

## Exam

### Question 1: A

R Code:

*Creating a function*

```
my_function = function(net_income, total_asset, shareholders_equity, net_sales) {  
  ROA = net_income / total_asset  
  ROE = net_income / shareholders_equity  
  profit_margin = net_income / net_sales  
  ratios = c(ROA, ROE, profit_margin)  
  names = c("ROA", "ROE", "Profit_margin")  
  final = cbind(ratios, names)  
  final  
}
```

# Question 1: B

```
my_function(10000, 60000, 40000, 1000000)
```

### Question 1: B

Output:

```
> my_function(10000, 60000, 40000, 1000000)  
      ratios      names  
[1,] "0.166666666666667" "ROA"  
[2,] "0.25"             "ROE"  
[3,] "0.01"             "Profit_margin"
```

As can be seen in the figure, the Return on Asset is 0.167, the Return on Equity is 0.25 and the profit margin is 0.01.

### Question 1: C

The Return on Asset (ROA) tells us the earnings generated from the company's asset, and the higher the better. It is difficult to say that if 0.167 is good for this company, because it depends on the company, the industry, the sector, region, types of customers, seasonal sales etc.

When it comes to Return on Equity, we want as investors to have a high return on equity. Again here, it is difficult to say if 0.25 is high or low, because we don't know the industry, sector etc., but usually we would compare the ROE with stock market return (for example S&P, Dow Jones, etc.). In this case 0.25 or 25% in return is a high return.

Finally, profit margin tells us the ratio between net income and net sales, and we want to have this ratio as high as possible. In this case, 0.01 seems to be low, but again it depends on the sector, industry, etc.

The conclusion is that I would invest in the company, which is mainly because of the high Return on Equity compared to the stock market. But it would be interesting to get historical data and compare the ratios with the company's competitors.

### Question 1: D

I would invest \$10,000 in the company. After my investment the asset (cash) will increase with \$10,000, and therefore the company's shareholder's equity will increase with \$10,000, and the new function will be:

```
# Before my investment
my_function(10000, 60000, 40000, 1000000)

# After my investment
my_function(10000, 70000, 50000, 1000000)
```

Comparison of the two } Therefore, the ROA will decrease from 0.167 to 0.142, the ROE will decrease from 0.25 to 0.20 and the profit margin will stay the same (0.01).

### Question 2: A

R Code

```
barto <- rexp(5000, rate = 0.8)
```

### Question 2: B

R Code

```
barto <- rexp(5000, rate = 0.8)

mean(barto)
var(barto)
```

Output

```
> mean(barto)
[1] 1.232584
> var(barto)
[1] 1.658553
```

### Question 2: C

R Code

```
exp(-new_lambda*5) # probability of higher than 5
1-exp(-new_lambda*3) # probability of lower than 3
exp(-new_lambda*3.5) # probability of higher than 3.5
```

Output

```
> exp(-new_lambda*5) # probability of higher than 5
[1] 0.01730915
> 1-exp(-new_lambda*3) # probability of lower than 3
[1] 0.9123069
> exp(-new_lambda*3.5) # probability of higher than 3.5
[1] 0.05845114
```

A)  $\exp(-\text{new\_lambda} * 5)$  # probability of higher than 5

---

0.017 = the probability that a passenger has to wait longer time than 5 is 0.017

B)  $1 - \exp(-\text{new\_lambda} * 3)$  # probability of lower than 3

---

0.912 = the probability that a passenger has to wait less than 3 is 0.912

C)  $\exp(-\text{new\_lambda} * 3.5)$  # probability of higher than 3.5

---

0.058 = the probability that a passenger has to wait more than 3.5 is 0.058

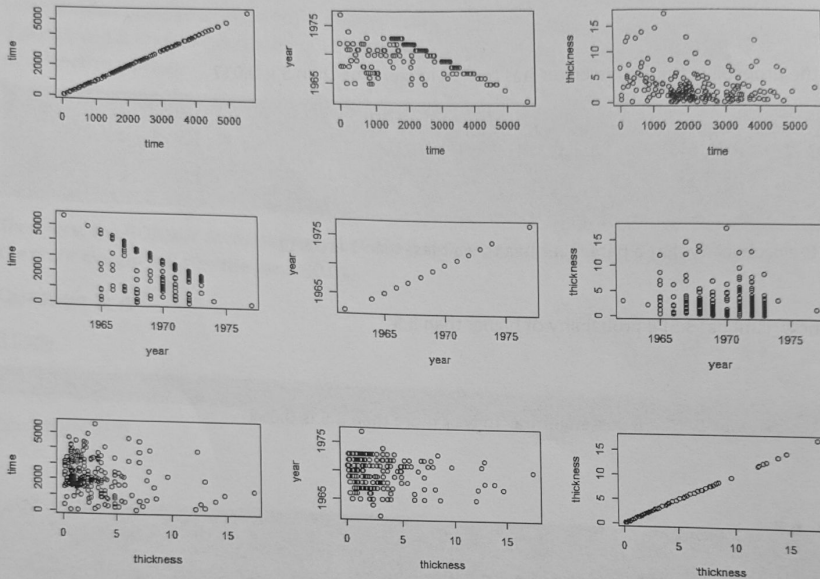
### Question 3

R Code:

```
library(MASS)
summary(Melanoma)
str(Melanoma)

my_df <- Melanoma[,c("time", "year", "thickness")]
matrix_dim <- ncol(my_df)
par(mfrow=c(3,3))
for (i in 1:ncol(my_df)){
  for (z in 1:ncol(my_df)){
    plot(x=my_df[,i], y=my_df[,z], xlab=colnames(my_df)[i],
         ylab=colnames(my_df)[z])
  }
}
```

Output



As can be seen in first graph in row 2; Y= time, X= year, there is a negative correlation between time and year, so as the years goes the time decrease.

#### Question 4: A

##### R Code

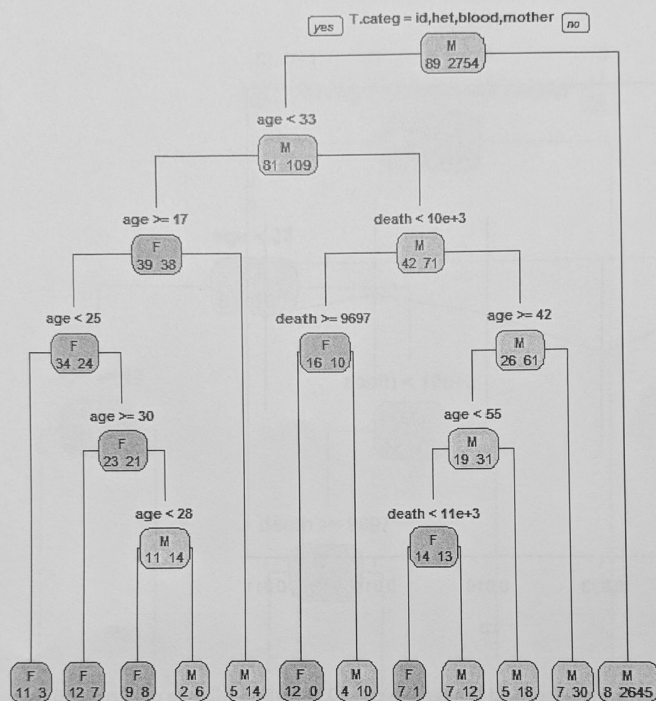
```
library(MASS)
library(rpart)
library(rpart.plot)
library(plotly)

summary(Aids2)

# Question 4:1

my_tree <- rpart(sex~age+T.categ+death, data = Aids2, method = "class")
rpart.plot(my_tree, type=1, extra=1, box.palette = c("pink", "green"))
```

##### Output



#### Question 4: B

##### R Code

```

library(MASS)
library(rpart)
library(rpart.plot)
library(plotly)

summary(Aids2)

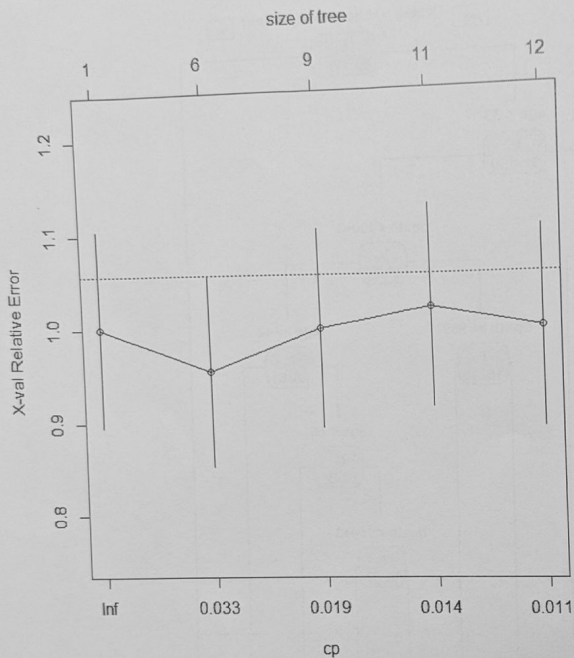
# Question 4:1

my_tree <- rpart(sex, age, T.categ, death, data = Aids2, method = "class")
rpart.plot(my_tree, type=1, extra=1, box.palette = c("pink", "green"))

plotcp(my_tree) # To see how big the tree should be. 0 = cp = 0.033

```

## Output



As can be seen in the figure, we should change the cp to 0.033 which will give us a tree of 6. We should therefore prune the tree.

## Question 4: C

### R Code

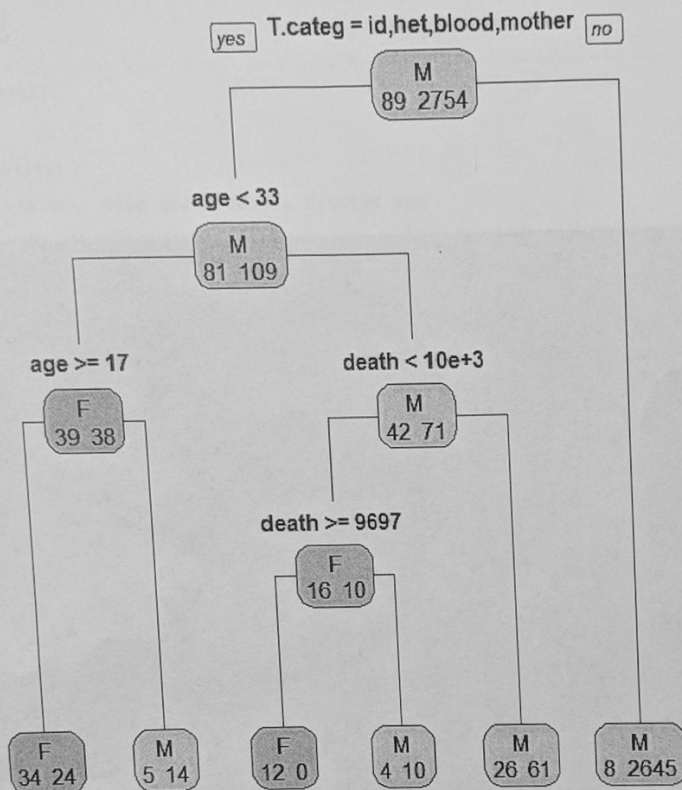
```
library(MASS)
library(rpart)
library(rpart.plot)
library(plotly)

summary(Aids2)

# Question 4:1

my_tree <- rpart(sex, age, T.categ, death, data = Aids2, method = "class", control = rpart.control(cp=0.033))
rpart.plot(my_tree, type=1, extra=1, box.palette = c("pink", "green"))
```

## Output



As can be seen in the classification tree, the split is between age and death, which means that these two variables are the two most important variables when it comes to classifying the transmission category.

If we look at the top of the tree, we can see that if there is a transmission or not. Here we see that for 2645 there were no transmission, but if there were a transmission, we can see that age lower and higher than 33 years is the first variable, to determine if it is a male or female.

For example, if the age is lower than 33 years old and death is higher than  $10e+3$ , there is a high probability of that being a male.



## Question 5

### R Code

```
title: "Exam - question 5"
author: "Kristian Kraack Nielsen"
date: "December 14, 2018"
output: html_document

## Aids 2

library(MASS)

head(Aids2,10)

summary(Aids2)

library(rpart)
library(rpart.plot)
library(plotly)

my_tree <- rpart(sex~age+T.categ+death, data = Aids2, method = "class", control =
rpart.control(cp=0.033))

rpart.plot(my_tree, type=1, extra=1,box.palette = c("pink","green"))
```

As can be seen in the classification tree, the split is between age and death, which means that these two variables are the two most important variables when it comes to classifying the transmission category.

If we look at the top of the tree, we can see that if there is a transmission or not. Here we see that for 2645 there were no transmission, but if there were a transmission, we can see that age lower and higher than 33 years is the first variable, to determine if it is a male or female. For example, if the age is lower than 33 years old and death is higher than 10e+3, there is a high probability of that being a male.

### Output

## Exam - question 5



Kristian Kraack Nielsen

December 14, 2018

## Aids 2

```
library(MASS)
```

```
head(Aids2,10)
```

```
##      state sex  diag death status T.categ age
## 1    NSW   M 10905 11081      D     hs    35
## 2    NSW   M 11029 11096      D     hs    53
## 3    NSW   M  9551  9983      D     hs    42
## 4    NSW   M  9577  9654      D    haem    44
## 5    NSW   M 10015 10290      D     hs    39
## 6    NSW   M  9971 10344      D     hs    36
## 7    NSW   M 10746 11135      D    other    36
## 8    NSW   M 10042 11069      D     hs    31
## 9    NSW   M 10464 10956      D     hs    26
## 10   NSW   M 10439 10873      D    hsid    27
```

```
summary(Aids2)
```

```
##      state      sex      diag      death      status
##  NSW :1780  F: 89  Min.   : 8302  Min.   : 8469  A:1082
## Other: 249  M:2754 1st Qu.:10163 1st Qu.:10672 D:1761
##  QLD  : 226      Median :10665  Median :11235
##  VIC  : 588      Mean   :10584  Mean   :10990
##                3rd Qu.:11103  3rd Qu.:11504
##                Max.   :11503  Max.   :11504
##
```

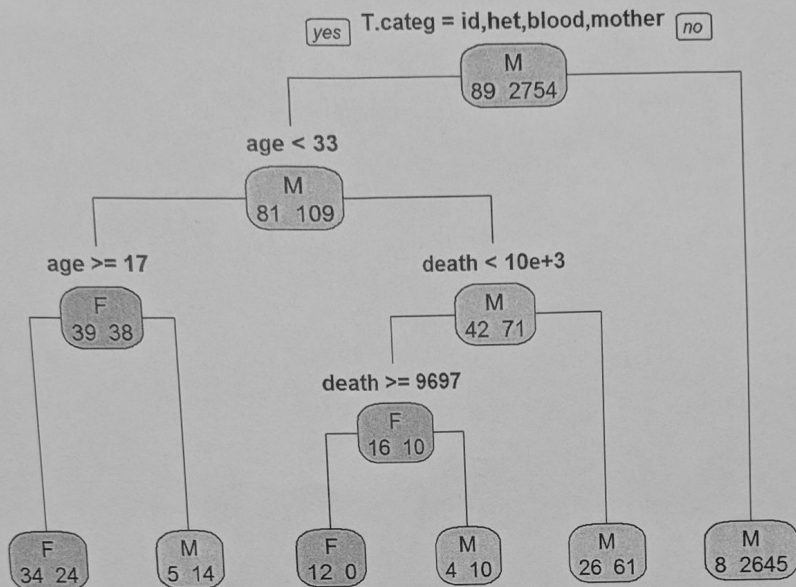
```
##      T.categ      age
##  hs      :2465  Min.   : 0.00
##  blood   : 94  1st Qu.:30.00
##  hsid    : 72  Median :37.00
##  other   : 70  Mean   :37.41
##  id      : 48  3rd Qu.:43.00
##  haem    : 46  Max.   :82.00
```

```

## (Other): 48
library(rpart)
library(rpart.plot)
library(plotly)
## Loading required package: ggplot2
##
## Attaching package: 'plotly'
## The following object is masked from 'package:ggplot2':
##
##     last_plot
## The following object is masked from 'package:MASS':
##
##     select
## The following object is masked from 'package:stats':
##
##     filter
## The following object is masked from 'package:graphics':
##
##     layout
my_tree <- rpart(sex~age+T.categ+death, data = Aids2, method = "class", control
= rpart.control(cp=0.033))

rpart.plot(my_tree, type=1, extra=1, box.palette = c("pink", "green"))

```



As can be seen in the classification tree, the split is between age and death, which means that these two variables are the two most important variables when it comes to classifying the transmission category. If we look at the top of the tree, we can see that if there is a transmission or not. Here we see that for 2645 there were no transmission, but if there were a transmission, we can see that age lower and higher than 33 years is the first variable, to determine if it is a male or female. For example, if the age is lower than 33 years old and death is higher than  $10e+3$ , there is a high probability of that being a male.