- Faculté des sciences économiques
- www.unine.ch/seco

Gestion, traitement et visualisation des données (5MI1014)

Filières concernées	Nombre d'heures	Validation	Crédits ECTS
Bachelor en droit	**Cours: 2 ph Exercice: 2 ph**	**Voir ci-dessous**	6
Bachelor en science des données	**Cours: 2 ph Exercice: 2 ph**	**Voir ci-dessous**	6
Bachelor en sciences économiques, orientation économie	**Cours: 2 ph Exercice: 2 ph**	**Voir ci-dessous**	6
Bachelor en sciences économiques, orientation management	**Cours: 2 ph Exercice: 2 ph**	**Voir ci-dessous**	6

ph=période hebdomadaire, pg=période globale, j=jour, dj=demi-jour, h=heure, min=minute

Période d'enseignement:

- Semestre Automne

Contenu

This is a hands-on course that integrates introductory programming, statistics and data science. Through out this course, we will formulate scientific hypotheses, design experiments, and collect and analyse data visually and through formal models. You are expected to supplement this course with homework, self-study and other courses in descriptive statistics and python programming, but no prior knowledge is assumed. Formal concepts will be introduced through class activities and examples.

The course will focus neither on statistical theory nor on programming. We will only lightly build those skills within the course. Our main aim will be to build scientific and statistical intuition through practical work. However, to achieve this you need to be able to independently pick up theoretical and programming skills. For that reason, it is necessary for you to do outside reading either (ideally) by taking statistics and programming courses in parallel, or through self-study.

** Learning goals:
Graphical comprehension:

1. Recognise structural eelemtns in a statistical graph (e.g. axis, symbols, labels) and evaluate the effectiveness (for perception and judgment) and appropriateness (for the type of data) of structural element
2. Translate relationships reflected in a graph to the data represented.
3. Recognise when one graph is more useful than another and organise/reorganise data to make an alternative representation.
4. Use context to make sense fo what is presented in a graph and avoid reading too much into any relationships observed.
5. Express creative thinking via the production of an innovactive graphical presentation.

Scentific goals: Statistics, analysis and experiment design

1. Understanding the randomness, variability and uncertainty inherent in a problem.
2. Developing clear statements of the problem/scientific research question; understanding the purpose of the answer.
3. Be able to perform a basic experiment design.

URLs	1) https://github.com/olethrosdc/digital-skills

- Faculté des sciences économiques
- www.unine.ch/seco

Gestion, traitement et visualisation des données (5MI1014)

4. Identify sources of bias in data collection and analysis.
5. Ensuring acquisition of high-quality data and not just a lot of numbers.
6. Understanding the process that produced the data, to provide proper context for analysis.
7. Allowing domain knowledge to guide both data collection and analysis.
8. Quantify uncertainty---and knowledge---visually.
9. Realise that all visualisations are models.
10. Be able to write simple python programs for data science workflows.

Forme de l'évaluation

The assessment is purely through in-class exercises, quizzes and homework assignments. There will be assignments spread over the semester, as well as a group project. The project will be performed in pairs.

For all assignment and the project, the following rubrik is used. Some of the assignments may not involve all parts.

Experiment design The first stage any project, no matter how small, is the experiment design and analysis. This includes a plan for how to collect data, one or two methodologies for analysing the data, and the development of a pipeline, preferrably in the form of a program, for collecting data and analysing it. In additional, the experiment design must be reproducible: This can be ensured by running the data collection and analysis pipeline on simulated data, and seeing if the results are as expected.

Computation Here you must instantiate the experiment design and analysis with concrete computations. For reproducibility, the computations you perform should be independent of the data you actually have. Correctness of the computations is the most important aspect, here. However, you should also take care to document why and how how you are doing the computations.

Graphics This addresses the creation of visualisations of your analysis. It is recommended to do this fully automatically, so that you can simply run your pipeline and get all the results you need. Be sure to quantify uncertainty.

Text Here you should explain in text what the graphics mean. Point out any interesting things you can see in the visualisation and try to explain it. Do not be overconfident, but quantify uncertainty properly.

Synthesis Here you should summarise the most important findings from your analysis. Be careful to not over-interpret your results. A lot of results can be imaginary and can be attributed to insufficient data, biased sampling, improper modelling or p-value hacking. Be sure to quantify uncertainty.

Pré-requis

Basic mathematics. Ability to use a computer.

URLs	1) https://github.com/olethrosdc/digital-skills

- Faculté des sciences économiques
- www.unine.ch/seco

Gestion, traitement et visualisation des données (5MI1014)

Forme de l'enseignement

Demonstrations, in-class exercises, live coding, discussion groups, assignments and group projects.

URLs	1) https://github.com/olethrosdc/digital-skills