

HSL Faculty, UiT The Arctic University of Norway, 8.2.2017	TEMPLATE FOR COURSE DESCRIPTIONS, THE HSL FACULTY Please see explanation to each point below. The template is based on requirements for modules within the UiT quality system.
Name	Bokmål: Statistics for linguistics with <i>R</i> bootcamp Nynorsk: English: Statistics for linguistics with <i>R</i> bootcamp
Course code and level	LIN-8011
Type of course	The course may be taken as a single course.
Scope of course	5 ECTS
Søknadsfrist	Ph.d.-students at UiT register for class and exam in Studentweb by June 1st. Other applicants: June 1 st Application code 9301 in Søknadsweb.
Required / recommended previous knowledge	PhD students or holders of a Norwegian Master's Degree of five years or 3+ 2 years (or equivalent) may be admitted. PhD students must upload a document from their university stating that they are registered PhD students. Holders of a Master's Degree must upload a Master's Diploma with Diploma Supplement / English translation of the diploma. Applicants from listed countries must document proficiency in English. To find out if this applies to you see the following list: http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Utenlandsk_utdannin_g/GSULista/2016/GSU_list_English_14112016.pdf For more information on accepted English proficiency tests and scores, as well as exemptions from the English proficiency tests, please see the following document: https://uit.no/Content/254419/PhD_EnglishProficiency_100913.pdf
Course contents	<i>Statistics for linguistics with R</i> is a hands-on introduction to statistical methods for both graduate students and seasoned researchers and is based on the second edition (2013) of Gries' textbook <i>Statistics for linguistics with R</i> . The course is mainly intended for linguists who already have a basic knowledge in statistics and some experience using <i>R</i> , and who wish to improve their proficiency in statistical analysis of linguistic data. Participants who are new to statistics and/or <i>R</i> should prepare beforehand by working through the readings listed below. The course puts a particularly strong emphasis on various kinds of fixed- and mixed-effects regression modeling as well the use of other predictive modeling techniques such as classification/conditional inference trees and (random) forests. The course features: <ul style="list-style-type: none"> • a brief recap of basic aspects of statistical evaluation as well as several descriptive statistics insofar as they facilitate later predictive modeling approaches; • a selection of monofactorial statistical tests for frequencies, means, and correlations and how they constitute special (limiting) cases of regression methods; • an exploration of different kinds of multifactorial regression modeling approaches as well as other techniques on the basis of both observational and experimental, published and unpublished data.

	For all modeling methods to be explored, we will discuss how to test their assumptions and visualize their results with instructive annotated statistical graphs. There also will be in depth discussion of different model selection strategies, how to interpret predictive modeling results (such as different kinds of interactions and contrasts), threats to the validity of modeling, etc.
Learning outcomes	<p>The students have the following learning outcomes:</p> <p>At the end of the course, participants will be able to understand any discussion of a regression model they come across in research literature and will be able to conduct their own fixed- and mixed-effects modeling analyses; time permitting, there will be a small section on how to write small statistical/visualization functions yourself.</p>
Relevance in the degree program	Should be provided, but not a requirement.
Teaching and working methods Teaching methods, scope and frequency should be described. Also provide information about the number of lectures / classes.	<p>This is a five-day intensive course that requires:</p> <ul style="list-style-type: none"> • the reading assignment (see Required Readings listed below) to be completed prior to the start of the course; • downloading and installing the software (which you will have been emailed about) via links and emailed instructions prior to course start; • testing that the software packages are functional on your computer prior to class. <p>The course will be taught in English and grading is done on a pass/no pass basis. The course will feature lecture-style teaching, with about half of the instructional time each day being hands-on work on a variety of different data sets. Data sets and (thousands of lines of) code will be provided to the participants, as will be a variety of helper functions that participants will be able to use for their own statistical applications. Also, we will discuss queries that were sent to <i>R</i> newsgroups as well as reviews of papers under review with an eye to help participants understand what mistakes to avoid. The course will consist of a morning and an afternoon teaching module from Monday through Friday of one week. It will run much longer than the typical “class”, hence the name bootcamp, starting at 9am and finishing at 5pm with a 1.5 hour break for lunch at midday, and 30-minute coffee breaks in the morning and afternoon.</p> <p>Course schedule:</p> <p>Day 1: 3-4 hours lecture: linear fixed-effects modeling; 2-3 hours practice</p> <p>Day 2: 3-4 hours lecture: generalized linear fixed-effects modeling; 2-3 hours practice</p> <p>Day 3: 3 hours lecture: linear mixed-effects modeling; 3 hours practice</p> <p>Day 4: 3 hours lecture: generalized linear mixed-effects modeling; 3 hours practice</p> <p>Day 5: 3-4 hours lecture: tree-based approaches; 2-3 hours practice</p> <p>(approx. 16 hours of teaching and 14 hours tutoring in total, yielding 5 classes and 30 hours)</p>
Practice	
Quality assurance of the course	The students will be expected to evaluate the overall quality of the lectures, relevance of the reading materials, student-instructor interaction and learning outcomes achieved. All course evaluation reports provided by students will be submitted to the Norwegian Graduate Researcher School in Linguistics and Philology (LingPhil)

	after the course. The template for course evaluation by students can be found at https://www.ntnu.edu/lingphil/course-proposals .
Coursework	
Assessment and exam	<p>PhD students will be awarded 5 ECTS if they</p> <ul style="list-style-type: none"> • read the required texts and download and test the software prior to the course; • attend all teaching sessions; • provide a written question each evening to the instructor, a selection of which will be used in the course to go over common queries; • complete one practical assignment of a data set provided by the instructor as the final assessment. <p>Any student with an interest in statistics for empirical research is encouraged to attend.</p> <p>The exam will be assessed on a Pass/Fail basis.</p>
Retake	Retake is offered in in the beginning of the following semester in cases of grade F or Fail. Deferred examination is offered in the beginning of the following semester if the student is unable to take the final exam due to illness or other exceptional circumstances. Registration deadline for retake is January 15 for autumn semester exams and August 15 for spring semester exams.
Syllabus	All registered applicants will receive a link to an on-line page where all readings can be downloaded. The reading assignment is to read all required readings and be familiar with recommended readings.
Language of instruction and examination	English