Practical Econometrics 2

Vygantas Butkus

Introductory Notes

This is not a textbook and it is not intend to be. Rather, it is short summary that emphases messages of lectures, gives corresponding references and reminds tasks (if any). During the course this note will be updated regularly.

Students assesments

The final score are going to be estimated according a formula given below. Give your self a time to understand it:

```
round(sum(min(6, sum(tasks, midExam, extraPoints)), goldenPoints, finalExam))
```

The variables are explained in more details:

- **tasks**: during the course you are going to have quite a number of tasks. I believe this is the main part of the course. The maximum sum of points is 4, but please note, that huge (unlimited) number of extra points can be received, if you demonstrate extra efforts in the tasks. In addition, please note that the **tasks** variable can be positive only if you supplied all the tasks that was required. If at least on of obligatory tasks are missing, the sum of total points are going to be 0. Thus, do all the tasks, event if the deadline is far over.
- **midExam**: in the middle of semester you are going to have mid exam. The possible outcome is 2 points.
- **extraPoints**: extra points are going to serve as system of encouraging. You can get extra points in many situations: taking challenges, answering unexpected question in lecture, giving good question and so on. Please note, that **extraPoints** can be negative as well. This could happen in the case of cheating (or something).
- **goldenPoints**: golden points are very similar to extra points, but to get those, you will have to put extra efforts in it. How to get them will be steaded in course.
- **finalExam** after the semester you are going to have final exam. The total points of the exam is 4.

Cultural aspects of the course:

1. During the course students are encourage to collaborate with each other. You can freely share thought, ideas and event source codes. In addition students can use all the material that can be found in the Internet or textbooks. Using the materials made by others are absolutely legal and have no negative side effect on your assesment. But there are two conditions that are strictly mandatory:

- You have to give all the credits to original source. Please write an explicit note with links and give some comments.
- You have to make sure you understand all the details of the code (or ideas) that you are using. You can be asked to comment it in great details. So, please use the material that you fully understand.

2. I am promise to do my best in answering all your question. You can ask freely during the lectures. Nevertheless, in practical assignments, your tasks is to try to find the answer by your self. So, in practical tasks, before asking any question please be prepare to report that efforts you have made to get the answer? For example: what has the Google query that you have used? That function you have already tried to apply?

3. Reading and assessing other students rapports is an important part of your learning. Please be honest to your colleagues and give the best insights. On the other hand, you should appreciate all the criticism that you get and accept it as important part of your learning.

4. Any initiative and communication is highly encouraged. Thus, If you have any ideas how to make course more productive or interesting, please let me know.

5. The tasks are always open for deeper analysis. If you have idea how to reveal topic in more detail please do it - it probably will get you some extra points. On the other hand, if you are not satisfied with the assignment that you get, you are free to give your reasons and negotiate.

Outline of the course

1. A short intro of the stuff that you already should know (at least most of it):

- Technical tools: R (RStudio), Rmd files, Git
- The basic properties and concepts in statistics
- Linear regression (with some extensions)

2. Linear regression in the content of time series

- Basics (concepts, the difference with classical regression)
- Trends (stochastic vs deterministic)
- Spurious regression
- Autoregressive model
- Transformations

3. Intro in time series

- Basics: stationarity, autocorrelation, AIF, PAIF
- The decomposition of time series
- Some common methods: exponential smoothing, moving average, filtering
- Overview of more advanced methods: HP, Kalman, State Space

4. ARIMA model

- AR(p)
- MA(q)
- RW
- ARIMA: specification, diagnostic, estimation.

5. Overview of other important models:

- Garch
- logit

Learning material

Selected textbooks that can be useful in this course could be found in `literature` folder. In addition, all references are given at the end o this document.

Moreover, some wonderful online material is available:

- Using R for Introductory Econometrics: http://www.urfie.net/
- Forecasting: principles and practice: https://www.otexts.org/fpp
- Introduction to Probability, Statistics, and Random Processes: https://www.probabilitycourse.com/

Chapter 1. Technical tools

Technical requirements

In the course you are going to use tree main software:

- R - environment for statistical computing and graphics
- RStudio - a set of integrated tools designed to help you be more productive with R.
- Git - distributed version control system

I assume you already know R and how to install it. If not check the links below:

- Installing R on Windows
- Installing R on Mac
- Installing R on Ubuntu (linux)

In addition to R, it's highly recommended that you install RStudio, which will make your experience with R much more enjoyable. If you need to install RStudio, you can do so here. Select the appropriate installer for your operating system.

Moreover, Git is also quite an important software. Git is a free and open source distributed version control system. It helps you not to get lost in versions of your project. In addition it is much easier to merge contributions among all team members and always have the newest version of the project. This is a system that professionals uses quite a lot (please admire it). Please visit Git home and install it in your system. Originally Git is command line program, on the other hand, Git might be used with graphical interface (GUI). Visit Gits guis page to select the one that looks best for you.

Task 1. Rmd in Github

The goal of your first task is to get familiar with R, Rmd files and Git.

1. Firstly, you have to sign up for Github account and then create a repository under the name `Econometrics`.

2. Go to the Google and find some awesome R features, packages, programs, graphs or something. Pick anything that you think is worth noting.

3. In `Econometrics` create a subdirectory `task1` and in this subdirectory create Rmd file with title `task1.Rmd`. Then write a very short summary about the topic you have choose. Please be polite and give the link to original material. Try to put some R code snippets and visualization that describes the main essence of your note. Try to customize something (if possible). And finally make sure that the file can be knited.

4. Commit it to Git and push to Github. Please note, that RStudio has small GUI for Git. It is avalible if you create a project (insted of plain Rmd file).

5. Write an email to me (Vygantas.Butkus@mii.vu.lt) with subject "[PE2][task1] Link to Github". And give the link to your Github account in the body.

Hint: A [short learning course on git](https://try.github.io/levels/1/challenges/1) could be useful. Please try it if you are not familial with Git.

Please note that those who demonstrate the greatest quality in their task will have actual opportunity to get their first extra points.

Task 2. Swirl course

Please visit swirl website. Swirl is a frameworks for interactive R learning. Your task is to complete R programming course that can be find in my GitHub accaunt.

For a first time, you have to install `swirl` package and the course material. Run R commands (be patient, this can take a while):

```r
install.packages('swirl')
library("swirl")
install_course_github("1vbutkus", "swirl", multi=TRUE)
```

To run a course you should type:

```r
library("swirl")
swirl()
```

Follow the instructions and select the "R Programming" course. Course contains 12 lessons. You have to do all of them. After Finishing each lesson, system will generate an email that you have to send to vygantas.butkus@mii.vu.lt (with exact subject and body content that will be generated by swirl).

Please note, that there are other courses. They are optional and can be really useful to increase knowledge and competition in R and data analysis. Please help your self.

Alternative. This task have an alternative. If you consider your self being advanced enough in R, then you can request a personal task. I believe this would be more interesting and more useful. Contact me personally if you consider doing a personal task.

Chapter 2. Probability and modeling

Literature: the main source is (Lavine 2013, Chapter 1 Probability). Alternative source is (Kerns 2011, ch. 4-6).

Key qualifications:

- understand main concepts of random variables (probability mass function, distribution function, quantile function, moments, covariance).
- know how to use R functions for classical random variables. Ex: dnorm, pnorm, qnorm, rnorm
- how to use model custom variables of the forms $f(x)$, where f is some known function, and x is some classical random variable.
- what is the principal of statistical tests?

- how to summarize modeling results and draw the conclusions.

Note: This chapter is about something that you already should know. So, please make sure you understand everything. If not, spend some time to catch up.

Planas:

- Ankstenes paskaitos neišpildyti akcentai:
- GIT iliustracija: status, add, commit, push, pull, clone, gitignore
- R: project, shiny, help(!)
- Atsitiktinių dydžių p,d,q,r funkcijos - teoriškai ir praktiškai.
- Mokomes generuoti dydžius pagal savo norus. Funkcija nuo atsitiktinių dydžių - taip pat atsitiktinis dydis!
- Mokomes daryti išvadas:
- Histograma - branduolinis tankis
- Vidurkis, Sklaida, kvantiliai
- Nestandartinių užklausų tikimybės
- Statistiniai testai (esminis visuotinis principas).

```
?dnorm
?dpois

# p ir d sarysis: p = \int d


# p ir q sarysis


# generavimas
```

Du žaidėja ridena kauliukus. Pirmasis žaidėjas meta du klausiukus, o antrasis tris. Vertinama išmestų kauliukų suma. Laimi tas žaidėjas, kurio išmestu taškų suma yra didesnė. Jeigu iškrenta lygiosios, tai laimi pirmasis žaidėjas. Kokia tikimybė, kad laimės antrasis žaidėjas?

```
# lets define the function that demonstrate dice roll


# apply this function


# lets find out the distribution of sum of two dices


# lets estimate the probability P(dice(3)>dice(2))
```

Užduotys: Pirmasis žaidėjas meta vieną kauliuką ir fiksuojamas atsivertusių taškų skaičius. Antrasis žaidėjas meta du klausiukus ir fiksuojamas dydis: $\max(x_1, x_2) - 1$, kur x_1 ir x_2 žymi pirmojo ir antrojo kauliukų iškritusių akių skaičių. Jeigu gautų taškų skaičius yra lygus, tai ekspermentas kartojamas iš naujo (ir taip kartojama tol, kol galiausiai nustatomas laimėtojas. Kokia tikimybė laimėti pirmajam žaidėjui?

Tarkime metama moneta. Jeigu iškrenta herbas, tai laimite 12 EU. Jeigu skaičius, tai pralošiate 10 EU. Bus žaidžiama n kartų ir gauti rezultatai sumuojami. Kokia tikimybė jums turėti teigiamą balansą po n kartų?

Kokis išlošių pasikliautiniai 95% intervalai. Pateikite atsakymus, kai $n = 1, 2, 5, 10, 20, 100$. Kaip atrodo balanso tankis? Kai n=100.

Kazino Ruletė. Kokia tikimybė po 1000 vienodų statymų turėti teigiamą balansą? Raskite apytikslę x reikšmę, kuriai galioja nelygybe $P(balansas < x) = 0.15$

Generuojame m atsitiktinių dydžių beta su idomiai parametrais. Iš šių atsitiktinių dydžių suskaičiuojama aritmetinis vidurkis. Kaip atrodo vidurkio tankis, kai n=100? Kok5 skirstį tai primena?

Ridename 3 klauliukus. Suskaičiuokite jų teorinį skirstinį ir patikrinkite jį su MC metodu.

Sudėtingasis: Žvejys žvejoja kiekviena darbo dieną. Ir taip jau stabiliai tęsiasi 5 metus. Vidutiniškai žmogus pagauna 20 žuvų per dieną. Tarkime, kad žuvų kiekio skirstinį gerai aprašo Poisono atsitiktinis dydis. Žuvų dydį gerai nusako normalusis atsitiktinis dydis. Pagaunamų žuvų dydžio pasikliautinis 90% intervalas yra [180, 750] (gramų). Pabandykite įvertinti kokį didžiausią laimikį yra pagavęs žvejys. Koks Jūsų įverčio pasikliautinis intervalas?

Futbolas generuojamas pagal eksponentini. Kaip atrodo iverčiu pasiskirstymas varžybų tėkmėje. Kaip atrodo, baigties skirstinys? Patikrinti skirstinius su formaliai testais.

Kazino ruletės analizė.

Fraktalo generavimas

Generuojamas m dyd=iu iš jų ksičiuojamas atsitiktinio dydžio vidurkis. Kaip atrodo vidurkio skirstinys aritmetinio vidurkio skirstinys, kai m=1,5,10, 20, 100, 1000. Ištirkite kai pradinis skirstinys yra normalus, betta, ir kosi.

Kurybine uzduotis: Pirnas ir issamus dpqr sarysiu paaiskinimas, pademostravimas

Salygine tikimybes generavimas - su teroristu testu

Chapter 3. Tiesinė regresija

Literatura:

Moketi:

Beveik tvarkinga tiesinė regresija:

Išskirtys, Multikolinerumas, Heteroskedastiskumas. Interpretacijos atskirimas nuo optimalios prognozes.

Kūribinga dalis: ištiesinimas plano matricos specifikavimas

Prezentacija: Mokėti tiksliai suformuluoti modeli paprasta lygtimi ir matriciniu pavidalu. Gauta lygti prašyti išreikštiniu pavidalu. Mokėti teisinga interpretuoti

Capter 2

MonteCarlo Tikimybe salygine tiimybe Vidurkis dispersija Salyginis vidurkis Marginaline tikimybe Napriklausomi ivykiai (atsitiktiniai dydziai) pasikliautiniai intervalai, p reiksme

Task 3

Pademonstruoti pnorm, dnorm qnorm ir rnornm sarysius. Charakteristiku skaiciavima: Momentu skaiciavima, disperisja, trecia momenta. Pasikliautiniu intervalu sudarymas, p-reisme.

Naudoti ne standartiniu parametrus. Empirine pasisikirtyo funkcija Isvestines Integralas

Task teorija ir montecarlo

Metami 2 ir 3 kauliukai. Kokia tikimybe, kad 2 laimes. rasti teorini ir praktini Ivertinima.

Task MonteCarlo

Funkcija nuo atsitiktinio dydzio generuoja atsitiktinius dydzius. Daug klasikiniu sunkcijus, abs, min, max,

Chalange MonteCarlo on MonteCarlo

References

Kerns, G. Jay. 2011. "Introduction to Probability and Statistics Using R."

Lavine, Michael. 2013. "Introduction to Statistical Thought."