

## Course Data-analysis and statistics

<b>Date:</b>	<b>5 Nov, 19 Nov, 10 Dec 2014</b>
<b>Time:</b>	<b>10.00 – 17.00 h.</b>
<b>Location:</b>	<b>Schiphol, Dinalog, Evert van de Beekstraat 356, SADC (2nd floor), Building C</b>
<b>Course leader:</b>	<b>Dr. Maarten Kroesen and Dr. Raymond Hoogendoorn</b>
<b>Days:</b>	<b>3</b>
<b>ECTS:</b>	<b>1 (attendance) / 3 (with assignment)</b>
<b>TUD GS credits:</b>	<b>2 (attendance) / 5 (with assignment)</b>
<b>Course fee:</b>	<b>Free for TRAIL/Beta/OML members, others please contact the TRAIL office</b>
<b>Registration:</b>	<b><a href="http://www.rstrail.nl">www.rstrail.nl</a></b>

### Objectives

After this course attendees are able to:

1. Know the various data collection techniques in the domains of transport, infrastructure and logistics and their strengths and weaknesses (in terms of internal and external validity, self-selection bias, measurement biases, etc.).
2. Recognize the probability density functions of various discrete (Bernoulli, Binomial, Geometric, Poisson) and continuous random variables (Uniform, Exponential, Normal).
3. Explain the basic principles behind statistical modelling (law of large numbers, Central Limit Theorem).
4. Choose appropriate data-analysis techniques\* (given a particular research question), correctly apply these techniques (by formulating statistical hypotheses, checking the statistical assumptions and deriving the test statistic) and interpret their results in meaningful ways.

\* The following data-analysis techniques will be treated: descriptive data analysis (mean, median, variance, standard deviation), univariate (one sample t-test, proportion test) and bivariate parametric tests (paired/independent samples t-test, ANOVA, Pearson correlation), non-parametric tests (chi-square, Kolmogorov-Smirnov, Wilcoxon rank, Kruskal-Wallis and Spearman rank correlation) and multiple regression (the basics).

### Course description

In this course attendees will actively work on solving concrete statistical problems in the domains of transport, infrastructure and logistics using various (parametric and non-parametric) data-analysis techniques. The course will extensively treat the basic principles behind statistical modelling so that attendees really understand what the results of statistical tests mean. To emphasize that good data quality is a necessary condition for any subsequent statistical testing, different data collections techniques will be treated (on day 1) on their strengths and weaknesses discussed.

### Assignment

Attendees have to:

1. Write a short essay describing the data collection method that they have applied (or are planning to) and reflect on the strengths and weaknesses of their method.
2. Apply several data-analysis techniques to their own data (or a given dataset) and report the results in a brief research report.

**Program**

Day 1 - Probability density functions, basic principles behind statistical modelling, descriptive statistics and bivariate data-analysis techniques

Day 2 - Data collections techniques in the domains of transport, infrastructure and logistics

Day 3 - Bivariate data-analysis techniques and multiple regression

**Course material**

Slides and online materials

**Prerequisite**

None