

## University College of Southeast Norway

### EXAM 002

### 9010 STUDY DESIGN AND STATISTICS

29.1.2016

Time: 0900-1400

Language: English

Number pages: 5 (including front page)

#### General instructions

- All written and electronic materials (class scripts, online resources) are allowed. Direct inquires and use of social media, telephone, etc. is NOT permitted.
- You are provided with necessary data files (5) and the empty template for the answer script at the start of the exam. Use the template answer script to enter all your R scripts and other answers to the exam questions. This is the ONLY file you will be submitting.
- Read the questions carefully. Even single-part questions may require more than one task/answer. Depending on the question/task provide your answers as R commands or as comments. For questions that involve writing R code, it has to be clear from the code how you arrived at your answers. Points may be deducted if an answer cannot be derived/repeated from your code. Note, even if you do not know the answer to a question, you may still get credit for partial answers.
- Name your final R script file with your name (e.g.: “*bob\_swenson.R*”) and provide the file to the Examiner AND email it to [andreas.zedrosser@hit.no](mailto:andreas.zedrosser@hit.no) BEFORE the end of the exam.

#### Data description

“hares”: Results from a population study where individual hares were monitored and various individual attributes (e.g. size in cm), survival, and source area were recorded.

“pike”: Study on pike (a fish; Norwegian: gjedde) collected from different lakes. Each animals length (in cm) and tissue mercury concentration (in parts-per-million, ppm) were recorded.

“ticks”: Ticks (Norwegian: flott) were collected from dogs. The weight of each tick, as well as the time it spent feeding on the dog were recorded.

“bird\_counts”: Birds survey at various locations. At each location the number of birds observed/heard, the temperature, and the habitat type are recorded.

“nonlinear\_data”: Simulated data with a non-linear relationship between x and y.

**! → Code for loading all data in R are provided in the template script.**



Fakultet for allmennvitenskaplege fag.



## Exam questions

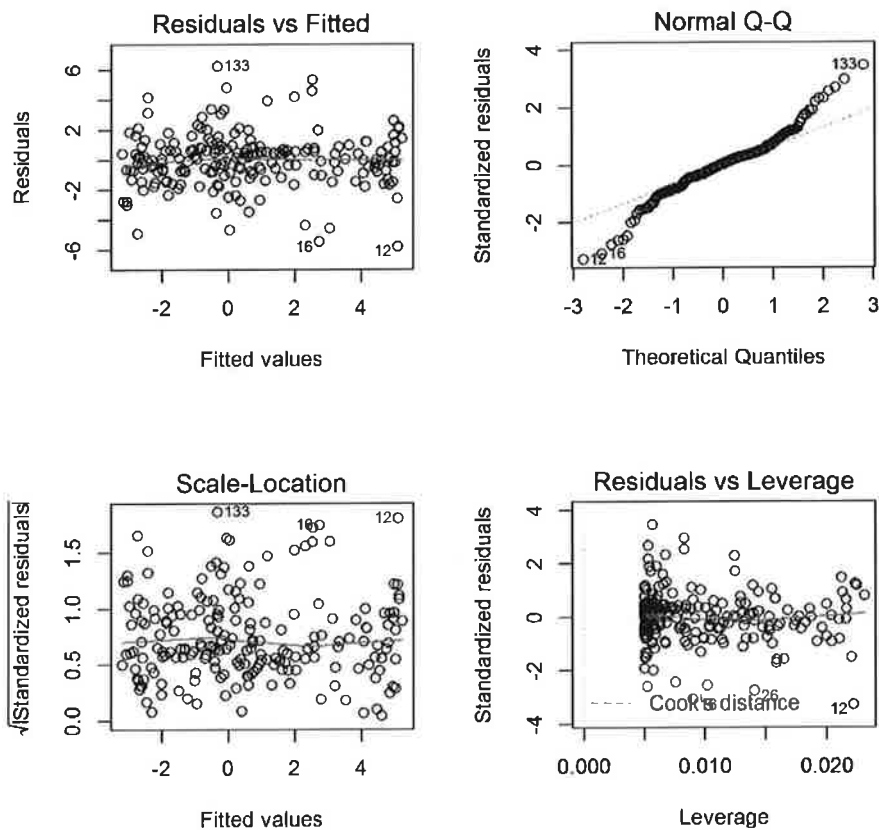
(Total: 7 questions, 32 points = 100 %)

1.

- A researcher is conducting bird surveys. She is visiting different locations in her study area and recording the number of birds observed or heard. Using linear regression and the "bird\_counts" data, you are to determine whether habitat type at a survey location and temperature influence the number of birds observed/heard. Fit an appropriate model to the data to answer this question, use AIC to select the best model, and display the model output. (2 pts)
- In 2-3 sentences, describe your conclusions based on the model output. (2 pts)

2.

- Inspect the diagnostic plots shown below. Which assumption of simple linear regression is being violated? Explain how you arrive at your answer. (1 pt)
- Describe a possible solution (1 sentence). (1 pt)



- 3.
- The "hares" data contains information about the survival (TRUE/FALSE) of hares during a monitoring study, as well as several individual attributes for each animal. You are to use linear regression to determine how "area" (hunted/not hunted) and "sex" (female/male) affect the probability of survival. Fit an appropriate model to the data to answer this question, use AIC to select the best model, and display the model output. **(2 pts)**
  - In 2-3 sentences, describe your conclusions based on the model output. **(2 pts)**
  - Using the model, predict survival (as a probability) for female hares in the hunted area. **(2 pts)**
- 4.
- Generate a density plot (with informative axis labels) to show the distribution of sizes of adult females hares (i.e. 2 years and older) in the hares data (use blue for the line color). **(2 pts)**
  - Generate random normally distributed data (N=10 000) with the same mean and standard deviation as the distribution of adult females hare sizes and add the resulting distribution to the plot (use red color for the line). **(2 pts)**
  - Does the size of adult female hares appear to be normally distributed? **(1 pts)**
- 5.
- A researcher is studying the blood-ingestion rate of ticks (proxy: weight of the tick) as a function of time spent on the host (weight ~ hours). Several dogs are exposed to a tick-rich environment until each has numerous ticks attached. Following exposure, ticks are removed at various times from each dog. The host dog's ID, the time until removal (in hours), and the weight of each tick (in microgram) are recorded. Based on this description of the study and inspection of the data "ticks", which assumption of the linear model do you conclude is being violated? **(1 pts)**
  - How could you account for this violation in your analysis? Implement 2 different solutions using regression models **(4 pts)**
- 6.
- The Environmental Protection Agency has determined that fish with a mercury concentration below 0.5 ppm (parts-per-million) are safe for human consumption. Using linear regression and the "pike" data, answer the following question: Approximately, below what size are pike safe to eat? **(4 pts)**
  - Generate a plot that illustrates the model prediction together with the data and shows the threshold for "safe" mercury concentration as a dashed line. **(2 pts)**

7.

- a. Fit a regression model for estimating the effect of x on y in the "nonlinear\_data" object. Implement the model so that it can account for the non-linear relationship between the two variables. **(2 pts)**
- b. Generate a plot that allows you to compare the model prediction with the underlying data. (i.e. overlay a scatterplot of the raw data with the fitted line from the model). **(2 pts)**