Statistics for Business and Economics I
Syllabus 2020–2021

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 *Statistics I Yourfirstname Yourlastname* in the subject line.

Teaching assistant: Selin Tansu Tunç (selin.tansu.tunc@vub.be). Pleinlaan 5, 5th floor, room 520 (PL5 5.20).

Learning platform: `http://canvas.vub.be`. The Canvas page for this course contains the syllabus, the assignment for the research project, the rules for written work, 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 (HOC) and Tutorials (WPO): As long as we are in Code Red, lectures and tutorials are on **Tuesdays 14:00–16:00**, via on-line streaming in Microsoft Teams. Live sessions of **tutorials** are about every three weeks, on Tuesday 23 February and Tuesday 16 March 2021, and (if we are still in Code Red after the spring break) on Tuesday 27 April and Tuesday 18 May 2021. Send questions in advance, by mail or in the discussion forum on Canvas. You can ask questions during the live sessiins via chat, but usually we'll only be able to answer these questions afterwards. We'll post weekly video clips on Canvas.

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— 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 eight 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.

For the tutorials and the research project you need some basic information technology skills: searching for information on the internet; downloading files from the internet; using a word processor.

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. (2015). Make sure you get the paperback Global Edition (ISBN 978-1-292-05869-6), which is considerably cheaper (about 68 euro on amazon.de) than the hardback US edition. Available from the Standaard Student Shop (B building, ground floor) and online (amazon.de, bol.com).

Calculator The Texas Instruments **TI-84 Plus CE-T** calculator is required for this course (about 120 euro:
`https://www.coolblue.be/nl/product/865206/texas-instruments-ti-84-ce-t-python-app.html`;
`https://www.amazon.de/Texas-Instruments-84PLCE-TBL-2E5/dp/B01086MEM2`;
`https://www.dreamland.be/e/nl/dl/texas-instruments-rekenmachine-ti-84-plus-ce-t-3342721`)
The older TI-84 Plus will do as well, but the TI-84 Plus CE-T is better value for money. You have to bring the calculator to all tutorials and to the examination. Keep the receipt and the original packaging; that makes it easier to sell the calculator next year.

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 with the RStudio front-end. R is open-source, free, and one of the most widely used tools in data science. You will continue using R and RStudio in Statistics II, Econometrics, and your research projects. R and RStudio are installed on the computers in the computer labs at the ground floor of B building (rooms B003 through B006). Here is a step-by-step guide to download, install, and run R and RStudio on your computer:

Installing R and RStudio (Windows) Point your browser to the R Project Homepage:

`http://www.r-project.org`

Follow the link "Download" > CRAN. Select a CRAN Mirror in Belgium (it doesn't matter which one). Select "Download R for Windows." Click on "base" or "install R for the first time." Click on "Download R 4.0.3 for Windows" (the version number may be more recent than 1.4.1103). Your browser starts to download the file R-4.0.3-win.exe (the version number may be more recent than 4.0.3) to your computer, probably to the Downloads directory. Once downloaded, double-click the file to install.

Then point your browser to the RStudio downloads page:

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

The first column shows

RStudio Desktop Open source license

Click on the "DOWNLOAD" button. Then click on the button "Download RStudio Desktop." Your browser starts to download RStudio-1.4.1103.exe (the version number can be more recent than 1.4.1103) to your computer, probably to your Downloads directory. Once downloaded, double-click the file to install.

Installing R and RStudio (macOS) Point your browser to the R Project Homepage:

`http://www.r-project.org`

Follow the link "Download" > CRAN. Select a CRAN Mirror in Belgium (it doesn't matter which one). Select "Download R for (Mac) OS X." Click on the link R-4.0.3.pkg (the version number may be more recent than 4.0.3). Your browser starts to download R-4.0.3.pkg (the version number may be more recent than 4.0.3) to your computer, probably to your Downloads folder. Once downloaded, double-click the package to install. The installer will put R.app in the Applications folder.

Then point your browser to the RStudio downloads page:

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

The first column shows

RStudio Desktop Open source license

Click on the "DOWNLOAD" button. Then click on the button "Download RStudio Desktop." Your browser starts to download RStudio-1.4.1103.dmg (the version number can be more recent than 1.4.1103) 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-1.4.1103" opens. 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-1.4.1103. 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

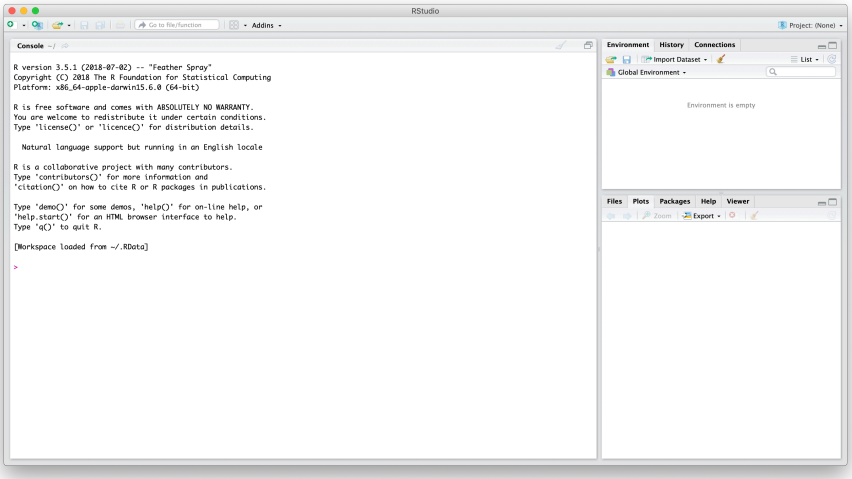

Figure 1: RStudio window (in macOS)

```
1 + 2
```

and press the return key. The Console pane will show the result. 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, and .exe files you downloaded.

Data sets Download the **data sets** from the textbook web site: `http://wps.pearsoned.co.uk/ema_ge_sharpe_business_3/250/64017/16388553.cw/index.html`. Follow the link Data Sets > Text Data Sets > Download all chapters as a .zip file (the direct link is `http://media.pearsoncmg.com/intl/ema/ema_ge_sharpe_business_3/data_sets/sbs3e_datasets_text.html`; also posted on the Canvas page). Double-click the .zip-file to unpack it. 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. Store the data files on the hard drive of your computer.

Word processor For the assignment you need access a computer that runs with a word processor. Every VUB student can download Microsoft Office and has access to Office 365 (`https://student.vub.be/en`).

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: `http://users.telenet.be/luc-hens/statistics-in-wolfram-alpha.html`.

Course assessment

This course uses a mix of lectures and in-class problems solved during the tutorials. Students are assessed on basis of a short research paper (in group) and a written exam. The final grade for the course will be calculated on the following weighting:

Paper	10 %
Examination	90 %

Percent means per hundred: 10% is equivalent to $10/100 = 0.10$, and 90% is equivalent to $90/100 = 0.90$. To compute your result (on a scale of 0 to 20) do:

$$\begin{array}{rl} & 0.10 \times \text{(result for the paper on a scale of 0 to 20)} \\ + & 0.90 \times \text{(result for the examination on a scale of 0 to 20)} \end{array}$$

Paper. You cannot redo the paper for the second examination session; in the second examination session you keep the grade for the paper obtained during the semester. The assignment for the paper is posted on the learning platform, with a detailed description of the criteria used to assess your work (the rubric). The paper should be submitted as a portable document format (pdf) file on the learning platform no later than the date specified in the course schedule. In accordance with Murphy's Law, computers are likely to crash the evening before the due date. Start in time. Make backups of your work. Computer problems are no valid reason for handing in work late. Late papers are allowed only in emergency cases, which must be documented by a physician or college official, in advance when possible. Otherwise the penalty for late work is 2 points (on a scale from 0 to 20) per calendar day late.

Examination. The examination usually consists of two parts. In the first part you have to solve one or more problems similar to the ones in the textbook (Sharpe et al., 2015); the second part is a set of multiple-choice questions. When solving the problems for the exam, you have to show that you understand which assumptions were made when developing the techniques you use, and what the limitations of the techniques are. You should be able to use the TI-84 calculator to do statistical calculations. To grade the multiple-choice questions we correct for guessing using the standard setting: that means that, for questions with five possible answers, a correct answer yields 1 point; no answer yields 0 points; and a wrong answer is penalized by $-¼$ of a point.

Bring the following things to the examination: your student ID, a mechanical pencil (soft lead: HB no. 2), an eraser, a pen, a ruler with a centimeter scale, the TI-84 calculator, and the laminated formula sheet. Make sure the batteries are charged (TI-84 CE-T) or bring spare batteries (TI-84 Plus). Put everything (except the formula sheet) in a 1 liter transparent plastic resealable (*Ziploc*) bag. **No** pencil case, smartphone, smartwatch, paper, tissue, or food. You can take a small bottle of water if you remove the label (a re-usable bottle is better). 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. Before you go to your seat, put your bags and coats away, take the Ziploc bag and walk in silence to the seats assigned by the proctors. 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

Class attendance is crucial: statistics is a difficult and sequential subject, and students missing several classes or tutorials rarely pass this course. 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 (if the course schedule below says so) your laptop with RStudio installed and the data sets downloaded. You are not allowed to use laptops, tablets, or smartphones in class: 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. (2015). 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).

Note concerning the coronavirus measures: The schedule below is for a normal semester. Due to the coronavirus measures (Code Red), live sessions for the tutorials are only once every three weeks (see p. 1). We post the videos on Canvas according to the schedule below. We recommend that you follow the schedule when studying the material.

Week 22 (starting Monday 8 February 2021)

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/`).

Data and decisions. Ch. 1. Exercises 2, 4, 6, 12, 16, 20, 26, 28, 32.

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

Homework assignment: 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.

Week 23 (starting Monday 15 February 2021)

Displaying and describing categorical data. Ch. 2, continued.

Introduction to R and RStudio. Bring your laptop computer to the tutorial. Make sure R and RStudio are installed on your laptop computer, and that you have downloaded the textbook data files (in text format).

Week 24 (starting Monday 22 February 2021)

Displaying and describing quantitative data. Ch. 3. Skip "Stem-and-leaf Displays" (pp. 80–81) and "Transforming skewed data" (pp. 100–102). Density histogram: see your class notes (not covered in the textbook). Exercises 2 (skip *d*), 4, 6, 8, 12, 14, 18, 26, 30.

Week 25 (starting Monday 1 March 2021)

Displaying and describing quantitative data. Continued. Exercises 42, 50. The exercises require RStudio so bring your laptop to the tutorial.

Week 26 (starting Monday 8 March 2021)

Correlation and Regression. Ch. 4. Skip sections 4.7, 4.8, 4.10, 4.11. Exercises 2, 4, 6, 9, 19, 14 (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, 68 (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 15 March 2021)

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 22 March 2021)

Randomness and probability. Continued.

Week 29 (starting Monday 29 March 2021)

Deadline: Submit the paper as a portable document format (pdf) file on the learning platform under 'Assignments' by Monday 29 March 2021 at midnight.

Random variables and probability models. Ch. 6. Skip "The Geometric Model" (pp. 220–221) and "The Poisson Model" (pp. 224–225). Exercises 1, 2, 4, 6, 10, 14, 16, 18; 26, 42.

Week 30 (starting Monday 5 April 2021)

Spring break—no class

Week 31 (starting Monday 12 April 2021)

Spring break—no class

Week 32 (starting Monday 19 April 2021)

Random variables and probability models Continued.

Week 33 (starting Monday 26 April 2021)

The normal distribution. Ch. 7. Skip section 7.6 (but do read the "What can go wrong" box at the bottom of p. 258—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 3 May 2021)

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.

Surveys and sampling. Ch. 8. 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 2, 4, 10, 14, 26, 28, 46.

Week 35 (starting Monday 10 May 2021)

Sampling distribution of a proportion and confidence interval for a proportion. Ch. 9. Exercises 2, 4, 8, 10, 12; 11 and 14 (they belong together), 16, 18 (in part *b,* use 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 the result using the 1-PropZInt function).

Week 36 (starting Monday 17 May 2021)

Sampling distribution of a mean and confidence interval for a mean when the sample is large. Ch. 11. A brief introduction. We limit ourselves to those cases when the sample is sufficiently large. Read sections 11.1 to 11.3, and "Cautions about Interpreting Confidence Intervals" (pp. 375–376). Skip the remainder of the chapter. For large samples only, the 95% confidence interval for a mean is approximately

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

Exercises 6, 28, 48 (adapted version — posted on Canvas).

Week 37 (starting Monday 24 May 2021)

Study week.

Weeks 38–42 (starting Monday 31 May 2021)—Final exams

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

Errata in Sharpe et al. (2015)

Missing units. The units ($) for the variable **Price** are missing in table 1.1 on p. 31. The values in column 4 should be: $5.99, $9.99, $9.99, $10.99, $11.99.

Missing units in calculations. Always carry the **units of measurement** in a calculation. That way the result also shows the units of measurement. E.g., the correct calculation of the z-score for the $\$\,3.5$ M house on p. 90 is:

$$z = \frac{\$\,3{,}500{,}000 - \$\,1{,}930{,}400}{\$\,914{,}500} \approx 1.72$$

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. Also use $\approx$ when rounding numbers, not $=$.

The example on p. 131 should be:

$$\bar{x} = 2004, s_x = 3.89$$

$$\bar{y} = 37.85 \text{ years}, s_y = 3.26 \text{ years}$$

The sum of the cross product of the deviations is found as follows:

$$\sum (x - \bar{x})(x - \bar{y}) \approx 147 \text{ years}$$

Putting the sum of the cross products in the numerator and $(n-1) \times s_x \times s_y$ in the denominator, we get

$$\frac{147 \text{ years}}{(13-1) \times 3.89 \times 3.26 \text{ years}} \approx 0.96$$

Sharpe et al. (2015) 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.