

## Probability and Statistics (O000139)

Course size (nominal values; actual values may depend on programme)  
 Credits 10.0 Study time 250 h Contact hrs 120.0 h

### Course offerings and teaching methods in academic year 2019-2020

A (year)	English	seminar: coached exercises	17.5 h
		lecture	50.0 h
		seminar: practical PC room classes	32.5 h

### Lecturers in academic year 2019-2020

Van Messem, Arnout KR01 lecturer-in-charge

### Offered in the following programmes in 2019-2020

	crdts	offering
<a href="#">Bachelor of Science in Environmental Technology</a>	10	A
<a href="#">Bachelor of Science in Food Technology</a>	10	A
<a href="#">Bachelor of Science in Molecular Biotechnology</a>	10	A

### Teaching languages

English

### Keywords

Probability theory, statistics, hypothesis testing, data analysis, regression, simulation of models, model selection, scientific thinking

### Position of the course

*First, the students are introduced to probabilistic and statistical concepts. In the second part, simulation methodologies for (bio)systems are discussed. Theory is illustrated with ample examples. The students learn to perform statistical techniques and correctly describe and interpret statistical data and output. They learn to distinguish between haphazard effects on the one hand and scientifically significant results on the other hand. The students will learn to critically read and evaluate results presented in scientific literature. The statistical software R is used throughout the course.*

### Contents

#### **Probability theory**

- Discrete and continuous random variables, random vectors
- Discrete and continuous distributions (binomial, Poisson, t, normal, ...), joint and marginal distribution
- Basic concepts: expectation, variance, standard error, correlation, probability measure, conditional probability, Bayes theorem, law of total probability, independence

#### **Statistics**

- Study design: randomized and observational studies, sample size calculation
- Descriptive statistics: histogram, stem-and-leaf plot, summary measures, density plot, QQ plot
- Point estimates and confidence intervals for the population mean, central limit theorem
- Hypothesis testing:
  - One-sided and two-sided, significance level, p-value, power, Type I and Type II error, power analysis
  - Z-test, one-sample and two-sample t-test, F-test, multiple testing (family-wise error rate, false discovery rate), ANOVA (comparison of multiple means), non-parametric tests, analysis of contingency tables

#### **Data analysis**

- *Principal component analysis*
- (\*) *Classification: linear and quadratic discriminant analysis*
- *Simple and multiple linear regression, principal component regression*

- (\*) *Cluster analysis*

### **Simulation of dynamical systems**

- *Model simulation: analytical vs. numerical, Monte Carlo*
- *Parameter estimation: objective functions, minimization algorithms, quality, validation*
- *Local and global sensitivity analysis*
- *Uncertainty analysis*
- *Model selection*

(\*) *Optional topics depending on time constraints*

### Initial competences

*Mathematics 1: Engineering Mathematics*

### Final competences

- 1 Select the appropriate experimental design for simple research questions
- 2 Translate a research question into a statistical analytical method
- 3 Perform statistical analyses or model simulations with the software R
- 4 Report results of statistical analyses in a concise and scientifically correct way
- 5 Critically analyze statistical information presented in scientific publications
- 6 Perform a parameter estimation for a given model structure
- 7 Identify different sources of uncertainty in a model
- 8 Quantify and compare the local and global sensitivity of model attributes
- 9 Select the optimal model from a set of models for a given system
- 10 Take responsibilities and initiative in group work

### Conditions for credit contract

Access to this course unit via a credit contract is unrestricted: the student takes into consideration the conditions mentioned in 'Starting Competences'

### Conditions for exam contract

Access to this course unit via an exam contract is unrestricted

### Teaching methods

Lecture, seminar: coached exercises, seminar: practical PC room classes

### Learning materials and price

Course notes and slides are available on Minerva

### References

*D. Moore D. and G. McCabe. Introduction to the Practice of Statistics, 5<sup>th</sup> edition  
R.C. Campbell (1989). Statistics for Biologists; Cambridge University Press; 3<sup>rd</sup> edition.*

*J. Fowler, L. Cohen, and P. Jarvis (1999). Practical Statistics for Field Biology; John Wiley & Sons; 2<sup>nd</sup> edition.*

*R.R. Sokal and F.J. Rohlf (1994). Biometry; W.H. Freeman & co; 3<sup>rd</sup> edition.*

*P.A. Vanrolleghem and D. Dochain D. (1998). Bioprocess Model Identification. In: Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes Eds.*

*B.A. Ogunnaike and Ray W.H. (1994). Process Dynamics, Modeling and Control. Oxford University Press.*

*D.M. Diez, C.D. Barr, and M. Çetinkaya-Rundel. OpenIntro Statistics; 3<sup>rd</sup> edition (<https://www.openintro.org/stat/>)*

### Course content-related study coaching

### Evaluation methods

end-of-term evaluation and continuous assessment

### Examination methods in case of periodic evaluation during the first examination period

Written examination with open questions, open book examination, skills test

### Examination methods in case of periodic evaluation during the second examination period

Written examination with open questions, open book examination, skills test

### Examination methods in case of permanent evaluation

Assignment, peer assessment

Possibilities of retake in case of permanent evaluation

examination during the second examination period is possible in modified form

Calculation of the examination mark

End-of-term evaluation: 60%

- Written examination with open questions, open book: 30%
- Computer based examination with open questions, open book: 30%

Assignment: 40%

- Group project on hypothesis testing: 20%
- Group project on linear regression: 10%
- Group project on simulation: 10%

Students not partaking in the continuous evaluation (i.e., handing in 3 projects of decent quality), can obtain a maximal score of 7/20 for the entire course, irrespective of their grade for the end-of-term evaluation.