

Mini Project

Statistical Learning Theory

Prof. G. De Nicolao

Academic Year 2020/2021

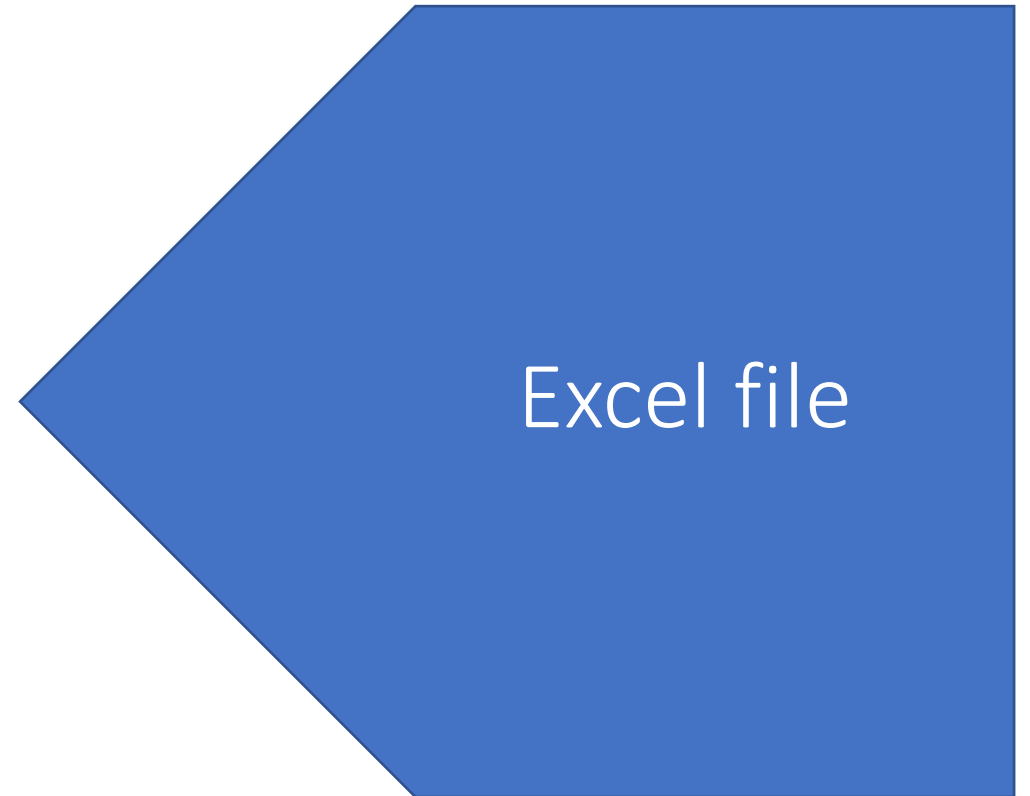
CASE STUDY 1

In a hypothetical toxicological in-vivo experiment (toxicity study) a substance X was administered to laboratory animals and three endpoints (endpoint 1, 2 and 3) were measured to define the reference point in order to set up health based guidance values (HBGV) for substance X. Eight groups of animals were tested, 4 groups of males and 4 groups of females. For each gender, one group was used as a control and each remaining group received one of three doses of substance X (30 mg/kg, 100 mg/kg and 150 mg/kg, administered to each animal daily). The endpoint values were measured for each animal 90 days after the beginning of the experiment. Information on the animals used in the experiment as well as summary information (mean and standard deviation, SD for each group of animals) on the endpoints measured at the end of the experiment is provided in **table 2**. It is requested to propose an appropriate statistical model and to analyse these data.

Table 2

Dose (mg/kg)	gender	Number of animals	Endpoint 1 mean	Endpoint 1 SD	Endpoint 2 mean	Endpoint 2 SD	Endpoint 3 mean	Endpoint 3 SD
0	male	9	570.4	75.6	45.2	1.5	18.3	0.5
30	male	10	548.5	44.3	44.6	1.7	17.9	0.5
100	male	10	533	72.3	43	3.8	16.5	0.6
150	male	9	488.1	42.5	42.8	1.7	16.3	0.8
0	female	10	329.8	46.8	44.7	1.3	19.1	0.5
30	female	10	319.7	24.3	44.1	1.6	19	0.4
100	female	10	293.3	20.1	41.5	1.6	17.9	0.5
150	female	10	312.4	43.1	42.1	1.2	17.3	0.6

response	number of animals	SD	dose	sex (0=M, 1=F)	endpoint
570.4	9	75.6	0	0	1
548.5	10	44.3	30	0	1
533	10	72.3	100	0	1
488.1	9	42.5	150	0	1
329.8	10	46.8	0	1	1
319.7	10	24.3	30	1	1
293.3	10	20.1	100	1	1
312.4	10	43.1	150	1	1
45.2	9	1.5	0	0	2
44.6	10	1.7	30	0	2
43	10	3.8	100	0	2
42.8	9	1.7	150	0	2
44.7	10	1.3	0	1	2
44.1	10	1.6	30	1	2
41.5	10	1.6	100	1	2
42.1	10	1.2	150	1	2
18.3	9	0.5	0	0	3
17.9	10	0.5	30	0	3
16.5	10	0.6	100	0	3
16.3	9	0.8	150	0	3
19.1	10	0.5	0	1	3
19	10	0.4	30	1	3
17.9	10	0.5	100	1	3
17.3	10	0.6	150	1	3



To do

1. Plot dose-response data for each pair endpoint-gender (6 plots) with error bars reflecting error size on response measurements.
2. Use subset selection to estimate separate models for the 3 endpoints using gender as categorical variable.
3. Use subset selection to estimate a unique model using gender and endpoint as categorical variables.
4. In points 2. and 3., heteroscedasticity of the data should be taken into account.
5. To be completed before Christmas holidays.