Mentoring a group for the
Applied Statistics Project

Each year, ENSAE Paris organizes a course called Applied Statistics Project. This course complements the theoretical courses in statistics, econometrics, surveys, time series and introduction to machine learning offered in the second year. It enables students to familiarize themselves with the various stages of statistical modeling (model development, estimation and testing) and apply them to real data.

The project is supervised by one or two tutors. The tutors propose a topic related to his/her professional occupation and provides the data to their selected group. Their role is then to coordinate and guide the project, by meeting the students regularly throughout the year. As a result, they must have statistical expertise. At the end of the project, they attend the oral presentation and participate in the final grading.

Most of the work will involve statistical modeling and the application of methods taught in the second year of ENSAE (linear and logistic regressions, standard time series models, random forests, etc.). Concepts taught in the third year (duration models, non-parametric methods, bootstrap, etc.) may also be introduced, with moderation.

The final grading will be based on a progress report (due mid-February 2025), a memoir and summary note (due mid May 2025), and an oral presentation (mid-May to mid-June). This enables students to learn how to present technical work clearly and concisely.

Educational Objectives

Through the choice of a tutor working in the following fields of statistics, economics, finance or even healthcare studies, ENSAE aims to train students in the following areas:

- Asking a scientific question adapted to a real database;
- Carry out a collective scientific approach (groups of 3 or 4 students), guided by a mentor, which includes a literature review, a statistical description of the data and the use of appropriate quantitative techniques (statistics and econometrics) to answer the problem;
- Synthesize this approach into a 20-page report for specialists and a summary note for non-specialists.

-

This project is *applied* in the sense that the work is carried out on real data, provided by the mentor, or at least on simulations. This is applied teaching, not a full-scale study, so the data files made available to students must be clean and of reasonable size.

It's important to stress the pedagogical nature of this project: the mentor must have the expertise to answer students' questions on technical and methodological aspects, and is paid for this. This is by no means a consulting mission, similar to the services provided by ENSAE's junior enterprise, for example.

The level of confidentiality applying to the data and the report must be defined between the mentor and the students. The latter are bound by contract by the obligation of professional discretion and are therefore responsible for the security of the files entrusted to them.

Practical Organisation

If you wish to propose a project, you must complete the *project presentation template*, attached to the e-mail, and fill in the following form:

https://projets-entreprise.ensae.fr/

Students form their own groups of three or four, and contact the mentors of projects that interest them directly. If there is mutual interest, the group and the mentor decide to work together and keep us informed by e-mail. We update the assignment directly online, so that the mentor is no longer contacted by other groups.

The project timetable is as follows:
- Mid-October: announcement of projects to students
- Mid-November: final group assignments
- Mid-February 2025: submission of a progress report
- Mid-May 2025: submission of the memoir and summary note
- Late May - early June 2025: oral defense

The hourly volume of teaching is as follows: ten two-hour follow-up sessions (an average of one and a half sessions per month), preferably on ENSAE premises. Gross remuneration (paid in August 2025) is around €1215 for one instructor, or €729 per instructor if there are two.

If you have any questions, please contact stat_app@ensae.fr and we'll get back to you as soon as possible.

PROJECT PRESENTATION TEMPLATE

SNCF electricity consumption forecast

Name(s) of the mentor(s)	Jean Martin
Organization	SNCF
Phone Number	+33 1 45 XX XX XX
Email address	Jean.martin@sncf.fr

DESCRIPTION OF THE PROJECT

Forecasting traction power consumption is one of the levers of efficiency in energy purchasing. Good anticipation of needs improves purchasing performance, as well as visibility in terms of the hedging strategy to be implemented on the financial markets.

High-voltage electricity consumption depends on the activity of the company's various entities (passenger and freight transport), but also on major exogenous factors such as temperature.
The aim of the proposed study is to develop an efficient model for forecasting electricity consumption at a relatively fine level of detail (hourly, for example).

The stakes are many, but above all financial. Hundreds of millions of euros are at stake.

Students will need to master Python software, as well as standard office software (LaTex, Pack office).

DATA

The data used in this study are of various kinds:
- Traffic data for various activities (train kilometers, tons towed, etc.);
- Electricity supply data for the national rail network;
- Meteorological data.

BIBLIOGRAPHY

Traction Electrique – *JM Allenbach / P. Chabpas / M. Comte / R. Kaller* – Presses polytechniques et universitaires romandes
Consommation française d'électricité caracteristiques et methode de prevision - *RTE*