Statistics for Business and Economics I
Syllabus 2024–2025

Luc Hens

Instructor: Luc Hens (luc.hens@vub.be). Office: Pleinlaan 5, 5th floor, room 515 (PL5 5.15). If you mail me or the teaching assistant concerning this course, write in the subject line: Statistics I Yourgivenname YOURFAMILYNAME what-this-mail-is-about.

Teaching assistants: Charles Malvaux (charles.marie.malvaux@vub.be, room PL5 514); Neamat Sidani (neamat.sidani@vub.be, room PL5 511).

Learning platform: The Canvas page for this course contains the syllabus, links to software and data files, and more.

Office hours: If you need help about this course, first talk to me or to the teaching assistant during the break or after class. My office hours are on Fridays 14:00–16:45, by appointment only: `calendly.com/luc-hens`.

Lectures: To be announced. In the schedule, lectures are indicated by their abbreviation in Dutch: HOC for *hoorcollege*.

Tutorials: check your schedule. In the schedule, tutorials are indicated by their abbreviation in Dutch: WPO for *werkcolleges, practica, oefeningen*.

Course description

This course covers descriptive statistics (the art of summarizing data) and introduces the student to inferential statistics (the art of using sample data to make numerical conjectures about problems involving a population). We learn how to display data in tables and charts (frequency tables, histograms, boxplots, time series plots, . . .); how to describe the shape of data distributions and summarize their center and spread (mean, median, standard deviation, interquartile range, . . .); and how to display, summarize and interpret the relationship between two variables (scatter plots, correlation, line of best fit). We then learn about the basic elements of probability needed for statistical inference (probability rules, joint and conditional probability, contingency tables, probability trees,. . .); about random variables (expected value and standard error; probability models). Finally, we study the sampling distributions of proportions and means and use the Central Limit Theorem to construct confidence intervals for population proportions and population means.

This course counts for 6 ECTS credits. One ECTS credit represents 25 to 30 hours of work, so the workload for a 6-credit course is 150 to 180 hours, or —spread over 15 weeks— about 11 hours per week. This includes the class meetings (two hours per week of lecture time and two hours of tutorials), so I expect you to work for this course about seven hours per week outside class.

Course prerequisites

This course has no college-level prerequisites. Math prerequisites do not go beyond the elementary algebra covered in Jacques (2009, or a more recent edition, Chapter 1: Linear Equations, pp. 13–92): graphs of linear equations (section 1.1); algebra (section 1.4); transposition of formulae (section 1.5) (check "The elementary math needed for STA101" on the Canvas page for this course). You should not take this course if you don't master these concepts and skills.

Learning objectives

This course aims at providing you with an understanding of descriptive statistics (displaying and describing data) and inferential statistics (making valid generalizations from sample data). At the end of the course, you are able to:

- distinguish between categorical and quantitative data;

- display and describe categorical and quantitative data;

- compute and interpret a coefficient of correlation and the line of best fit;

- apply the rules of probability;

- work with random variables and probability models (binomial model, normal model, . . .);

- explain the properties of the sampling distribution of a proportion or a mean (and the conditions under which those properties hold);

- find (if appropriate) a confidence interval for a proportion or a mean using data from a large sample, and interpret the meaning of the confidence interval;

- explain the limitations of statistical methods;

- use statistical software and a scientific calculator to do statistical computations (enter data, generate descriptive statistics and graphs, compute confidence intervals);

- communicate the results of statistical work; more specifically, write up the results of statistical analysis in a report consisting of a non-technical abstract aimed at decision makers, so that they can improve their decisions, and a main section aimed at peers explaining the technical details and exact interpretation of the results. The report is formatted in APA Style.

Course materials

Textbook The required textbook is Sharpe et al. (2021). Make sure you get the paperback Global Edition (ISBN 978-1-292-26931-3), which is considerably cheaper (about 70 euro) than the hardback US edition. Available from the Standaard Student Shop (B building, ground floor) and online (`https://www.amazon.de/-/en/Tyler-Cowen/dp/1319383033/`; `https://www.bol.com/be/nl/p/modern-principles-of-economics/9300000024574590/`). If you use a previous edition it is your own responsability to verify where it differs from the current edition. Appendix B has a correspondence table between the previous and the current edition.

Calculator For the tutorials and exams you will need a Texas Instruments **TI-84** calculator. I recommend the TI-84 Plus CE-T, but other versions of the TI-84 are fine, too. You will also use the TI-84 in Statistics II. The statistical functions of the TI-84 that we'll use in Statistics I are on the formula sheet. Here are some sites where you can buy a TI-84 CE-T (about 140 euro: `https://www.coolblue.be/nl/product/865206/texas-instruments-ti-84-ce-t-python-app.html`; `https://www.amazon.de/-/en/Instruments-Python-Graphic-Calculator-Display/dp/B08F3PBLCL`; `https://www.dreamland.be/e/nl/dl/texas-instruments-rekenmachine-ti-84-plus-ce-t-python-972868`. Keep the receipt and the original packaging. That makes it easier to resell the calculator later. If you already have a **TI-Nspire** and what to use it for Statistics I, it is up to you to figure out how to find the statistical functions.

Statistical software A spreadsheet program is useful to do simple calculations and keep track of lists, but not to do serious statistical work. Practitioners of statistics and data scientists use specialized statistical software such as R, SPSS, or Stata, or a general-purpose programming language such as Python. In this course I will introduce you to the statistical environment R. R is open-source, free, and is (together with Python) the most widely used programming environment in data science. You will continue using R in Statistics II, Econometrics, and your research projects. Appendix A is a step-by-step guide to download, install, and run R in RStudio, an integrated development environment (IDE) for R, on your computer.

Data sets Because any statistical software and any text editor on any computer can read plain text files, plain text files are the preferred format to store data. The data sets used in the textbook are here: `https://dasl.datadescription.com/datafiles/`. I also posted them as a compressed (.zip) file on Canvas, which you can download and then unpack by double-clicking the zip file. Store the data files on the hard drive of your computer.

Additional Materials

Practice is important to learn statistics. The answers to the odd-numbered end-of-chapter problems of Sharpe et al. (2015) are in the back of the book. Students who wish to work additional exercises can find hundreds of solved

exercises in Kazmier (2004) (or a more recent edition). Available in the VUB library, call number: 311.17 G KAZM 2004.

Dalgaard (2008) is a helpful guide to the R Statistical Environment. Available in the VUB library, call number: 004.9 G DALG 2008. You can do statistical calculations with RStudio or with a statistical calculator like the TI-84, but also with the "knowledge engine" WolframAlpha (`wolframalpha.com`). This document explains how to do statistical calculations in WolframAlpha: `https://luc-hens.github.io/statistics-in-wolfram-alpha.html`.

Course assessment

The course is assessed with a written exam that consists partially or wholly of multiple-choice questions. We correct for guessing using a higher cut-off; that means that, for an exam consisting of 20 multiple choice questions with four answer options, you have to answer at least 12 questions correctly to get a passing grade. You should be able to use the TI-84 calculator to do statistical calculations.

As for multiple-choice questions, remember that to find the right answer you often need to explicitly write out your work in long form like we taught you to do in the tutorials. The exam booklet has sufficient space to do so on the back of the previous page. Start by writing down the givens. Then write down what is asked. Then write down your work, step-by-step, with brief declarative sentences explaining key steps. You do this for yourself (not for us: we won't read it), but it's essential to find the right answer and review your work at the end of the exam.

Bring the following things to the examination: your student ID, a mechanical pencil (soft lead: HB no. 2), an eraser, a blue or black pen, some colored pencils (not red), a ruler with a centimeter scale, the TI-84 calculator, and the laminated formula sheet. Make sure the batteries of the calculator are charged (TI-84 CE-T) or bring spare batteries (TI-84 Plus). Clear the memory of the TI-84 before coming to the exam. Put everything (except the formula sheet) in a 1 liter transparent plastic resealable (*Ziploc*) bag. You can take a small bottle of water if you remove the label (a re-usable bottle is better). **No** pencil case, smartphone, smartwatch, paper, or food. Switch off your smartphone and put it in your bag before the exam: you are not allowed to take a smartphone or smartwatch to your desk. Walk in silence to the seat that is assigned to you by one of the proctors, store your bag and coat below your seat and out of reach. Wait until everyone is seated before starting the exam.

Don't mail or call us to inquire about your grade—we don't communicate grades by e-mail or by telephone. The administration communicates the grades.

Preparation for class

This course uses a mix of lectures and in-class problems solved during the tutorials. Class attendance is crucial: statistics is a difficult and sequential subject, and students missing several classes or tutorials rarely pass this course. Attendance and active participation in the tutorials is compulsory. During the tutorials, concepts are reviewed, you can ask questions, and you get to practice

problems with immediate feedback. Consult the learning platform regularly (at least once a week). Carefully read the materials indicated in the course schedule before coming to class. Statistics is a sequential subject: new topics build on concepts introduced before, so it is crucial to keep up with the material as we go along and to regularly review concepts. During the tutorials you will work problems, sometimes individually, sometimes in small teams. I expect you to actively work the problems, and be prepared to briefly present the results of your work to the other students. **Bring to class**: a mechanical pencil (with soft leads: HB no. 2); an eraser; a ruler with a centimeter scale; and A4-sized paper with 5 mm squares (notebooks of the *Atoma* brand allow you to easily add, remove, and reorganize pages). To the **tutorials** you should bring the above plus: the textbook (because we'll work end-of-chapter exercises); the laminated formula sheet; the TI-84 calculator; and your laptop camputer with R and RStudio installed and the data sets downloaded. You are not allowed to use tablets or smartphones in class, and you are only allowed to use a laptops to use R in selected problems: researchers have shown that they hinder classroom learning for both users and nearby peers (Sana et al., 2013). Take notes during the lectures, on paper.

Academic Honesty

Academic dishonesty is not tolerated. I will communicate every case of cheating, plagiarism, or other forms of academic dishonesty cases to the Dean. If you refer to other work (someone else's or your own), provide appropriate references and citations.

Course schedule

Chapter numbers refer to Sharpe et al. (2021) unless indicated otherwise. Appendix B shows the correspondence between Sharpe et al. (2015) and Sharpe et al. (2021). Consult the learning platform before every class meeting for announcements and possible schedule changes. Review the chapters that we already covered before coming to class. Work the listed exercises after we covered the chapter. I strongly encourage you to practice statistics by working additional exercises at home, individually or in group. The back of the textbook has solutions to all odd-numbered exercises. Your teaching assistant will be glad to help you when you get stuck (but always bring your written preparation).

Week 22 (starting Monday 10 February 2025)

How the course is organized. Read Joseph Stromberg (2014, August 21). Why you should take notes by hand—not on a laptop. *Vox* (`http://www.vox.com/2014/6/4/5776804/note-taking-by-hand-versus-laptop`); Cindi May (2017, 11 July). Students are better off without a laptop in the classroom. *Scientific American* (`https://www.scientificamerican.com/article/students-are-better-off-without-a-laptop-in-the-classroom/`). Do the following before the first tutorial (WPO): Buy the textbook. Install R and RStudio on your laptop. Go to the textbook web site and download the data sets in text

format to your laptop (the link is on Canvas). Buy the Texas Instruments **TI-84 Plus CE-T** calculator. Download the **formula sheet** from the learning platform and print it or put the pdf on a USB-stick. Go to a copy shop and have the formula sheet printed on thick paper; write your name at the top right and have it laminated. Always bring the formula sheet to the tutorials. You can use the formula sheet during the exam, but only when it is laminated.

Module 1: Data and decisions. Ch. 1. Exercises 2, 4, 6, 12, 16, **19**, **23**, **25**.

Module 2: Displaying and describing categorical data. Ch. 2. Exercises 2, 4, 6, 8, 10, 28, 40, **46**.

Tutorial: Bring your laptop with R and RStudio installed and the datasets on the drive. Make sure uout laptop is charged. During the tutorial you will get an introduction to R and RStudio.

Week 23 (starting Monday 17 February 2025)

Module 2: Displaying and describing categorical data, continued.

Week 24 (starting Monday 24 February 2025)

Module 3: Displaying and describing quantitative data. Ch. 3. Skip "Transforming skewed data" (pp. 111–113). Density histogram: see your class notes (not covered in the textbook). Exercises 2, 4, 6, 8, 12, 14, 18, 23, **30**.

Week 25 (starting Monday 3 March 2025)

Module 3: Displaying and describing quantitative data, continued. Exercises **42**, **48**. The exercises require RStudio so bring your laptop to the tutorial.

Week 26 (starting Monday 10 March 2025)

Module 4: Correlation and regression. Ch. 4. Skip sections 4.7, 4.8, 4.10, 4.11. Exercises **2**, 4, 6, 10, 14 (use the information from question 12 that $r = 0.988$; first without the statistical functions of the TI-84; then check the results using the `LinReg` function of the TI-84), **20**, 24, 26, 30, 32, **60**, **70** (use the `LinReg` function of the TI-84 and RStudio). The exercises require RStudio so bring your laptop to the tutorial.

Week 27 (starting Monday 17 March 2025)

Module 5: Randomness and probability. Ch. 5. Skip section 5.9 (Bayes' rule). Exercises 2, 4, 6, 8, 10, **12**, 14, 56, 58.

Week 28 (starting Monday 24 March 2025)

Module 5: Randomness and probability, continued.

Week 29 (starting Monday 31 March 2025)

Module 6: Random variables and probability models. Ch. 6. Skip "The Geometric Model" (p. 234) and "The Poisson Model" (pp. 238–239). Exercises 1, 2, 4, 6, 10, 14, **16**, 18, 26, 42.

Week 30 (starting Monday 7 April 2025)

Spring break. No class.

Week 31 (starting Monday 14 April 2025)

Spring break. No class.

Week 32 (starting Monday 21 April 2025)

Monday 22 April 2024 is a legal holiday. No class.

Module 6: Random variables and probability models, continued.

Week 33 (starting Monday 28 April 2025)

Thursday 1 May 2024 is a legal holiday. No class.

Module 7: The normal distribution. Ch. 7. Skip section 7.6. Do read the "What can go wrong" box at the bottom of p. 272—it does not belong to section 7.6). Exercises **2**, 4, 6, **10**, 12, 18, **28** (use the 68-95-99.7 rule), **30** (use the 68-95-99.7 rule), 32, 34, 36, **40**, 42, 60.

Week 34 (starting Monday 5 May 2025)

Module 7: The normal distribution, continued. Homework assignment: for the problems assigned for chapter 7 that required the TI-84, use RStudio replicate the results you obtained with the TI-84. Print out the console window with the results and bring it to class.

Module 8: Observational studies, surveys, and experiments. Ch. 8; Ch. 9 (only sections 9.1 and 9.2 pp. 312-318, 335) (I will post these pages on Canvas for students who use the third edition of the textbook). Watch "Leading questions," a fragment from *Yes Prime Minister*, series 1 episode 2 (The Ministerial Broadcast), BBC, 1986 (`https://youtu.be/G0ZZJXw4MTA`) (video and transcript of the dialogue). Exercises for Ch. 8: 2, 4, 12, 16, 28, 30, 48. Exercises for Ch. 9: **2**, **4**.

Week 35 (starting Monday 12 May 2025)

Module 9: Sampling distribution of a proportion and confidence interval for a proportion. Ch. 9. Exercises 2, 4, 8, 10, 12; 11 & 14 (they belong together), 16, 18 (use in part *b* 99.7% instead of 99%), 48, 50, **58**. For

all exercises that involve computing a confidence interval, first work the exercises without using the 1-PropZInt function of the TI-84; then check that you get (approximately) the same result using the 1-PropZInt function.

Week 36 (starting Monday 19 May 2025)

Module 10: Sampling distribution of a mean and confidence interval for a mean when the sample is large. Ch. 11. Only sections 11.1 to 11.3 and "Cautions about Interpreting Confidence Intervals" (p. 393). We limit ourselves to those cases when the sample is sufficiently large to use the normal approximation. In such cases, the 95% confidence interval for a mean is approximately

$$\bar{y} \pm 2 \times SE(\bar{y})$$

Exercises **4**, **6**, **8** & **14** (for part (a), find the 95% confidence interval and assume that you can use the normal approximation; skip part (c)), **26**.

Week 37 (starting Monday 26 May 2025)

Study week.

Weeks 38–42: Final exams

Written examination (date and room to be announced). See the instructions for the examination above.

Appendix A Installing R and RStudio

Installing R and RStudio (Windows)

Point a web browser to the download page of *Posit*, the firm that makes RStudio:

    `https://posit.co/download/rstudio-desktop/`

At he left-hand side you see: "1: Install R." Click the button "DOWNLOAD AND INSTALL R." The browser opens a new tab with the download page of R (*The Comprehensive R Archive Network, CRAN)*. Select "Download R for Windows." Click on "Download R 4.2.2 for Windows" (the version number may be more recent than 4.2.2). Your browser starts to download the file R-4.2.2-win.exe (the version number may be more recent) to your computer, probably to the Downloads directory. Once downloaded, double-click the file to install.

Then point your browser back to RStudio Desktop page:

    `https://posit.co/download/rstudio-desktop/`

On the right-hand side you see: "2. Install RStudio." Click the button "DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS." Your browser starts to download RStudio-2022.12.0-353.exe (the version number can be more recent than 2022.12.0-353) to your computer, probably to your Downloads directory. Once downloaded, double-click the file to install.

Installing R and RStudio (macOS)

First check which chip your Mac has (older Macs have an Intel chip, newer Macs have an Apple M chip) by going to the apple icon in the menu bar (top left) and select: About This Mac > Chip. Write down whether your Mac has an Intel or an Apple M chip. You will need this information in a moment.

Point a web browser to the download page of *Posit*, the firm that makes RStudio:

    https://posit.co/download/rstudio-desktop/

At he left-hand side you see: "1: Install R." Click the button "DOWNLOAD AND INSTALL R." The browser opens a new tab with the download page of R (*The Comprehensive R Archive Network, CRAN*). Select "Download R for macOS." If your Mac has an Apple M chip click on "R-4.2.2-arm64.pkg" (the version number can be more recent than 4.2.2); If your Mac has an Intel chip click on "R-4.2.2.pkg" (the version number can be more recent than 4.2.2). Your browser starts to download the file R-4.2.2.dmg (the version number may be more recent) to your computer, probably to the Downloads directory. Once downloaded, double-click the file to install. The installer will put R.app in the Applications folder.

Then go in browser back to the tab with the RStudio downloads page:

    https://rstudio.com/products/rstudio/download/

On the right-hand side you see: "2. Install RStudio." Click the button "DOWNLOAD RSTUDIO DESKTOP FOR MAC." Your browser starts to download RStudio-2022.12.0-353.dmg (the version number can be more recent than 2022.12.0-353) to your computer, probably to the Downloads directory. Once downloaded, double-click the file to install. Once downloaded, double-click the .dmg file to install. A window "RStudio-2022.12.0-353" opens (the version number can be more recent). In the left side of the window there's an alias of the Applications folder; in the right side there's an icon that says RStudio.app. Drag the RStudio.app icon to the alias of the Applications folder. The installer will now put RStudio.app in the Applications folder. Go to the Applications folder and drag the RStudio icon to the Toolbar to have easy access to RStudio. Go to Finder and open a new Finder window (File > New Finder Window). Scroll down the left panel until you see, under "Locations": RStudio-2022.12.0-353 (the version number can be more recent). Click on the eject button.

Running R in RStudio

After you installed R and RStudio on your PC or Mac, locate the RStudio icon in your Applications folder or Toolbar in macOS or where your applications are listed in Windows. Click the RStudio icon. RStudio starts up and shows a window very much like figure 1.

The pane at the left is called Console. In the Console pane some text appears: the version number of R and some information about the program, followed by a prompt (>) and a blinking cursor, waiting for your input. After the prompt, type
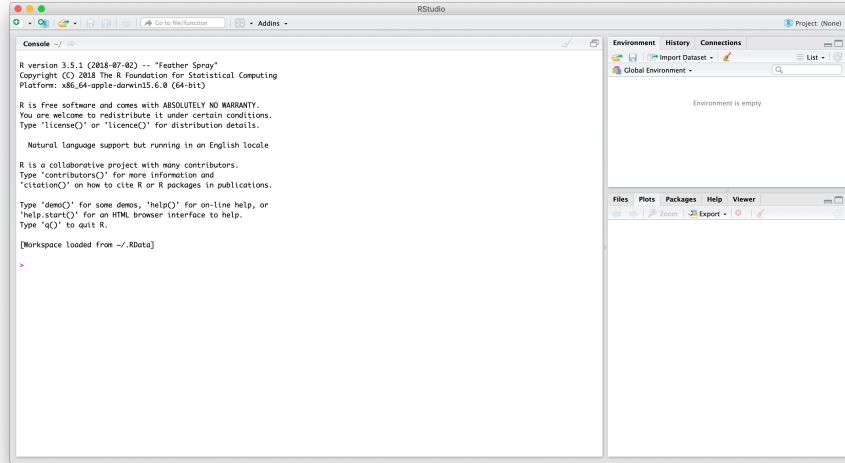
```
1 + 2
```

Figure 1: RStudio window (in macOS)

and press the return key. The Console pane will show the result:

```
[1] 3
```

The result is preceded by `[1]` to indicate that this is the first line of the result—some results run over more than one line.

You are ready to crunch numbers! You can now remove the .dmg, .pkg, or .exe files you downloaded.

Appendix B Correspondence between Sharpe et al. (2015, 2021)

Sharpe et al. (2015)	Sharpe et al. (2021)
Chapter 1	Chapter 1
Chapter 2	Chapter 2
Chapter 3	Chapter 3
Chapter 4	Chapter 4
Chapter 5	Chapter 5
Chapter 6	Chapter 6
Chapter 7	Chapter 7
Chapter 8	Chapter 8 + new Chapter 9 (Experiments)
Chapter 9	Chapter 10
Chapter 11	Chapter 11

The exercises in the schedule above are for Sharpe et al. (2021) and largely correspond to Sharpe et al. (2015). Sometimes the number of the exercise is

somewhat different or the data have been updated. Exercises that are new in Sharpe et al. (2021) are indicated in **bold**.

Appendix C   Errata in Sharpe et al. (2021)

Missing units in tables.   The units (\$) for the variable **Price** are missing in table 1.1 on p. 37. The values in column 4 should be: \$5.99, \$9.99, \$9.99, \$10.99, \$11.99. Alternatively, report units in the table header: **Price (\$)**.

Missing units in calculations.   Always carry the **units of measurement** in a calculation. That way the result will show the correct units of measurement. E.g., the correct calculation of the z-score of the \$340,000 house on p. 101 is:

$$z = \frac{\$\,340\,000 - \$\,175\,000}{\$\,55\,000} = 3.00$$

Now it is clear that the units of the numerator and the denominator are the same (\$), and that the result of the fraction has no units. In the example on p. 150 (top) units are missing, too. The equation should be:

$$\hat{\text{Price}} \approx \$\,13\,439 + 113.12\ \$/\text{sq.ft} \times (\text{Living Area in sq. ft})$$

The equation on p. 151 (top) should be:

$$\hat{\text{Price}} \approx \$\,13\,439 + 113.12\ \$/\text{sq.ft} \times 3000\ \text{sq.ft} \approx \$\,352\,799$$

Note that the units of the outcome now result from the computation. Moreover, the interpretation of the coefficient is much clearer now that you can see which units it has.

Rounding and approximations.   Use $\approx$ when rounding or approximating numbers, not $=$. See the example in the previous paragraph.

Sharpe et al. (2021) wrongly omit units elsewhere, and wrongly use the equality sign ($=$) when they should use $\approx$.

References

Dalgaard, P. (2008). *Introductory Statistics with R*. Springer, Berlin, 2nd edition.

Jacques, I. (2009).   *Mathematics for Economics and Business*.   Financial Times/Prentice Hall, London, 6th edition.

Kazmier, L. J. (2004). *Schaum's Outline of Theory and Problems of Business Statistics*. Schaum's Outline Series. McGraw-Hill, New York, 4th edition.

Sana, F., Weston, T., and Cepeda, N. J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, 62:24–31.

Sharpe, N. R., De Veaux, R., and Velleman, P. (2015). *Business Statistics*. Pearson Education, 3rd edition.

Sharpe, N. R., De Veaux, R., and Velleman, P. (2021). *Business Statistics*. Pearson Education, 4th edition.