

Parametric Statistics

Sofia Triantafyllou

sof.triantafyllou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

October 5, 2021

Course Info

E-mail: `sof.triantafillou@gmail.com`

Web: <https://polyhedron.math.uoc.gr/2122/moodle/enrol/index.php?id=15>

Class Hours: Tuesday, Friday 11.00-13.00 (A214)

Office Hours: Wednesdays 13.00-15.00

Extra lecture: Thursday 9.00-11.00 (A212)- counts as office hours
when we do not have a lecture

Office: B316, zoom id: 925 694 5046

Course Materials

The course notes are based on the following books:

(LW) All of Statistics, Larry Wasserman, 2004

(DGS) Probability and Statistics, Morris DeGroot and Mark Schervish (International Edition), 2014 (4th edition)

Course notes will be available on the class website. The material can also be found in Greek in the following books:

- ▶ Εισαγωγή στη Στατιστική ΜΕΡΟΣ Ι, Δαμιανού Χ., Κούτρας Μ.
- ▶ Στατιστική συμπερασματολογία, (Τόμοι Ι & ΙΙ), Ρούσσας Γεώργιος Γ., Σταματέλος Γεώργιος.

Grading

- ▶ 2 midterm exams, 30% each
- ▶ Final exam, 30%
- ▶ (Group) homework assignment, 20%

Prerequisites

Required preliminary math tools are elementary probability, calculus and basic linear algebra. There are no formal prerequisites for this class, but I would advise that you have passed Calculus I, Calculus II and Probability Theory.

What is statistics

Statistics and probability are concerned with describing uncertainty. In both disciplines, we have some **experiments** that can have **multiple outcomes** that occur with some **probability**.

In your probability course, you start with a distribution (say, a normal distribution with a mean μ and a variance σ^2) and predict features of future observations $x = (x_1, x_2, \dots, x_n)$.

What is statistics

In statistics we observe the data x and then try to understand the properties of the data-generating process.

In parametric statistics, we assume that the process can be described using a finite number of parameters θ and try to identify these parameters.

Example

We are measuring the height of the students in the class. We assume that they follow a normal distribution. Which are the parameters we need to identify?

Things we will learn in this course

- ▶ Experiments, Sample Spaces, Events
- ▶ Probability
- ▶ Independence, Conditional Probability
- ▶ Random Variables and their distributions
- ▶ Expectation, variance
- ▶ Convergence

Things we will learn in this course

Statistical Inference

- ▶ Point Estimation
- ▶ Hypothesis testing
- ▶ Prediction

Data Collection

We will perform an ANONYMOUS survey.

You have been given an empty card. Write your height and your gender on the card (if you do not want to answer, you can leave any or both fields blank).

What our data looks like

Height	Gender
1.75	F
\vdots	\vdots
1.68	M

Assume the experiment (random process) is picking a student at random and measuring their height. Let's discuss:

- ▶ The sample space.
- ▶ The random variable denoting the height of the student.
- ▶ Our population and our sample.

Let's do some statistical inference

Height	Gender
1.75	F
\vdots	\vdots
1.68	M

- ▶ Point estimation: Assuming the data follow a Normal $N(\mu, \sigma^2)$ distribution, what is the "best guess" for μ and σ ?
- ▶ Hypothesis testing: Is the mean height of females (significantly) smaller than the height of males? (Do we have the data to test this hypothesis?)
- ▶ Assume we had also measured every student's height at age 2. Can we use height at age 2 to predict adult height?

Statistical inference using R

- ▶ Install R: <https://cran.r-project.org/>
- ▶ Install R-studio:
<https://www.rstudio.com/products/rstudio/download>
- ▶ Download class_data_heights.xlsx from the class website
- ▶ Run the commands in Rscript0.R to see the summary of our data.