Final Project Assignment

Background

Diabetes mellitus (DM) is a metabolic disease associated with high blood sugar levels over a prolonged period. Type 1 diabetes mellitus (DM1) is caused by autoimmune destruction of insulin-producing cells of the pancreatic islets, leading to insulin deficiency. Onset of symptoms frequently occurs in childhood. Untreated diabetes leads to many long-term complications, including those of ophtalmologic nature. The eye problems are caused by damage to the blood vessels in the retina[1] of the eye, which can lead to gradual vision loss and blindness. This condition is known as diabetic retinopathy (DR).

Study Design

A study was conducted in the University Hospital Motol to assess the changes in the corneal[2] cells of DM1 patients, and their associations with the degree of diabetic retinopathy. The study included 60 DM1 patients and 20 healthy subjects. DM1 patients were divided into three groups according to the degree of DR (none, mild, severe). Corneal cells of all subjects were evaluated using in vivo confocal microscopy.

There were several variables of interest ("outcomes") describing (i) cell density in various layers of cornea, and (ii) density and shape of nerve fibers in the cornea.

Cell density variables were

- Basal epithelial cell density (cells at the thin surface layer of the cornea).
- Anterior stroma cell density (cells from the thick middle layer of the cornea)

These variables were measured at one location within each eye. They are expressed as the number of cells observed in 1 mm^2 of a corneal layer.

Nerve fiber variables were

- Nerve fiber length, i.e., the total length of all nerve fibers observed in a selected spot with area of 0.1 mm^2.
- Nerve fiber density, i.e., the number of nerve fibers observed in that spot.
- Nerve fiber branch density, i.e., the number of nerve branchings observed in that spot.

These variables were measured at two locations within each eye. Thus, each participant had two measurements of each cell density variable and four measurements of each nerve fiber variable.

Dataset

The dataset comes in R format and includes three dataframes. Dataframe `subj` includes information on subjects. Dataframe `celld` contains corneal cell density measurements. Dataframe `nfib` contains corneal nerve fiber measurements. The dataframes are linked together by a unique subject ID. A detailed description of these dataframes is included at the end of this assignment.

The dataset also contains a named character vector `varlab` that stores variable labels. E.g., the label for variable `v` can be accessed by `varlab["v"]`.

[1] sítnice
[2] rohovka

The Task

Conduct an appropriate analysis that answers the following questions. The outcomes are the two corneal cell densities (basal epithelial cell density and anterior stroma cell density) and the three characteristics of corneal nerve fibers (nerve fiber length, nerve fiber density, and branch density).

1. Do the outcome means differ between DM1 patients and healthy subjects?
2. Do the outcome means depend on age, duration of diabetes, and level of glycated hemoglobin[3] (HbA1c)? If so, how?
3. Are the outcome means associated with the degree of diabetic retinopathy?

Requirements

Create an electronic report in the pdf format summarizing the solution to the problem, the results and conclusions. Describe the manipulation with the data set that preceded the analysis (variable transformations and recoding, treatment of potential missing values, omitted observations). Include a concise descriptive analysis (descriptive tables, figures) targeted towards answering the questions of interest. Describe the model(s) used to address the study questions and explain how the model was selected. Display the model formula in the report, list the assumptions. Justify the validity of the model (is there a better model than this? do all the assumptions hold?) Provide *an explicit answer to each question of interest*; show estimated effects of important variables with confidence intervals; include p-values for relevant hypotheses. Explain which methods were used to obtain the results. Include a short discussion of your approach to the analysis (its strengths and weaknesses) and of the meaning and reliability of the results. Attach in a separate file the R code used to perform data manipulations and to obtain the results included in the report.

Due date: Two working days before the scheduled date of project evaluation.

[3] high levels of glycated hemoglobin indicate poor long-term control of blood sugar

Variable Coding Tables

Dataframe `subj`

This dataframe includes 80 observations and 7 variables.

Variable Name	Variable Label	Variable Coding/Units
pid	Participant ID	integer
age	Participant's age	years
dg.yr	Year of DM1 diagnosis	integer
dmdur	Duration of diabetes	integer
cc	Diabetes status	factor, 2 levels
grp	Diabetic retinopathy group	factor, 4 levels
hba1c	Glycated hemoglobin	mmol/mol

Dataframe `celld`

This dataframe includes 160 observations and 4 variables.

Variable Name	Variable Label	Variable Coding/Units
pid	Participant ID	integer
eye	Which eye (L/R)	factor
epit	Basal epithelial cell density	cells/mm^2
ps	Anterior stroma cell density	cells/mm^2

Dataframe `celld`

This dataframe includes 320 observations and 6 variables.

Variable Name	Variable Label	Variable Coding/Units
pid	Participant ID	integer
eye	Which eye (L/R)	factor
occ	Occasion (1/2)	integer
nfl	Nerve fiber length	100 μm per mm^2
hn	Nerve fiber density	num. fibers per 0.1 mm^2
v	Nerve fiber branch density	num. branches per 0.1 mm^2