

Seminar 3: Pancreatic Cancer

Start Assignment

- Due 16 Nov by 23:59
- Points 1
- Submitting a file upload
- File types pdf, r, and rmd

General Instructions

Written solutions to all the tasks must be submitted before the **deadline: Sunday November 16 at 23:59** (as pdf files in Canvas, preferably including the most important parts of your code).

You are also expected to prepare an oral presentation of your solution for each of the tasks. The presentation should aim at taking 15 minutes, to leave room for questions and discussion. At the seminar, a member of your group (randomly chosen by the teacher) will be asked to present the solution.

We expect each of you to spend around 15h on these tasks; please plan your time and meetings keeping this in mind.

Please note that the course book (and the lectures) are not your only sources of information. There are lots of information available in the research/educational literature and on the Internet about statistics as well as R and its possibilities. You are very much encouraged to try and find methods not mentioned in the book or the lectures.

Background: Pancreatic Cancer

Pancreatic cancer is one of the deadliest forms of cancers due to late-stage diagnosis. The incidence of pancreatic cancer is increasing globally. It is thought that this is due to ageing and changes in environmental and lifestyle factors (such as alcohol consumption , smoking, diet, activity, *etc.*).

You are part of an EU funded research consortium, studying the epidemiology of pancreatic cancer and repurposing of existing cancer treatments.

Task 1: The EUxPancreas Cohort Study

A large cross-sectional study, named EUxPancreas, was carried out to study the incidence of

pancreatic cancer in five different metropolitan regions across Europe, all with population samples of approximately 2 Million inhabitants.

The dataset, [Data_T1.csv \(https://canvas.kth.se/courses/55919/files/9706663?wrap=1\)](https://canvas.kth.se/courses/55919/files/9706663?wrap=1), [↓ \(https://canvas.kth.se/courses/55919/files/9706663/download?download_frd=1\)](https://canvas.kth.se/courses/55919/files/9706663/download?download_frd=1), contains the following variables:

Variable	Description
NewCases	Number of new cases of pancreatic cancer over one year.
Npopulation	Population number of the subgroup (region-sex-age strata).
Region	5 levels: regions 1 to 5.
AgeGroup	3 levels: 20-39 years, 40-59 years, 60-79 years.
Sex	2 levels: male, female
CLIstd	Standardised composite lifestyle index (CLI): accounting for e.g., physical activity, diet, sleep, stress management, alcohol consumption.
SmokingPrevalence	[fraction]
BMImedian	Body Mass Index, median[kg/m ²]


- Explore the dataset.
- Develop a model to examine the trends in new cases of pancreatic cancer across the recorded population variables.
- Analyse the model outputs and examine model performance.

Task 2: Tumour Models in Mice for Studying Treatment Effect

Organic chemists and drug metabolism-pharmacokinetic researchers have identified an experimental treatment for pancreatic cancer, a compound named ATC10X. Preclinical data suggests that the compound has the advantage of being mainly renally cleared. Because of this few interactions with other treatments are expected, meaning that the treatment could potentially be easier to manage.

Researchers have carried out a study of treatment effect of ATXC10X. A parallel preclinical pharmacodynamic (PD - the study of drug effect over time) study have been carried out in mice with induced tumours and xenograft mice with human cancer. These as two different disease models of


cancer in mice that are used in pharmaceutical research.

The dataset, [Data_T2.csv \(https://canvas.kth.se/courses/55919/files/9706664?wrap=1\)](https://canvas.kth.se/courses/55919/files/9706664?wrap=1)  (https://canvas.kth.se/courses/55919/files/9706664/download?download_frd=1) , contains the following columns: ID: subject id; Time: Time from initiation of study [days]; DV: tumour volume [mm³]; Treatment: treatment group = 1, control = 0; Model: 0 = mice with induced tumours, 1 = xenograft mice.

Study the effect of the experimental treatment on tumour growth.

- Explore the data visually.
- Propose and develop a statistical model to analyse the effects of treatment across the two mice models.
- Analyse the statistical model results.
- Carry out predictions to explore the impact of treatment and disease model.
- Comment on the ability to carry out predictions using your model, within and beyond the ranges of the data (interpolation and extrapolation).

Task 3: Treatment Effect in Late-Stage Pancreatic Cancer Patients

Following promising preclinical results, a clinical study was carried out in late stage pancreatic cancer patients to evaluate the experimental treatment ATC10X against the gold-standard treatment. The study is reported in the data-file, [Data_T3.csv \(https://canvas.kth.se/courses/55919/files/9706665?wrap=1\)](https://canvas.kth.se/courses/55919/files/9706665?wrap=1)  (https://canvas.kth.se/courses/55919/files/9706665/download?download_frd=1) , and includes the following variables:

ID: patient ID; TreatmentGroup: Treatment/Control; Age [yrs]; Sex: Male/Female; ECOG_PS: ECOG Performance Status; GFR: glomerular filtration rate [mL/min/1.73m²]; Time: time of event [weeks]; Event: 1: deceased, 0: survived.

- Analyse the data visually.
- Determine a modelling approach to the estimate treatment effect.
- Analyse the data.
- Interpret and discuss the impact of the different variables on treatment response.

Task 4: Adverse Effects following Treatment

Additional data on adverse effects have been made available for the treatment group in Task 3.

The data file, [Data_T4.csv \(https://canvas.kth.se/courses/55919/files/9706666?wrap=1\)](https://canvas.kth.se/courses/55919/files/9706666?wrap=1)  (https://canvas.kth.se/courses/55919/files/9706666/download?download_frd=1)

canvas.kth.se/courses/55919/files/9706666/download?download_frd=1 , contains IDs from the treatment group in the Task 3 dataset and reported cases of neutropenia (a potentially serious adverse effect linked to the experimental treatment).

- Combine the datasets in Task 3 and Task 4.
- Carry out an analysis of the factors associated with neutropenia.
- Interpret the results, reason around the potential causes for your findings.

What additional information would you like to collect? How would you design a follow-up study?