Chapter 2: simple regression

We start again by loading the data.

```r
# Load the datasets from the RData file
load("../dataCreated/listings_clean.RData")

# Display the first few rows of the cleaned dataset
head(listings_clean)
```

```
##        id                          listing_url   scrape_id last_scraped
## 1   73155   https://www.airbnb.com/rooms/73155 2.024063e+13   2024-06-25
## 2   77592   https://www.airbnb.com/rooms/77592 2.024063e+13   2024-06-25
## 3  101526  https://www.airbnb.com/rooms/101526 2.024063e+13   2024-06-25
## 4  539905  https://www.airbnb.com/rooms/539905 2.024063e+13   2024-06-25
## 5  763422  https://www.airbnb.com/rooms/763422 2.024063e+13   2024-06-25
## 6 1322782 https://www.airbnb.com/rooms/1322782 2.024063e+13   2024-06-25
##        source                           name
## 1 city scrape    Apartment in centre of Rotterdam
## 2 city scrape      Charming, Cozy Rotterdam Center
## 3 city scrape       Apartment Dependance Rotterdam
## 4 city scrape Sunny garden apt in Kralingen/R'dam
## 5 city scrape        City appartment: b&b Walenburg
## 6 city scrape  Spacious room close to city centre
##
## 1 This bright comfortable one bedroom apartment with a architectural thouch has a true cosy feeling!
## 2
## 3
## 4
## 5                                    A lovely, comfortable stay in the city centre of Rotterd
## 6
##
## 1                                    The apartment is in a quiet side street of the Witte de W
## 2
## 3                                    Lively young neighborhood. Bistro on the corner, Lots of
## 4
## 5 The northern side of Rotterdam Centre is an up-coming area. Here you will find authentic old pre-wa
## 6
##                                                                        picture_
## 1 https://a0.muscache.com/pictures/miso/Hosting-73155/original/de9253d1-f8c9-4d29-8ef4-2b17680086ab.
## 2                     https://a0.muscache.com/pictures/e96512a1-b4c7-463e-bd52-3c99ed5f8e19
## 3                           https://a0.muscache.com/pictures/679013/0bb0d33d_original
## 4                           https://a0.muscache.com/pictures/6557859/860ad071_original
## 5                           https://a0.muscache.com/pictures/51971409/2309608f_original
## 6                           https://a0.muscache.com/pictures/66768021/f135743b_original
##   host_id                          host_url host_name host_since
## 1  381163  https://www.airbnb.com/users/show/381163       Kim 2011-02-07
## 2  416305  https://www.airbnb.com/users/show/416305       Jos 2011-03-01
## 3  531869  https://www.airbnb.com/users/show/531869      Loes 2011-04-24
```

```
## 4 2652971 https://www.airbnb.com/users/show/2652971     Michael 2012-06-16
## 5 4024319 https://www.airbnb.com/users/show/4024319  Machteld 2012-10-31
## 6 6926282 https://www.airbnb.com/users/show/6926282      Marion 2013-06-15
##            host_location
## 1 Rotterdam, Netherlands
## 2 Rotterdam, Netherlands
## 3 Rotterdam, Netherlands
## 4          New York, NY
## 5 Rotterdam, Netherlands
## 6 Rotterdam, Netherlands
##
## 1
## 2
## 3                                                            I am a mimeplayer as a professi
## 4 This apartment is my "pied a terre" in Rotterdam acquired many years ago.  Kralingen, where the ap
## 5
## 6
##   host_response_time host_response_rate host_acceptance_rate host_is_superhost
## 1     within an hour                100%                  82%             FALSE
## 2 within a few hours                100%                   0%             FALSE
## 3     within an hour                100%                 100%             FALSE
## 4                N/A                N/A                  N/A             FALSE
## 5     within an hour                100%                 100%              TRUE
## 6       within a day                 70%                  33%             FALSE
##
## 1 https://a0.muscache.com/im/pictures/user/User-381163/original/7cddb6dc-4731-43ec-9b5a-9055818b02da
## 2                     https://a0.muscache.com/im/pictures/user/b6041d69-c701-4b57-acbf-13074144089
## 3                     https://a0.muscache.com/im/pictures/user/eb0aa9f1-5fbd-4bd7-ae91-bfb3786fffe
## 4                         https://a0.muscache.com/im/users/2652971/profile_pic/1339865622/original
## 5                         https://a0.muscache.com/im/users/4024319/profile_pic/1351757885/original
## 6                     https://a0.muscache.com/im/pictures/user/5142c904-7e23-4a97-9984-3249cf8bd44
##
## 1 https://a0.muscache.com/im/pictures/user/User-381163/original/7cddb6dc-4731-43ec-9b5a-9055818b02da
## 2                     https://a0.muscache.com/im/pictures/user/b6041d69-c701-4b57-acbf-13074144089
## 3                     https://a0.muscache.com/im/pictures/user/eb0aa9f1-5fbd-4bd7-ae91-bfb3786fffe
## 4                         https://a0.muscache.com/im/users/2652971/profile_pic/1339865622/original
## 5                         https://a0.muscache.com/im/users/4024319/profile_pic/1351757885/original
## 6                     https://a0.muscache.com/im/pictures/user/5142c904-7e23-4a97-9984-3249cf8bd44
##   host_neighbourhood host_listings_count host_total_listings_count
## 1               <NA>                   1                         1
## 2               <NA>                   1                         2
## 3               <NA>                   1                         1
## 4               <NA>                   1                         1
## 5               <NA>                   2                         3
## 6               <NA>                   3                         4
##               host_verifications host_has_profile_pic host_identity_verified
## 1            ['email', 'phone']                 TRUE                   TRUE
## 2            ['email', 'phone']                 TRUE                   TRUE
## 3 ['email', 'phone', 'work_email']                 TRUE                   TRUE
## 4            ['email', 'phone']                 TRUE                   TRUE
## 5            ['email', 'phone']                 TRUE                   TRUE
## 6            ['email', 'phone']                 TRUE                   TRUE
##                       neighbourhood neighbourhood_cleansed
## 1 Rotterdam, South Holland, Netherlands         Stadsdriehoek
```

```
## 2                                   <NA>           Middelland
## 3 Rotterdam, South Holland, Netherlands            Dijkzigt
## 4                                   <NA>        Kralingen Oost
## 5 Rotterdam, South Holland, Netherlands            Blijdorp
## 6                                   <NA>        Nieuwe Westen
##   neighbourhood_group_cleansed latitude longitude
## 1             Rotterdam Centrum 51.91683   4.48294
## 2                   Delfshaven 51.91490   4.46322
## 3             Rotterdam Centrum 51.91246   4.46586
## 4          Kralingen-Crooswijk 51.92292   4.51435
## 5                       Noord 51.93021   4.46853
## 6                   Delfshaven 51.92127   4.45630
##                       property_type         room_type accommodates bathrooms
## 1               Entire rental unit Entire home/apt            2       1.5
## 2               Entire rental unit Entire home/apt            2       1.0
## 3                     Entire condo Entire home/apt            2       1.5
## 4                     Entire condo Entire home/apt            2       1.0
## 5 Private room in bed and breakfast    Private room            2       1.0
## 6              Private room in home    Private room            2       1.0
##   bathrooms_text bedrooms beds
## 1    1.5 baths        1    1
## 2        1 bath        1    2
## 3    1.5 baths        1    1
## 4        1 bath        1    1
## 5 1 private bath        1    0
## 6  1 shared bath        1    1
##
## 1 ["Coffee maker", "Private patio or balcony", "Hot water kettle", "Hot water", "Mosquito net", "Ref:
## 2
## 3
## 4
## 5                                                    ["Private patio or balcony", "Hot wate:
## 6
##   price minimum_nights maximum_nights minimum_minimum_nights
## 1   125              7           1125                      7
## 2   150              2            730                      2
## 3   140              7            365                      7
## 4   125              4             90                      4
## 5   110              2             90                      1
## 6    70              3              8                      3
##   maximum_minimum_nights minimum_maximum_nights maximum_maximum_nights
## 1                      7                   1125                   1125
## 2                      2                    730                    730
## 3                      7                   1125                   1125
## 4                      4                     90                     90
## 5                      2                   1125                   1125
## 6                      3                      8                      8
##   minimum_nights_avg_ntm maximum_nights_avg_ntm calendar_updated
## 1                      7                   1125               NA
## 2                      2                    730               NA
## 3                      7                   1125               NA
## 4                      4                     90               NA
## 5                      2                   1125               NA
## 6                      3                      8               NA
```

```
##   has_availability availability_30 availability_60 availability_90
## 1             TRUE              19              37              55
## 2             TRUE               0               0               0
## 3             TRUE               7              22              51
## 4             TRUE               0               0               0
## 5             TRUE               0               9              20
## 6             TRUE               7              11              32
##   availability_365 calendar_last_scraped number_of_reviews
## 1              301            2024-06-25               120
## 2                0            2024-06-25                18
## 3              148            2024-06-25                21
## 4                0            2024-06-25                24
## 5               65            2024-06-25               382
## 6              122            2024-06-25               137
##   number_of_reviews_ltm number_of_reviews_l30d first_review last_review
## 1                     5                      1   2011-05-29  2024-06-07
## 2                     0                      0   2011-04-12  2022-10-02
## 3                     3                      2   2011-06-03  2024-06-03
## 4                     0                      0   2012-10-13  2022-09-11
## 5                    33                      5   2013-06-10  2024-06-12
## 6                    19                      1   2015-03-28  2024-06-05
##   review_scores_rating review_scores_accuracy review_scores_cleanliness
## 1                 4.94                   4.97                      4.97
## 2                 4.78                   4.72                      4.78
## 3                 4.86                   4.89                      5.00
## 4                 4.92                   4.83                      4.83
## 5                 4.81                   4.89                      4.91
## 6                 4.91                   4.87                      4.66
##   review_scores_checkin review_scores_communication review_scores_location
## 1                  4.95                        4.95                   4.96
## 2                  4.89                        4.61                   4.83
## 3                  5.00                        5.00                   4.89
## 4                  5.00                        5.00                   4.96
## 5                  4.95                        4.93                   4.88
## 6                  4.96                        4.95                   4.85
##   review_scores_value                 license instant_bookable
## 1                4.86 0599 D9BA 806A 9A56 BDBB            FALSE
## 2                4.56                  Exempt            FALSE
## 3                4.79 0599 51CC CA3A 88B0 CEF1            FALSE
## 4                4.79 0599 F642 774A E974 C0A9            FALSE
## 5                4.78 0599 0D64 133D 0093 457A            FALSE
## 6                4.90 0599 7F3D 2D0E ECD3 0C97            FALSE
##   calculated_host_listings_count calculated_host_listings_count_entire_homes
## 1                              1                                           1
## 2                              1                                           1
## 3                              1                                           1
## 4                              1                                           1
## 5                              1                                           0
## 6                              3                                           0
##   calculated_host_listings_count_private_rooms
## 1                                            0
## 2                                            0
## 3                                            0
## 4                                            0
```

```
## 5                                                            1
## 6                                                            3
##   calculated_host_listings_count_shared_rooms reviews_per_month
## 1                                            0              0.75
## 2                                            0              0.11
## 3                                            0              0.13
## 4                                            0              0.17
## 5                                            0              2.84
## 6                                            0              1.22
```

Regress price on `review_scores_rating`.

```
# Run a linear regression of price on review_scores_rating
model <- lm(price ~ review_scores_rating, data = listings_clean)

# Display the summary of the regression model
summary(model)
```

```
##
## Call:
## lm(formula = price ~ review_scores_rating, data = listings_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -103.098  -44.641   -7.383   32.454  162.867
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)             76.86      64.11   1.199    0.231
## review_scores_rating    11.21      13.49   0.831    0.407
##
## Residual standard error: 59.04 on 431 degrees of freedom
## Multiple R-squared:  0.001599,   Adjusted R-squared:  -0.0007174
## F-statistic: 0.6903 on 1 and 431 DF,  p-value: 0.4065
```

Run a regression of price on how many people can be accommodated.

```
# Run a linear regression of price on accommodates
model_accommodates <- lm(price ~ accommodates, data = listings_clean)

# Display the summary of the regression model
summary(model_accommodates)
```

```
##
## Call:
## lm(formula = price ~ accommodates, data = listings_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -173.716  -39.084   -5.084   31.397  155.397
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  106.565      5.248  20.306  < 2e-16 ***
## accommodates   7.759      1.475   5.262 2.25e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 57.28 on 431 degrees of freedom
## Multiple R-squared:  0.06037,    Adjusted R-squared:  0.05819
## F-statistic: 27.69 on 1 and 431 DF,  p-value: 2.251e-07
```

Now we use the Strgazer package that we loaded in the beginning to produce a nice table with three regression results.

```r
# 1. Run a level-level regression of price on accommodates
level_level <- lm(price ~ accommodates, data = listings_clean)

# 2. Run a log-level regression of log(price) on accommodates
log_level <- lm(log(price) ~ accommodates, data = listings_clean)

# 3. Run a log-log regression of log(price) on log(accommodates)
log_log <- lm(log(price) ~ log(accommodates), data = listings_clean)

# 4. Create a table with stargazer displaying all three results
stargazer(level_level, log_level, log_log,
          type = "text",
          title = "Regression Results",
          column.labels = c("Level-Level", "Log-Level", "Log-Log"),
          dep.var.labels = "Price",
          covariate.labels = c("Accommodates", "Constant"),
          omit.stat = c("f", "ser"),
          digits = 3)
```

```
## 
## Regression Results
## ===========================================
##                     Dependent variable:
##              -------------------------------
##                  Price          log(price)
##              Level-Level Log-Level Log-Log
##                  (1)        (2)      (3)
## -------------------------------------------
## Accommodates   7.759***   0.055***
##               (1.475)    (0.012)
## 
## Constant                            0.409***
##                                     (0.042)
## 
## Constant      106.565***  4.591***  4.361***
##               (5.248)    (0.044)   (0.046)
## 
## -------------------------------------------
## Observations     433        433      433
## R2             0.060      0.045    0.180
## Adjusted R2    0.058      0.043    0.178
## ===========================================
## Note:            *p<0.1; **p<0.05; ***p<0.01
```