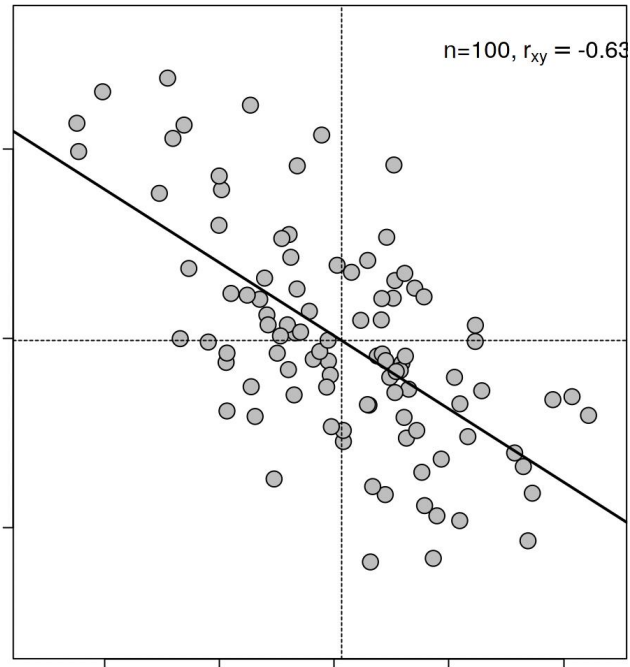


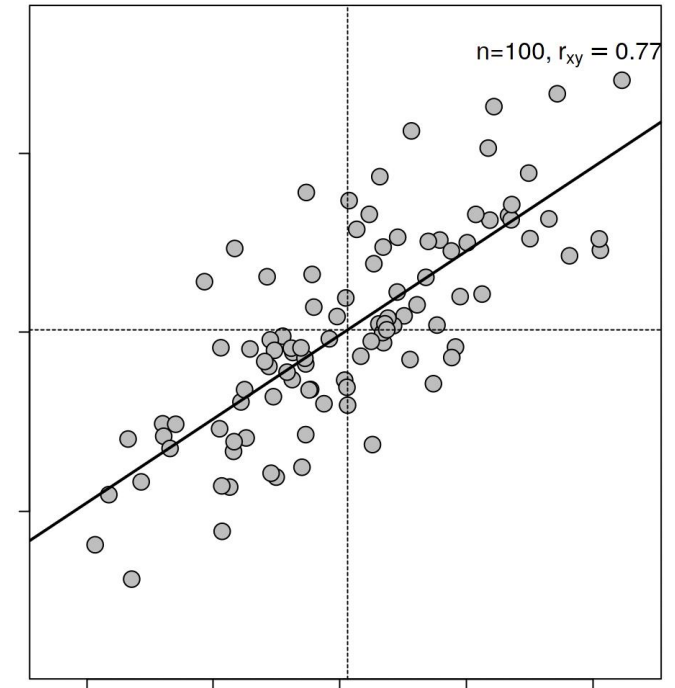
Fundamentals of Econometrics Models



Vicenç Soler

`v.soler@tbs-education.org`

~~`vincent.soler@tbs-education.org`~~



Quick reminder

What is a dependent variable?

It is the variable that we try to **explain or predict**. Usually, denoted by the letter Y.

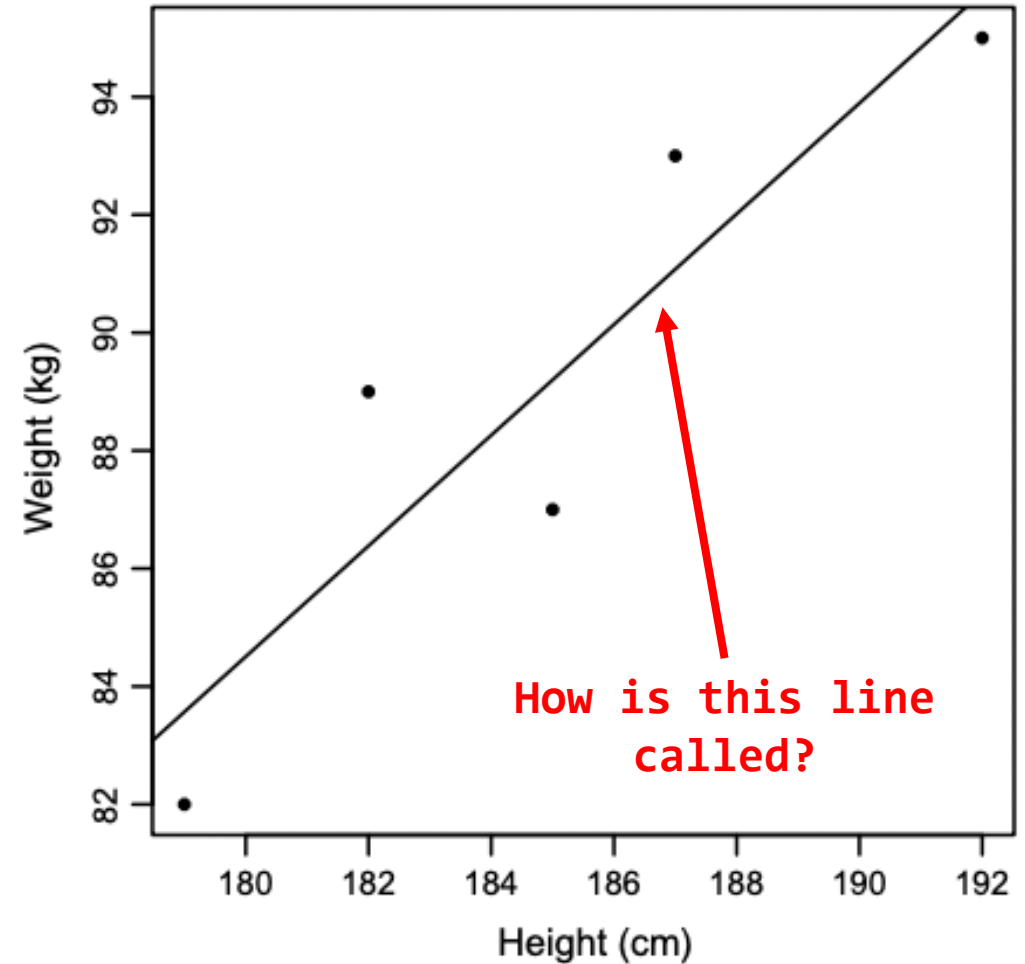
Quick reminder

What is an independent variable?

It is the variable that we use to **explain or predict** our dependent variable. Usually, denoted by the letter X.

Quick reminder

Regression line



Quick reminder

In this linear equation...

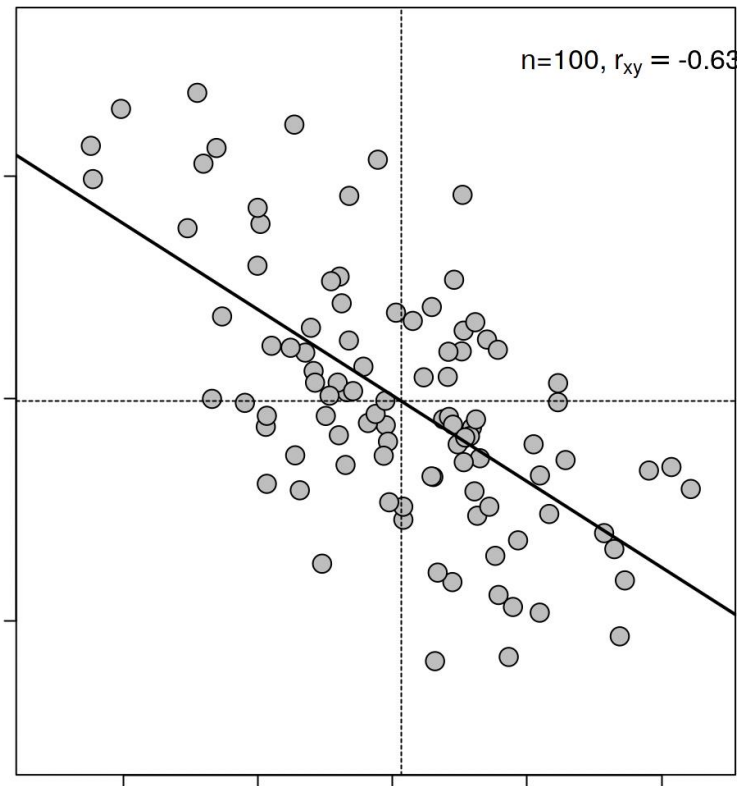
$$y = a + \beta x$$
$$\textit{Weight} = -83.47 + 0.94 * \textit{Height}$$

What is the slope? And the intercept?

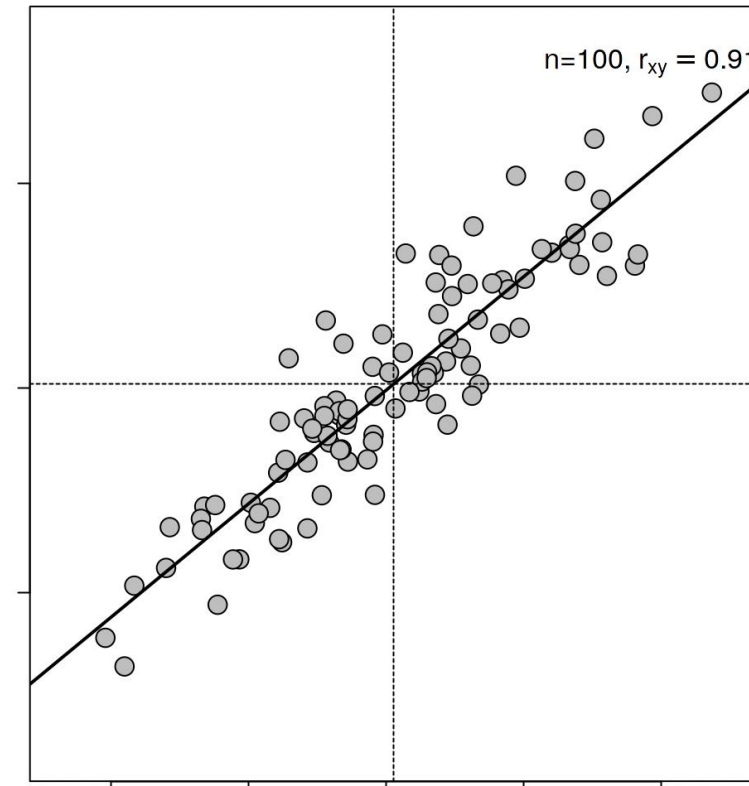
Slope = 0.93 (or β) and Intercept = -83.47 (or a)

Quick reminder

Which linear regression has a smaller error?

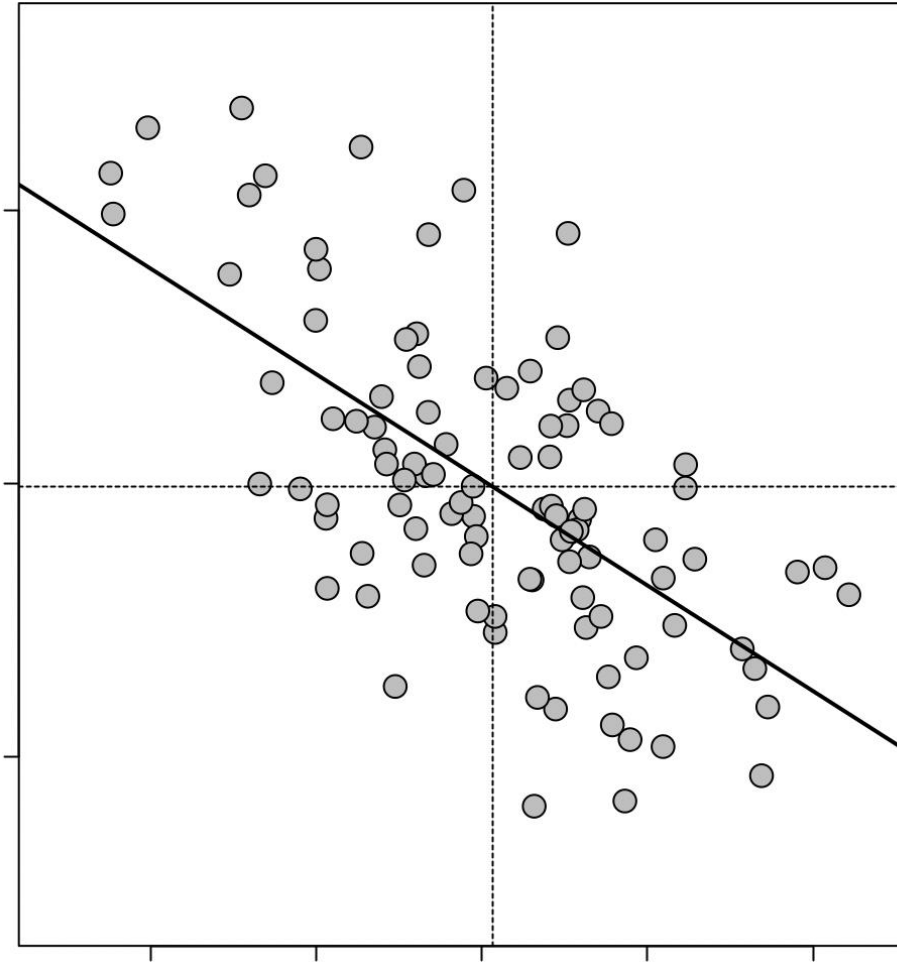


Bigger error



Smaller error

Quick reminder



Is there a correlation?

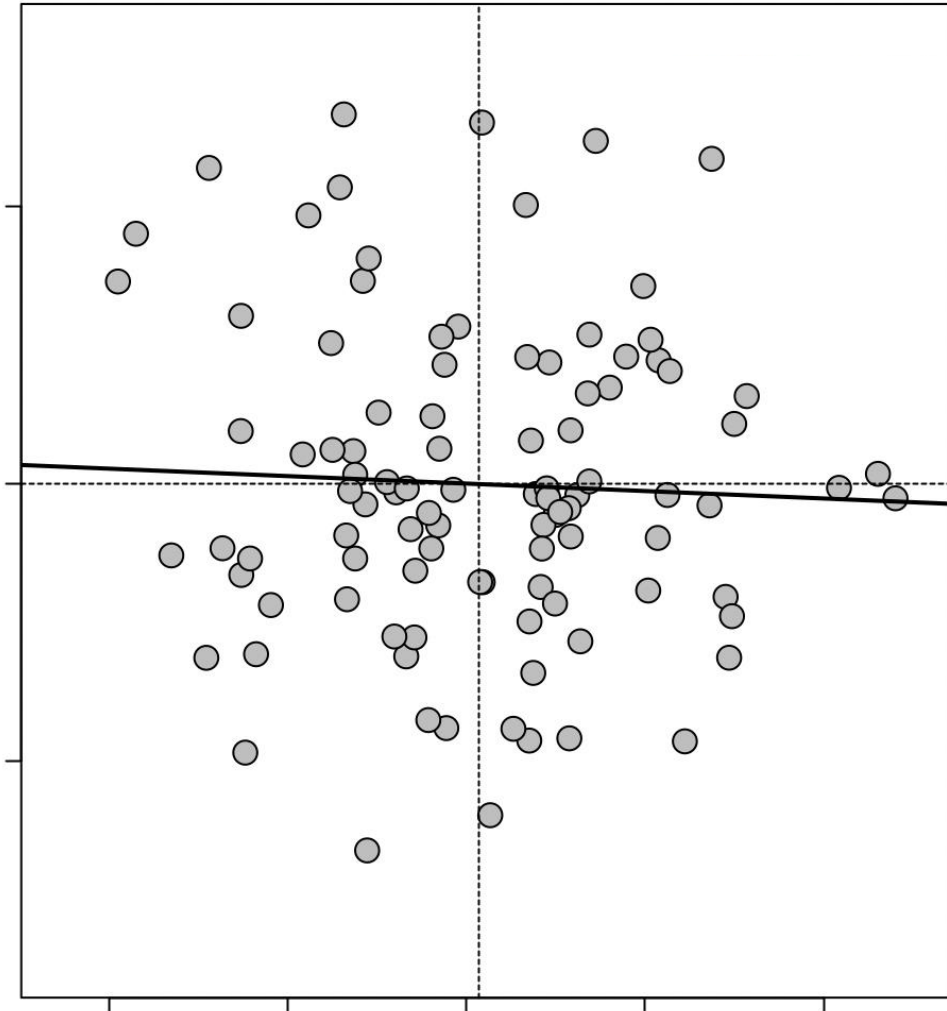
If so, positive or negative?

Weak or strong?

Can you guess the correlation coefficient (between -1 and 1)?

$$R = -0.63$$

Quick reminder



Is there a correlation?

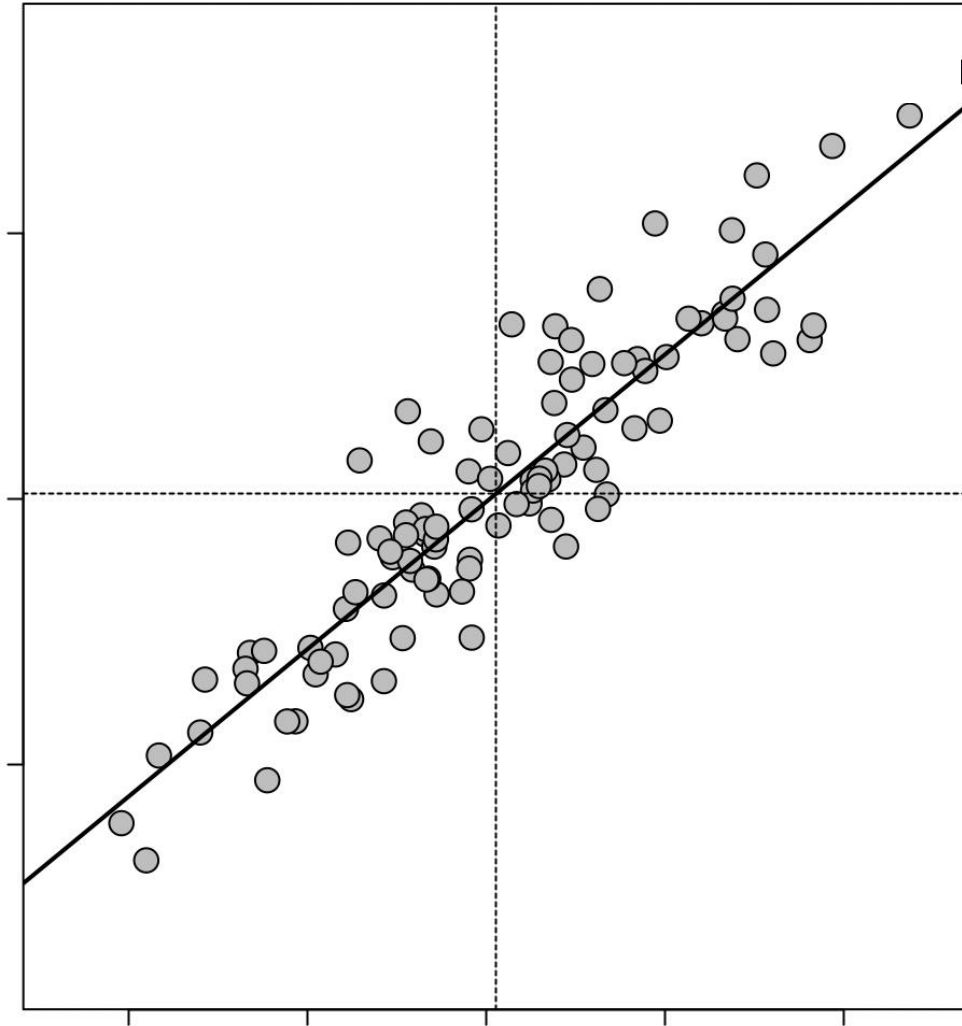
If so, positive or negative?

Weak or strong?

Can you guess the correlation coefficient (between -1 and 1)?

$$R = -0.04$$

Quick reminder



Is there a correlation?

If so, positive or negative?

Weak or strong?

Can you guess the correlation coefficient (between -1 and 1)?

$R = 0.91$

Practice for the final project

```
Call:
lm(formula = Mshare ~ TropPremium + Trop + MMAid + Aldi, data = dataset)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-8.9771 -2.9296 -0.6031  2.3847 16.6914
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.4101     8.2955   1.375   0.172
TropPremium   8.3962     6.9659   1.205   0.231
Trop         -4.1488     6.9982  -0.593   0.554
MMAid        -8.5738     0.7558 -11.343 < 2e-16 ***
Aldi          4.1263     0.7390   5.584 1.57e-07 ***
```

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

```
Residual standard error: 4.519 on 116 degrees of freedom
Multiple R-squared:  0.5627,    Adjusted R-squared:  0.5476
F-statistic: 37.32 on 4 and 116 DF, p-value: < 2.2e-16
```

| | Dependent variable: | | | |
|---------------------|----------------------------|--------------------------|------------------------|--------------------------|
| | Resistance (1) | Resistance (2) | Resistance (3) | Resistance (4) |
| Cement | 0.988*** (0.027) | | | 0.846*** (0.056) |
| Additives | | 64.266*** (2.682) | | |
| Water | | | 0.196*** (0.067) | |
| Additives | | | | 15.160*** (5.284) |
| Water | | | | -0.073 (0.065) |
| Constant | 0.738 (7.150) | 97.890*** (6.694) | 211.884*** (14.240) | 15.925 (14.846) |
| Observations | 804 | 804 | 804 | 804 |
| R2 | 0.617 | 0.417 | 0.011 | 0.623 |
| Adjusted R2 | 0.617 | 0.417 | 0.009 | 0.621 |
| Residual Std. Error | 37.778 (df = 802) | 46.616 (df = 802) | 60.740 (df = 802) | 37.563 (df = 800) |
| F Statistic | 1,293.782*** (df = 1; 802) | 574.388*** (df = 1; 802) | 8.700*** (df = 1; 802) | 439.929*** (df = 3; 800) |

Note: *p<0.1; **p<0.05; ***p<0.01

```
> model_parameters(z1, summary = TRUE)
```

| Parameter | Coefficient | SE | 95% CI | t(116) | p |
|-------------|-------------|------|-----------------|--------|--------|
| (Intercept) | 11.41 | 8.30 | [-5.02, 27.84] | 1.38 | 0.172 |
| TropPremium | 8.40 | 6.97 | [-5.40, 22.19] | 1.21 | 0.231 |
| Trop | -4.15 | 7.00 | [-18.01, 9.71] | -0.59 | 0.554 |
| MMAid | -8.57 | 0.76 | [-10.07, -7.08] | -11.34 | < .001 |
| Aldi | 4.13 | 0.74 | [2.66, 5.59] | 5.58 | < .001 |

```
Model: Mshare ~ TropPremium + Trop + MMAid + Aldi (121 observations)
Residual standard deviation: 4.519 (df = 116)
R2: 0.563; adjusted R2: 0.548
```

Practice for the final project



**Create groups of 3 or 4 people
and sit next to each other**

Practice for the final project

The main objective is to apply, **at least,
a multiple linear regression**

**Other techniques will add extra points
to your grade**

Practice for the final project

Minimum requirement

A multiple linear regression

Extra points

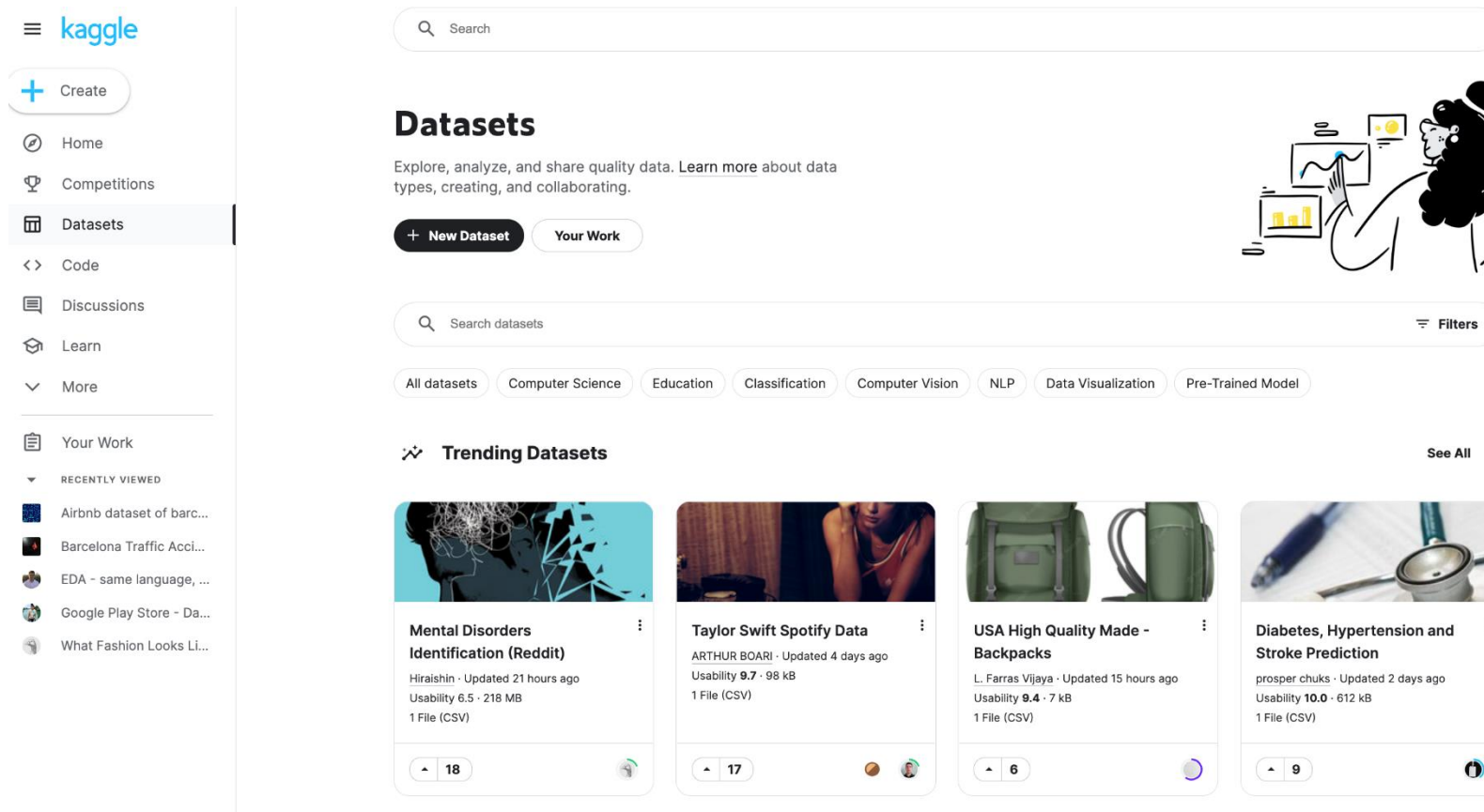
Several simple linear regressions

scatter plots

(Nice) data visualization using R...









Practice for the final project

Go to <https://www.kaggle.com/datasets>



The screenshot shows the Kaggle Datasets homepage. On the left is a sidebar with navigation links: Home, Competitions, Datasets (selected), Code, Discussions, Learn, and More. Below these are 'Your Work' and 'RECENTLY VIEWED' items. The main content area has a search bar at the top. Below the search bar is the 'Datasets' section with a description: 'Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.' There are buttons for '+ New Dataset' and 'Your Work'. To the right is an illustration of a person looking at data visualizations. Below this is another search bar and a 'Filters' button. A row of category tags includes 'All datasets', 'Computer Science', 'Education', 'Classification', 'Computer Vision', 'NLP', 'Data Visualization', and 'Pre-Trained Model'. The 'Trending Datasets' section is highlighted with a star icon. It shows four dataset cards: 'Mental Disorders Identification (Reddit)' by Hiraishin, 'Taylor Swift Spotify Data' by ARTHUR BOARI, 'USA High Quality Made - Backpacks' by L. Farras Vijaya, and 'Diabetes, Hypertension and Stroke Prediction' by prosper chuks. Each card displays the dataset name, creator, update time, usability score, file size, and number of files. The first card has 18 votes, the second 17, the third 6, and the fourth 9.

Practice for the final project

| | | |
|---|--|--------------------|
|  | Predicting Credit Card Approvals devzohaib · Updated 16 days ago Usability 10.0 · 2 Files (other) · 10 kB | ▲ 14 Bronze ... |
|  | Coffee_Data_CoffeeReview Hanif Al Irsyad · Updated 18 days ago Usability 6.5 · 2 Files (CSV) · 2 MB | ▲ 20 Bronze ... |
|  | Chemicals That May Contribute to Disease The Devastator · Updated a month ago Usability 9.4 · 1 File (CSV) · 503 kB | ▲ 17 ... |
|  | Website Satisfaction Survey hayri · Updated 23 days ago Usability 10.0 · 1 File (CSV) · 3 kB | ▲ 16 Bronze ... |
|  | The Hunger Games Dataset The Devastator · Updated 8 days ago Usability 9.4 · 1 File (CSV) · 666 B | ▲ 12 ... |
|  | World Crime Index Ahmad Jalal Masood · Updated 17 days ago Usability 9.4 · 1 File (CSV) · 8 kB | ▲ 2 ... |
|  | MARVEL SNAP : Decks and cards Jean-Michel D. · Updated 10 days ago Usability 10.0 · 2 Files (CSV) · 775 kB | ▲ 11 Bronze ... |
|  | Personality classification Type Dataset MD Shariful Islam · Updated 2 days ago Usability 7.6 · 1 File (CSV) · 1 kB | ▲ 7 Bronze ... |

Practice for the final project

- 1. Identify a dataset you are interested in** (you cannot repeat next Monday!)
- 2. Select your dependent variable**
Try to explain your dependent variable using the independent variables in the dataset
Visualize some relationships using scatter plots
- 3. Explain your results in business terms**
What are the implications of your results? What would you recommend to do?
- 4. Prepare a 10 minutes presentation**

What should you include in the presentation?



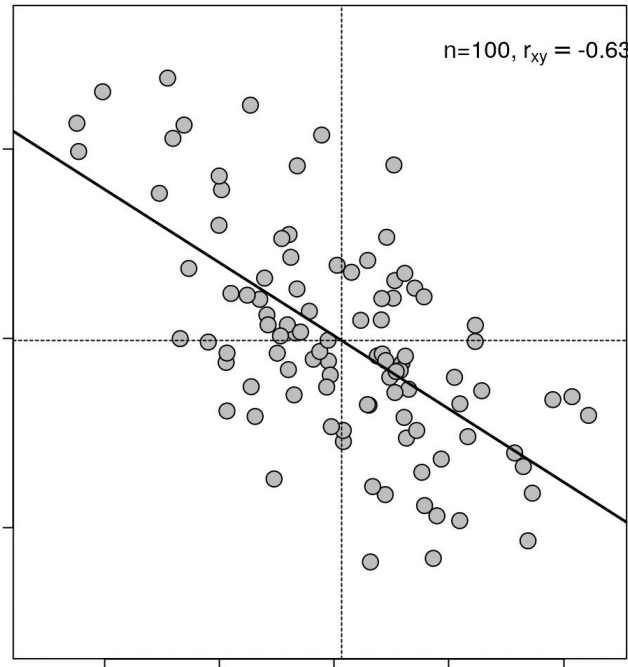
- 1. Context of your data (organization/company, problem to solve, why is this important, etc.)**
- 2. What variables do you have in the dataset?**
- 3. Show the multiple regression results in a table (tell us about what is and what is not significant, how big is the effect, etc.)**
- 4. Show some data visualization (using R)**
- 5. Business recommendations**

Practice for the final project



QUESTIONS?

Fundamentals of Econometrics Models



Vicenç Soler

`v.soler@tbs-education.org`

~~`vincent.soler@tbs-education.org`~~

