

MATH 204 - ASSIGNMENT 2

Deadline: 11.59pm, 7th April 2026

Please submit your solutions via myCourses in pdf format as a single pdf file.

Note: You may include code and output to explain and justify your answers. If you do not include code, you must explain in detail how the numerical results and conclusions were derived.

1. In a study of 111 patients who have undergone a kidney transplant and subsequently experienced organ rejection, the time (in months) until signs of rejection were diagnosed was recorded. The influence of genetic factors on time to rejection was studied. Two predictors were measured at time of transplantation:
 - **Genotype:** The genotype of the patient at a specific genomic locus. With alleles G and g, the three possible genotypes forming the three levels of this **factor predictor** are
 - gg (labelled 0)
 - Gg (labelled 1)
 - GG (labelled 2)
 - **Expression:** The level of expression of a second gene believed to be involved in immune defence. This variable is a **continuous covariate**.

The response (or dependent) variable is **Time**, the time in months until rejection is diagnosed.
The data are available at

<https://www.math.mcgill.ca/dstephens/204/Assignments/A2/Rejection.csv>

Using the techniques studied in lectures for subgroup analyses, find the most suitable model for predicting time until rejection based on the genetic information. Assess the fit of your selected model.

- | | |
|---------------------|---------|
| (a) Model fitting | 8 Marks |
| (b) Model selection | 8 Marks |
| (c) Model checking | 4 Marks |

```
#R Code for data entry and elementary processing
fname<-"https://www.math.mcgill.ca/dstephens/204/Assignments/A2/Rejection.csv"
Rejection<-read.csv(fname)
Rejection$Genotype<-as.factor(Rejection$Genotype)
str(Rejection)

## 'data.frame': 111 obs. of 3 variables:
## $ Genotype : Factor w/ 3 levels "gg","Gg","GG": 1 1 1 1 1 1 1 1 1 1 ...
## $ Expression: num 8.04 7.79 8.08 6.68 8.23 ...
## $ Time : num 17 18.5 19.1 19 20.4 ...

nrow(Rejection)

## [1] 111

table(Rejection$Genotype)

##
## gg Gg GG
## 38 42 31
```

2. In an experimental study of the benefits of three dietary regimen on muscle development, 36 subjects aged 25-35 and 36 subjects aged 36-45 were recruited. After six weeks on study, the subjects underwent a strength test, and the result was recorded. At the end of study, the body mass of the subjects was also recorded. The variables recorded in the data set are

- **Age:** Age group
 - 25-35 (labelled 1)
 - 36-45 (labelled 2)
- **Regimen:** Dietary regimen followed
 - Diet 1
 - Diet 2
 - Standard
- **Mass:** body mass (in kg) at end of study

The response (or dependent) variable Y is the strength test result at end of study. The data are available at

<https://www.math.mcgill.ca/dstephens/204/Assignments/A2/DietStudy.csv>

Using the techniques studied in lectures, find the most suitable model for predicting test strength. Assess the fit of your selected model.

- | | |
|---------------------|---------|
| (a) Model fitting | 8 Marks |
| (b) Model selection | 8 Marks |
| (c) Model checking | 4 Marks |

```
#R Code for data entry and elementary processing
fname<-"https://www.math.mcgill.ca/dstephens/204/Assignments/A2/DietStudy.csv"
DietStudy<-read.csv(fname)
DietStudy$Age<-as.factor(DietStudy$Age)
DietStudy$Regimen<-factor(DietStudy$Regimen, levels=c('Standard','Diet 1','Diet 2'))
str(DietStudy)

## 'data.frame': 72 obs. of 4 variables:
## $ Age      : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ Regimen: Factor w/ 3 levels "Standard","Diet 1",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Mass    : num  67.2 64.8 68 66.9 69.5 69.1 70.3 64.4 65.8 64.1 ...
## $ Y       : num  198 201 198 196 201 ...

nrow(DietStudy)

## [1] 72

table(DietStudy$Age, DietStudy$Regimen)

##
##      Standard Diet 1 Diet 2
## 1          12     12     12
## 2          12     12     12
```

3. The anticoagulant heparin is used for preventing blood clots of patients after surgery, and for the treatment and prevention of deep vein thrombosis. The most commonly used form of the drug requires careful monitoring, as misuse may lead to possible adverse outcomes. In a real clinical study concerning the efficacy of heparin, levels of the metabolite heparin sulphate in the blood of eligible recruited patients were monitored.

The file

<https://www.math.mcgill.ca/dstephens/204/Assignments/A2/HeparinDataSet.csv>

contains information on 90 patients (classified in the study via biological sex as 37 female, 53 male).

The data file contains five columns:

- **Sex:** Sex of patient (Female or Male) ;
- **Age:** Age of patient (years);
- **Weight:** Weight of patient (kg);
- **Dose:** Initial dose of heparin received (Low or High);
- **Heparin:** level of heparin sulphate in blood sample, measured 3 hours after dosing (relative scale).

Using these data, after selecting a suitable model, address the following questions (justifying your answers in each case):

- Model selection step. 8 Marks
- Does the Age of the patient influence the level of heparin sulphate in the blood ? 4 Marks
- On average, do Male patients have a lower or higher level of heparin sulphate in their blood-stream compared to Female patients of the same weight and age, when both are dosed initially with a High dose ? 4 Marks
- Predict the level of heparin sulphate that would be detected in the blood of a 40 year old male, weighing 100 kg, who is dosed on the Low dose. 4 Marks

```
#R Code for data entry and elementary processing
fname<-"https://www.math.mcgill.ca/dstephens/204/Assignments/A2/HeparinDataSet.csv"
Heparin<-read.csv(fname)
Heparin$Sex<-factor(Heparin$Sex,levels=c('Female','Male'))
Heparin$Dose<-factor(Heparin$Dose,levels=c('Low','High'))
str(Heparin)

## 'data.frame': 90 obs. of 5 variables:
## $ Sex : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
## $ Age : int 51 48 65 44 43 54 44 36 43 55 ...
## $ Weight : num 146.9 115.1 83.5 94.9 122.5 ...
## $ Dose : Factor w/ 2 levels "Low","High": 2 2 1 2 2 1 1 2 2 1 ...
## $ Heparin: num -1.65 1.49 1.76 1.32 -2.86 ...

nrow(Heparin)

## [1] 90

table(Heparin$Sex, Heparin$Dose)

##
## Low High
## Female 22 31
## Male 15 22
```