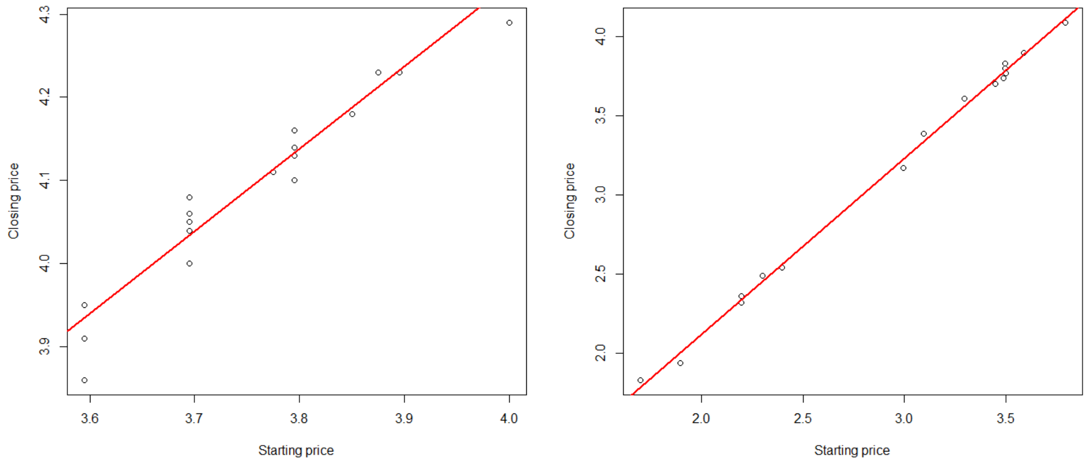

Figure 40: Scatter plots and fitted regressions of the closing price in terms of starting price for two sets of housing units.

7 Multiple linear regression: the descriptive approach

We return to the situation where we have $K+1$ numerical variables $x_1, x_2, \cdots, x_K$ and y on a set s of n elements as illustrated in Table 24. With this information we want to describe y as a function of the x-variables, i.e. we want to express y as $y = f(x_1, x_2, \cdots, x_K)$. We will restrict ourselves to the case where we want to express y as a linear function of the x-variables,

$$y = f(x_1, x_2, \cdots, x_K) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K.$$

We illustrate the idea with an example.

Example 58. Let s be the set of $n = 20$ housing units considered in Example 57. Now we are going to consider a bigger set of variables which is shown in Table 30. For the time being, let y_i = closing price of the ith unit, x_{1i} = area (is squared meters) of the ith unit and x_{2i} = starting price of the ith unit. This means that we want to explain the closing price as a linear function of the size of the unit and the starting price. $\square$

Once again, we are looking for the linear function that minimizes the distance from the fitted regression (a hyperplane in this case) to the observed points. Our criterion for measuring distance is the sum of squares error —SSE—, which in this case is given by

$$SSE = \sum_s e_i^2 \qquad \text{where} \qquad e_i = y_i - \hat{y}_i \qquad \text{and} \qquad \hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_K x_{Ki}.$$

The solution to this problem is, again, the least squares regression. We show the formula for the coefficients $b_0, b_1, \cdots, b_K$ of the least squares regression in this case just for completeness, we will not be using this formula. We will rely on Excel for making the calculations. We will concentrate on the interpretation of the results.

Definition 59. Let $(x_{1i}, x_{2i}, \cdots, x_{Ki}, y_i)$ for $i = 1, 2, \cdots, n$ be the observations of K independent (explanatory) variables x and one dependent variable y on a set s of n elements. Let also

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ 1 & x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix} \qquad \text{and} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Municipality	Balcony	Area	Room	Starting	Closing
Järfälla	0	88.0	4	3.590	3.90
Stockholm	1	61.0	3	2.995	3.17
Upplands Väsby	1	75.0	3	3.495	3.83
Upplands Väsby	0	93.0	4	3.495	3.80
Other	0	69.0	3	3.500	3.77
Huddinge	1	83.0	3	2.195	2.36
Järfälla	1	73.0	3	2.195	2.32
Järfälla	1	72.0	3	2.195	2.32
Järfälla	0	78.0	3	3.450	3.70
Järfälla	1	61.0	3	1.895	1.94
Upplands Väsby	1	62.0	3	2.395	2.54
Järfälla	0	150.0	3	1.699	1.83
Upplands Väsby	1	79.0	3	2.195	2.36
Other	0	72.0	3	2.995	3.17
Stockholm	1	73.5	3	3.295	3.61
Upplands Väsby	1	62.0	3	2.395	2.54
Huddinge	1	70.0	3	3.795	4.09
Järfälla	1	75.0	3	3.490	3.74
Stockholm	1	81.0	3	3.095	3.39
Stockholm	1	83.0	3	2.300	2.49

Table 30: Housing units dataset

The hyperplane that minimizes the SSE is given by the least squares regression,

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K$$

with $[b_0 \quad b_1 \quad \cdots \quad b_K]' = b$ and

$$b = (X'X)^{-1}X'Y. \qquad \square$$

The interpretation of the coefficients resembles the interpretation given in the simple linear case, with some differences. The intercept b_0 indicates the value of the dependent variable y that is expected when all the independent variables $x_1, x_2, \cdots, x_K$ are equal to zero. Each slope b_k indicates that —having all other explanatory variables unchanged— on average, a unit change in x_k is associated to a change of b_k in the dependent variable y.

If one explanatory variable, say x_k, is a dummy variable that indicates with one the presence of some characteristic and with zero the absence of the characteristic, its associated coefficient b_k is interpreted as the expected additional effect on the dependent variable y due to the presence of the characteristic of interest.

Example 60. Consider the dataset of $n = 20$ housing units introduced in Example 58. The least squares regression that explains the closing price y in terms of the size x_1, the starting price x_2 and a dummy variable x_3 which indicates the presence of a balcony gives $b_0 = -0.3343$, $b_1 = 0.001774$, $b_2 = 1.132$ and $b_3 = 0.04665$, i.e.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 = -0.3343 + 0.001774 x_1 + 1.132 x_2 + 0.04665 x_3.$$

This means that an unit of 0 m^2 and with a starting price of 0 SEK and without a balcony is expected to have a closing price of -0.33 million SEK. Clearly, this value does not make sense. But, in this case, it does not make sense either to consider a unit with area equal to 0 or a starting price of 0. Regarding the slopes, we have that, having all other variables the same, a unit that is 1 m^2 bigger than another is expected to have a closing price 0.001774 million SEK higher. Also, having all other variables the same, a unit with a starting price 1 million SEK higher than another is expected to have a closing price 1.132 million SEK higher. Finally, having all other variables the same, a unit with balcony is expected to have a closing price 0.04665 higher than one unit without a balcony. $\square$

Categorical variables as independent variables

Taking a look at Table 30, it makes sense to believe that the municipality may be a good explanatory variable for the closing price of a housing unit. In order to make use of a categorical variable x as explanatory variable in a regression we have to express this variable as a set of dummy variables. In Section 1.3 we saw that a categorical variable with K categories can be rewritten as $K-1$ dummy variables $x_1, x_2, \cdots, x_{K-1}$ where

$$x_{ki} = \begin{cases} 1 & \text{if } x_i \text{ takes the } k\text{th category} \\ 0 & \text{otherwise} \end{cases}$$

The Kth category, for which no dummy variable is created, is considered to be the *reference category*.

A categorical variable x can be used in a multiple regression by making use of these $K-1$ dummy variables. The coefficients are then interpreted as follows. The intercept is the effect due to the reference category, the coefficient associated to the kth dummy variable is the additional effect due to the kth category.

Example 61. Consider the dataset of $n = 20$ housing units introduced in Example 58. In order to explain the closing price y in terms of the municipality, we have to rewrite the municipality as a set of dummy variables. As the municipality takes $K = 5$ categories (Huddinge, Järfälla, Stockholm, Upplands Väsby, Other), we will create 4 dummy variables: indicating whether the unit is located in Huddinge ($= x_1$), Järfälla ($= x_2$), Stockholm ($= x_3$) or Upplands Väsby ($= x_4$). Note that we are choosing the category Other as the reference one.

Using Excel, the least squares regression that explains the closing price y in terms of the four dummy variables x_1, x_2, x_3 and x_4 gives $b_0 = 3.47$, $b_1 = -0.25$, $b_2 = -0.65$, $b_3 = -0.31$ and $b_4 = -0.46$, i.e.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 = 3.47 - 0.25x_1 - 0.65x_2 - 0.31x_3 - 0.46x_4,$$

which means that housing units in Other municipalities are expected to have a closing price of around 3.47 millions, housing units in Huddinge are expected to cost one quarter of a million less, housing units in Upplands Väsby are expected to cost half a million less, and so on. $\square$

Once we have found the coefficients, $b_0, \cdots, b_K$, we can obtain the n *fitted values*, these are the approximations to the observed y values made by the multiple regression, i.e.

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_K x_{Ki} \qquad (i = 1, 2, \cdots, n).$$

We can also find the n *residuals*, these are the distances from the true y values to the fitted y values, i.e.

$$e_i = y_i - \hat{y}_i \qquad (i = 1, 2, \cdots, n).$$

Example 62. Consider the dataset of $n = 20$ housing units in Example 60 where we fitted a multiple regression that explains the closing price y in terms of the size x_1, the starting price

x_2 and a dummy variable x_3 that indicates whether the unit has a balcony. Using Excel we found that $b_0 = -0.3343$, $b_1 = 0.001774$, $b_2 = 1.132$ and $b_3 = 0.04665$, therefore we obtain the fitted values $\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}$ and the residuals $e_i = y_i - \hat{y}_i$ shown in Table 31. $\square$

i	x_{1i}	x_{2i}	x_{3i}	y_i	$\hat{y}_i$	e_i
1	88.0	3.59	0	3.90	3.89	0.01
2	61.0	3.00	1	3.17	3.21	-0.04
3	75.0	3.50	1	3.83	3.80	0.03
4	93.0	3.50	0	3.80	3.79	0.01
5	69.0	3.50	0	3.77	3.75	0.02
6	83.0	2.19	1	2.36	2.34	0.02
7	73.0	2.19	1	2.32	2.33	-0.01
8	72.0	2.19	1	2.32	2.32	0.00
9	78.0	3.45	0	3.70	3.71	-0.01
10	61.0	1.90	1	1.94	1.97	-0.03
11	62.0	2.40	1	2.54	2.53	0.01
12	150.0	1.70	0	1.83	1.85	-0.02
13	79.0	2.19	1	2.36	2.34	0.02
14	72.0	3.00	0	3.17	3.18	-0.01
15	73.5	3.29	1	3.61	3.57	0.04
16	62.0	2.40	1	2.54	2.53	0.01
17	70.0	3.80	1	4.09	4.13	-0.04
18	75.0	3.49	1	3.74	3.80	-0.06
19	81.0	3.10	1	3.39	3.36	0.03
20	83.0	2.30	1	2.49	2.46	0.03

Table 31: Fitted values and residuals in the dataset of 20 housing units

Interactions

Just for convenience let us consider a multiple regression with two explanatory variables, i.e. we want to explain one variable y in terms of two variables x_1 and x_2. If the observations look more or less like the three dimensional scatter plot on the left panel of Figure 41 then it is evident that the dependent variable is adequately described as a linear function of x_1 and x_2, i.e. a plane. On the right panel, however, it is apparent that y can still be adequately described as a function of x_1 and x_2 but this function is not linear anymore. One option is to add the *interaction* term between x_1 and x_2 to the regression, this simply means to add the product of x_1 and x_2 to the model, i.e we fit the regression:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2. \tag{15}$$

We have fitted a plane to the dataset on the right panel of Figure 41, this yields a poor fit as shown on the left panel of Figure 42. On the other hand, when we fit a model of the type (15) we obtain a much better fit, as shown on the right panel of Figure 42. Needless to say, this was a simulated dataset, so it is not surprising that we obtain this excellent fit. In practice usually things may not be so nice, but adding an interaction may be useful in many occasions.

Let us consider the regression that explains y in terms of a numerical variable x_1 and a dummy variable x_2. Let us discuss the following four regressions:

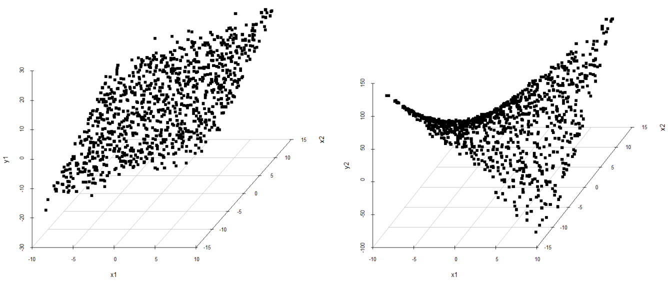

Figure 41: 3D Scatter plots of three variables x_1, x_2 and y.

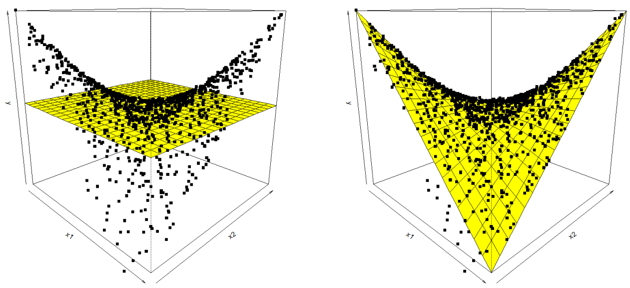

Figure 42: 3D Scatter plots and fitted regressions of three variables x_1, x_2 and y. Without interaction (left panel) and with interaction (right panel).

1. Regression 1. $\hat{y} = b_0 + b_1x_1$. Here we are simply ignoring the dummy variable x_2 and we are fitting a regression that explains y in terms of x_1. The top–left panel of Figure 43 illustrates this regression.

2. Regression 2. $\hat{y} = b_0+b_1x_1+b_2x_2$. When the dummy variable $x_2 = 0$ we get $\hat{y} = b_0+b_1x_1$ but when the dummy variable $x_2 = 1$ we get $\hat{y} = b_0 + b_1x_1 + b_2 = (b_0 + b_2) + b_1x_1$. Thus the effect of introducing a dummy variable in the regression is that we are fitting two parallel lines (same slope, different intercept) to the data: one for the elements with $x_2 = 0$ and another one for the elements with $x_2 = 1$. The top–right panel of Figure 43 illustrates this regression.

3. Regression 3. $\hat{y} = b_0 + b_1x_1 + b_2x_1x_2$. When the dummy variable $x_2 = 0$ we get $\hat{y} = b_0+b_1x_1$ but when the dummy variable $x_2 = 1$ we get $\hat{y} = b_0+b_1x_1+b_2x_1 = b_0+(b_1+b_2)x_1$. Thus the effect of introducing the interaction between a dummy variable and a numerical variable in the regression (but not the "main effect" of the dummy variable) is that we are fitting two lines having the same intercept but a different slope to the data. The bottom–left panel of Figure 43 illustrates this regression.

4. Regression 4. $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2$. When the dummy variable $x_2 = 0$ we get $\hat{y} = b_0 + b_1x_1$ but when the dummy variable $x_2 = 1$ we get $\hat{y} = b_0 + b_1x_1 + b_2 + b_3x_1 = (b_0 + b_2) + (b_1 + b_3)x_1$. Thus the effect of introducing the interaction between a dummy variable and a numerical variable in the regression as well as the "main effect" of the dummy variable is that we are fitting two lines with different intercepts and different slopes to the data. The bottom–right panel of Figure 43 illustrates this regression.

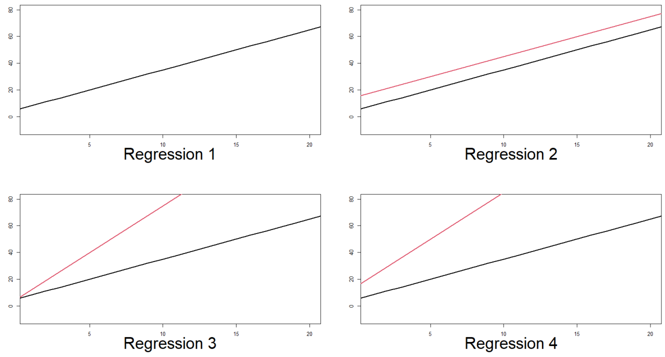

Figure 43: Plots of four regressions

Thus, simply by adding interactions we can fit very flexible models. Let us illustrate the idea with some examples.

Example 63. Some variables were measured on a set of 4137 students in a university in the US. We want to fit a regression that explains the Grade Point Average (GPA, $= y$) in terms of their SAT score ($= x_1$) and a dummy variable indicating if the individual is an athlete ($= x_2$). Figure 44 shows the scatter plot between the SAT score x_1 and the GPA y. Athletes are shown as red dots, and non-athletes as black dots. The regression line

$$\hat{y} = 0.66 + 0.19x_1$$

is overplotted in green. The resulting coefficient of determination is $R^2 = 16.7\%$. Visually, we do not see any difference between athletes and non-athletes. Thus, this simple model seems to adequately describe the data. In addition, fitting the more complex model that includes x_2 and the interaction between x_1 and x_2 yields $R^2 = 16.7\%$, so adding more terms to the regression do not yield a better fit. □

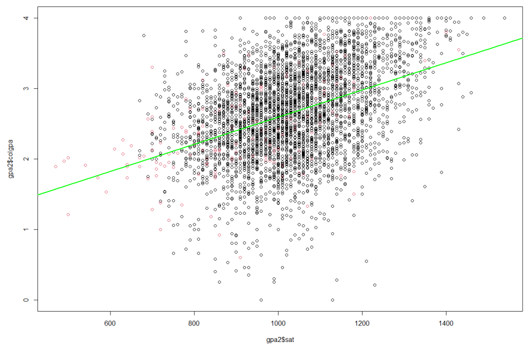

Figure 44: Scatter plot of SAT scores vs. GPA of 4137 students. Red: athletes; Black: non-athletes

Example 64. Considering the same population as in Example 63 a second regression was fitted. Now we want to explain the GPA ($= y$) in terms of the "total hours through fall semester" (*tothrs*, apparently the number of hours studying) ($= x_1$) and a dummy variable indicating if the individual is an athlete ($= x_2$). Figure 45 shows the scatter plot between *tothrs* x_1 and GPA y. Athletes are shown as red dots, and non-athletes as black dots. The regression

$$\hat{y} = 2.53 + 0.0025x_1 - 0.29x_2$$

was estimated. The resulting lines for athletes and non-athletes are overplotted in red and black, respectively.

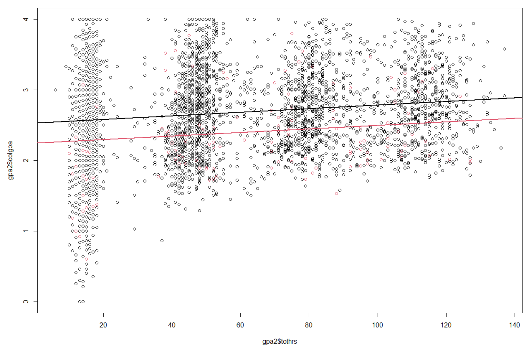

Figure 45: Scatter plot of *tothrs* vs. GPA of 4137 students. Red dots: athletes; Black dots: non-athletes

Why this model and not the simpler one that explains y in terms of only x_1 or a more complicated one that includes interaction? It may not seem evident from the scatter plot that there is a difference between athletes and non-athletes, however, take into account that the fitted model yields $R^2 = 2.7\%$. The "simpler model" yields $R^2 = 1.8\%$ so the chosen model fits much better (relatively speaking) than the simpler one. On the other hand, the model including the interaction yields $R^2 = 2.7\%$, so it fits the data as well as our chosen one. □

Example 65. Let us consider the set of $N = 392$ automobiles again. We want to explain the *autonomy* i.e. the miles per gallon ($= y$) in terms of the *horsepower* ($= x_1$) and a dummy variable indicating if the automobile is American ($= x_2$). Figure 46 shows the scatter plot between horsepower x_1 and mpg y. American automobiles are shown as red dots, and non-American automobiles are shown as black dots. The regression

$$\hat{y} = 47.34 - 0.23x_1 - 12.86x_2 + 0.11x_1x_2$$

was estimated. The resulting lines for American and non-American automobiles are overplotted in red and black, respectively.

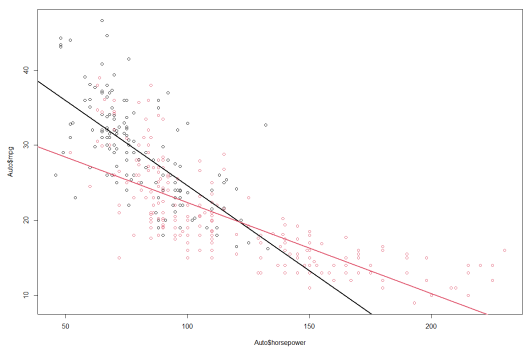

Figure 46: Scatter plot of horsepower vs. mpg of 392 automobiles. Red dots: American; Black dots: non-American

This time it is more evident (at least to me) that the fitted model describes the data better than a simpler one. This is supported by the R^2: the simpler model that includes x_1 only yields $R^2 = 60.6\%$. Adding the dummy variable R^2, we get $R^2 = 65.1\%$. Whereas adding the interaction between them we obtain $R^2 = 67.2\%$.

Explanatory power of a multiple regression equation
Once we have obtained the least squares regression, i.e. once we have found the coefficients $b_0, b_1, \cdots, b_K$ to be used in the linear function

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K,$$

the next natural step is to evaluate how good does this function describe the set of n observations. As in the simple linear regression case, in order to do this, we use the coefficient of determination R^2.

Definition 66. The *coefficient of determination* of a multiple linear regression, denoted by R^2, is

$$R^2 = \frac{SSR}{SST} \qquad \text{or equivalently} \qquad R^2 = 1 - \frac{SSE}{SST},$$

where

- $SST = \sum_s (y_i - \bar{y}_s)^2$ is the *sum of squares total*;
- $SSE = \sum_s (y_i - \hat{y}_i)^2$ is the *sum of squares error*;
- $SSR = \sum_s (\hat{y}_i - \bar{y}_s)^2$ is the *sum of squares regression*

and

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_K x_{Ki}. \qquad \square$$

R^2 have the same properties and interpretation as in the case of the simple linear regression, i.e. it takes values between 0 and 1, where 0 indicates that the regression does not fit well to the data and 1 indicates a perfect linear association between the explanatory variables and y. R^2 measures the proportion of variability in the dependent variable y that is explained by the independent variables x.

In the simple linear regression case, we saw that R^2 is equal to the squared coefficient of correlation between x and y, $r_{xy,s}$. In the multiple linear regression case, there is a similar result.

Result 67. Let R^2 be the coefficient of determination of the least squares regression of a variable y in terms of $x_1, x_2, \cdots, x_K$ and let $r_{y\hat{y},s}$ be the coefficient of correlation between y and the fitted values $\hat{y}$. Then, the following relation holds

$$R^2 = r^2_{y\hat{y},s}. \qquad \square$$

The correlation between y and the fitted values $\hat{y}$ is known as *coefficient of multiple correlation.*

Example 68. Let us return to the regression fitted in Example 60 where we explained the closing price y in terms of the size x_1, the starting price x_2 and the indicator of having a balcony x_3. The fitted values $\hat{y}$ and residuals e were found in Example 62. The SSE is

$$SSE = \sum_s (y_i - \hat{y}_i)^2 = \sum_s e_i^2 = 0.01^2 + (-0.04)^2 + \cdots + 0.03^2 = 0.01371.$$

Taking into account that $\bar{y}_s = 3.04$, the SST is

$$SST = \sum_s (y_i - \bar{y}_s)^2 = (3.90 - 3.04)^2 + (3.17 - 3.04)^2 + \cdots + (2.49 - 3.04)^2 = 10.42.$$

Thus, the coefficient of determination is

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{0.01371}{10.42} = 99.868\%,$$

which means that 99.868% of the variation in y is being explained by the three explanatory variables. □

Every explanatory (independent) variable that is added to a multiple linear regression will increase (or at least leave unchanged) the value of the coefficient of determination R^2, disregarding if this new variable has no explanatory power on y. Many statisticians see this as an undesired characteristic of R^2. For this reason, an alternative statistic is sometimes preferred. This statistic is called the *adjusted coefficient of determination*.

Definition 69. The *adjusted coefficient of determination* of a multiple linear regression, denoted by $\bar{R}^2$, is

$$\bar{R}^2 = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)}. \qquad \square$$

Example 70. Let us illustrate the fact that R^2 tends to increase with any new variable that is added to the regression disregarding if this variable has real explanatory power. In Example 68 we found that the coefficient of determination of the regression that explains the closing price y in terms of the size x_1, the starting price x_2 and the indicator of having a balcony x_3 is $R^2 = 99.868\%$. We add a fourth explanatory variable which is simply random observations from a uniform distribution. By construction, this variable has no explanatory power on y, however, the value of the coefficient of determination in this case was $R^2 = 99.870\%$.

On the other hand, the adjusted coefficient of determination for the regression with only three explanatory variables is $\bar{R}^2 = 99.844\%$, whereas the one for the regression where we added a fourth variable which is simply noise is $\bar{R}^2 = 99.835\%$, which is smaller than the previous one, indicating that the added variable is counterproductive in the regression. □

In Section 6 we mentioned that R^2 can be used as a threshold for deciding if a variable x is useful as explanatory variable for y. A similar reasoning can be used in multiple linear regression in order to choose the variables to keep in the equation. Let us say that a regression with $K-1$ explanatory variables $x_1, x_2, \cdots, x_{K-1}$ has been fitted and yields a coefficient of determination equal to R^2_{K-1}. We want to know if it is worth adding a Kth variable x_K to the regression. Then we fit the regression with all K variables and we compare its coefficient of determination R^2_K against R^2_{K-1}. If the increment is larger than some pre-specified threshold, say one percentage unit, we keep the variable x_K, otherwise we remove it from the regression.

Example 71. A first regression explaining the closing price y in terms of a dummy variable indicating the presence of a balcony x_1 has been fitted. It yields a coefficient of determination $R^2 = 8.33\%$.

We plan to add the size x_2 to the regression. We will keep x_2 if R^2 increases by at least 5 percentage units. We obtain a coefficient of determination $R^2 = 23.68\%$. As the increment is more than 5 percentage units, we decide to keep the size x_2 into the regression. □

It is important to take into account that R^2 is not additive. Let us say that the coefficient of determination of a regression that explains y in terms of x_1 is R^2_1 and the coefficient of determination of a regression that explains y in terms of x_2 is R^2_2. The coefficient of determination of a regression that explains y in terms of both x_1 and x_2 is, in general, not $R^2_1 + R^2_2$. It can be either larger or smaller.

Summary and R output

Let us summarize all the statistics that we have defined in the frame of a multiple linear regression. We have defined:

- the coefficients $b_0, b_1, \cdots, b_K$. (**Note:** in this course you will not be asked to compute these coefficients "by hand". They will either be given or we will use R for computing them);
- the fitted values $\hat{y}_i$ and the residuals $e_i = y_i - \hat{y}_i$;
- the sums of squares: sum of squares error —SSE—, sum of squares regression —SSR— and sum of squares total —SST—;
- the coefficient of multiple correlation $r_{y\hat{y},s}$, the coefficient of determination R^2 and the adjusted coefficient of determination $\bar{R}^2$;
- the equation for predicting the value of the dependent variable y for a given set of values of the independent variables x.

Figure 47 shows the output of the multiple linear regression on the set of $n = 20$ dwelling units that has been used as a running example in Section 7. There are some quantities that we are not familiar with yet. We will introduce them later in the course. Let us describe what we should know from this output at this point:

- Under *Residuals:* we see some summary statistics of the residuals $e_i = y_i - \hat{y}_i$, namely, the minimum, the three quartiles and the maximum.
- The rows of the table *Coefficients:* show some information about the estimated coefficients. By now we are only familiar with the actual coefficients that are found in the column *Estimate*.
- The coefficient of determination R^2 is found under the name *Multiple R–squared*. The adjusted coefficient of determination $\bar{R}^2$ is found under the name *Adjusted R–squared*

References

R. De Veaux, P. Velleman, and D. Bock. *Stats: data and models.* Pearson/Addison Wesley Boston, 2021.

```
Call:
lm(formula = colgpa ~ tothrs * femath, data = gpa2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.55670 -0.46296 -0.01662  0.45293  1.45073 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.521999   0.018364 137.334   <2e-16 ***
tothrs        0.002479   0.000289   8.577   <2e-16 ***
femath       -0.165189   0.172659  -0.957    0.339    
tothrs:femath  0.002631   0.002679   0.982    0.326    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6528 on 4133 degrees of freedom
Multiple R-squared:  0.01836,	Adjusted R-squared:  0.01765 
F-statistic: 25.77 on 3 and 4133 DF,  p-value: < 2.2e-16
```

Figure 47: R output of a multiple linear regression on the dataset of dwelling units.