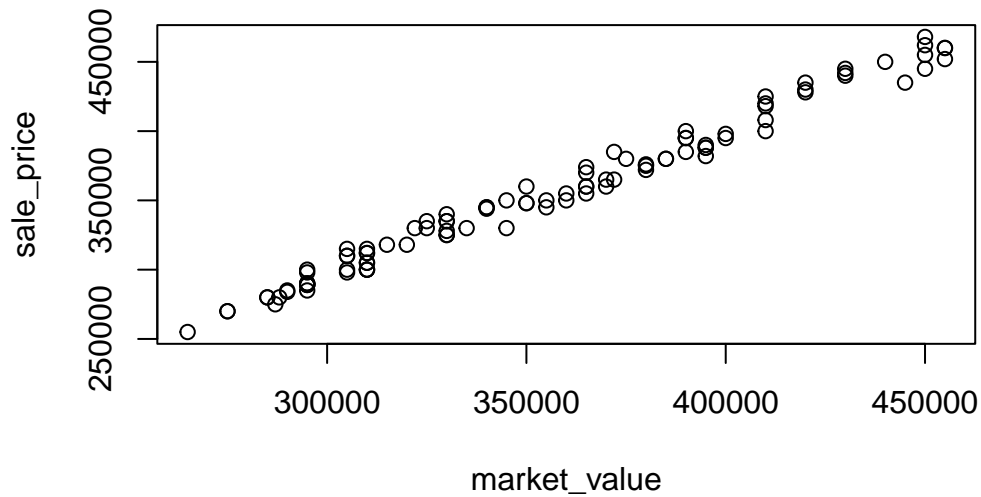
**Exercise 1** (Learning the mechanics.). Use the method of least squares to fit a straight line to these six data points:

x	1	2	3	4	5	6
y	2	4	5	4	2	7

- (a) What are the least squares estimates of  $\beta_0$  and  $\beta_1$ ?
- (b) Plot the data points and graph the least squares line on the scatterplot.

**Exercise 2** (Predicting home sales price.). Real estate investors, homebuyers, and homeowners often use the appraised (or market) value of a property as a basis for predicting sale price. Please look at the provided dataset [MARKET.csv](#). The first five and last five observations of the data set are listed in the accompanying table.

- (a) Propose a simple linear model to relate the appraised market value  $x$  to the sale price  $y$ .
- (b) A scatterplot of the data is shown below. Does it appear that a straight-line model will be an appropriate fit to the data?
- (c) A R simple linear regression printout is also shown below. Find the equation of the best-fitting line through the data on the printout.
- (d) Interpret the  $y$ -intercept of the least squares line. Does it have a practical meaning for this application? Explain.
- (e) Interpret the slope of the least squares line. Over what range of  $x$  is the interpretation meaningful?
- (f) Use the least squares model to estimate the mean sale price of a property appraised at \$300,000.



Call:

```
lm(formula = sale_price ~ market_value)
```

Residuals:

Min	1Q	Median	3Q	Max
-14674	-5480	-1287	6300	13408

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.928e+04	5.019e+03	-3.841	0.000217 ***
market_value	1.053e+00	1.399e-02	75.256	< 2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7234 on 98 degrees of freedom

Multiple R-squared: 0.983, Adjusted R-squared: 0.9828

F-statistic: 5663 on 1 and 98 DF, p-value: < 2.2e-16

**Exercise 3.** A study shows that during a certain sport the mean heart rate  $y$  and the maximal oxygen uptake  $x$  might have relations. The dataset `SPORTHR.csv` shows  $y$  (expressed as a percentage of maximum heart rate) and  $x$  ( $V_{O_2\max}$ ). The data are shown in the table.

player	vo2max	meanHR
1	140	68.2
2	150	71.1
3	160	74.4

4	170	76.5
5	180	78.8
6	185	80.1
7	190	82.4
8	200	84.6

- Find the equation of the least squares line.
- Give a practical interpretation (if possible) of the  $y$ -intercept of the line.
- Give a practical interpretation (if possible) of the slope of the line.

**Exercise 4** (Spreading rate of spilled liquid.). A researcher studied the rate at which a spilled liquid will spread across a surface. The mass (in pounds) of the spill after a period of time ranging from 0 to 60 minutes is recorded and shown below (based on the dataset `SPILLS.csv`). Do the data indicate that the mass of the spill tends to diminish as time increases? If so, how much will the mass diminish each minute?

time_min	mass_lb
0	330.0
5	325.3
10	319.8
15	315.2
20	309.7
25	305.1
30	300.3
35	295.4
40	290.2
45	284.7
50	279.9
55	274.8
60	269.5

**Exercise 5** (Sweetness of orange juice.). To study the sweetness of orange juices, researchers collect some data on the sweetness index ( $y$ ) and the amount of pectin ( $x$ ) in the orange juice (in  $g/L$ ). The dataset is `ORANGEJUICE.csv`.

sample	pectin	sweetness
1	0.15	6.8
2	0.20	7.1
3	0.25	7.4
4	0.30	7.9
5	0.35	8.2
6	0.40	8.6

7	0.50	9.1
8	0.60	9.5
9	0.70	10.1
10	0.80	10.4
12	0.90	10.9
13	1.00	11.3

- Find the values of  $SSE$ ,  $s^2$ , and  $s$  for this regression.
- Estimate  $\sigma^2$ , the variance of the random error term in the model.
- Estimate  $\sigma$ , the standard deviation of the random error term in the model.
- Explain why it is difficult to give a practical interpretation to  $s^2$ , the estimate of  $\sigma^2$ .
- Give a practical interpretation of the value of  $s$ .