Business Intelligence II: Summarisk løsning af eksamenscase 2025

Indlæsning af data:

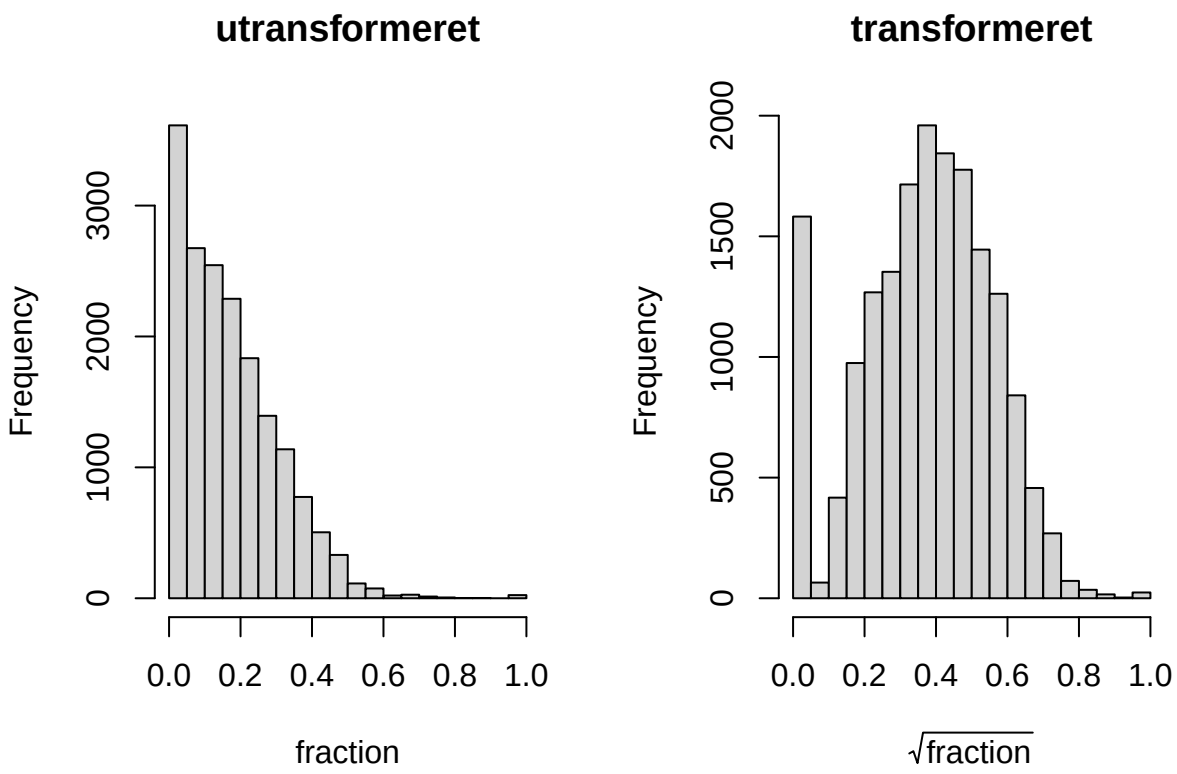
```
rentalbikes <- read.csv('rentalbikes.csv')
str(rentalbikes)
```

```
## 'data.frame':    17379 obs. of  12 variables:
##  $ fraction   : num  0.188 0.2 0.156 0.231 0 ...
##  $ season     : chr  "spring" "spring" "spring" "spring" ...
##  $ year       : int  2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 ...
##  $ month      : chr  "Jan" "Jan" "Jan" "Jan" ...
##  $ weekday    : chr  "Saturday" "Saturday" "Saturday" "Saturday" ...
##  $ hour       : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ holiday    : chr  "No" "No" "No" "No" ...
##  $ temperature: num  9.84 9.02 9.02 9.84 9.84 ...
##  $ feelslike  : num  14.4 13.6 13.6 14.4 14.4 ...
##  $ windspeed  : int  0 0 0 0 0 6 0 0 0 0 ...
##  $ humidity   : int  81 80 80 75 75 75 80 86 75 76 ...
##  $ weather    : chr  "clear" "clear" "clear" "clear" ...
```

Nogle grafer og nogle summaries:

Fordelingen af `fraction` med og uden kvadratrodstransformation:

```
par(mfrow=c(1,2))
# histogrammer:
hist(rentalbikes$fraction, main='utransformeret', xlab='fraction')
hist(sqrt(rentalbikes$fraction), main='transformeret', xlab=expression(sqrt(fraction)))
```



Ekstreme værdier:

```
# antal observationer (timer) hvor _ingen_ brugere er "lejlighedsvise"
sum(rentalbikes$fraction==0)
```

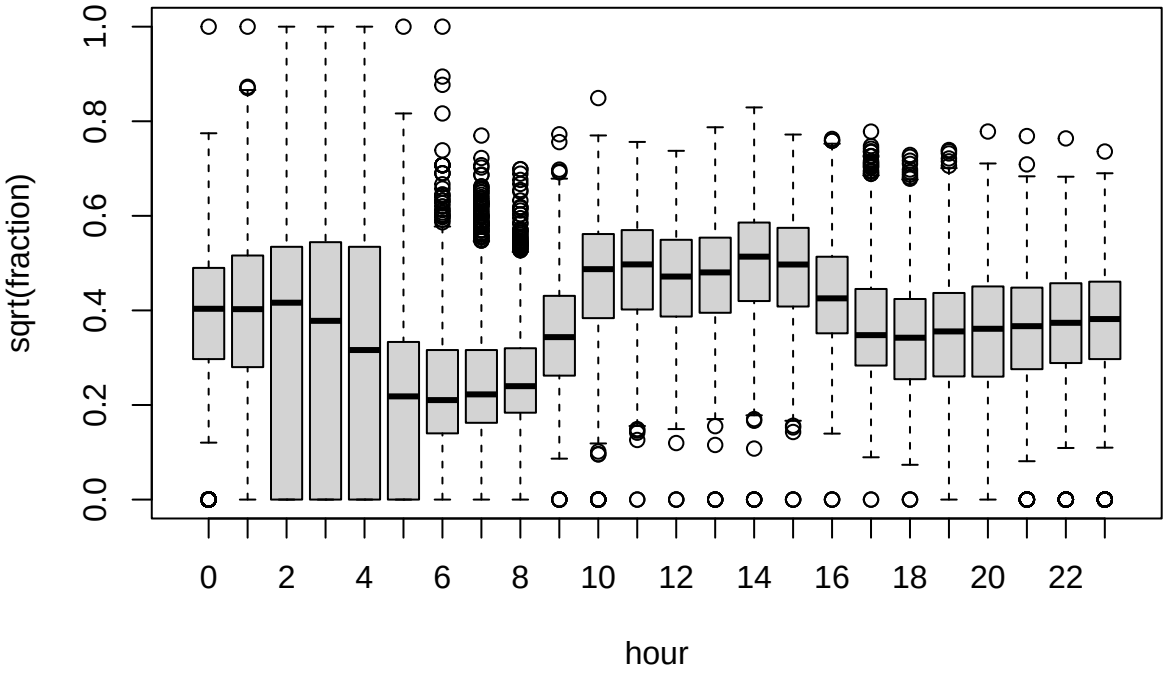
```
## [1] 1581
```

```
# antal observationer (timer) hvor _alle_ brugere er "lejlighedsvise"
sum(rentalbikes$fraction==1)
```

```
## [1] 24
```

Boxplots af `fraction` som funktion af `hour`:

```
boxplot(sqrt(fraction)~hour, rentalbikes)
```



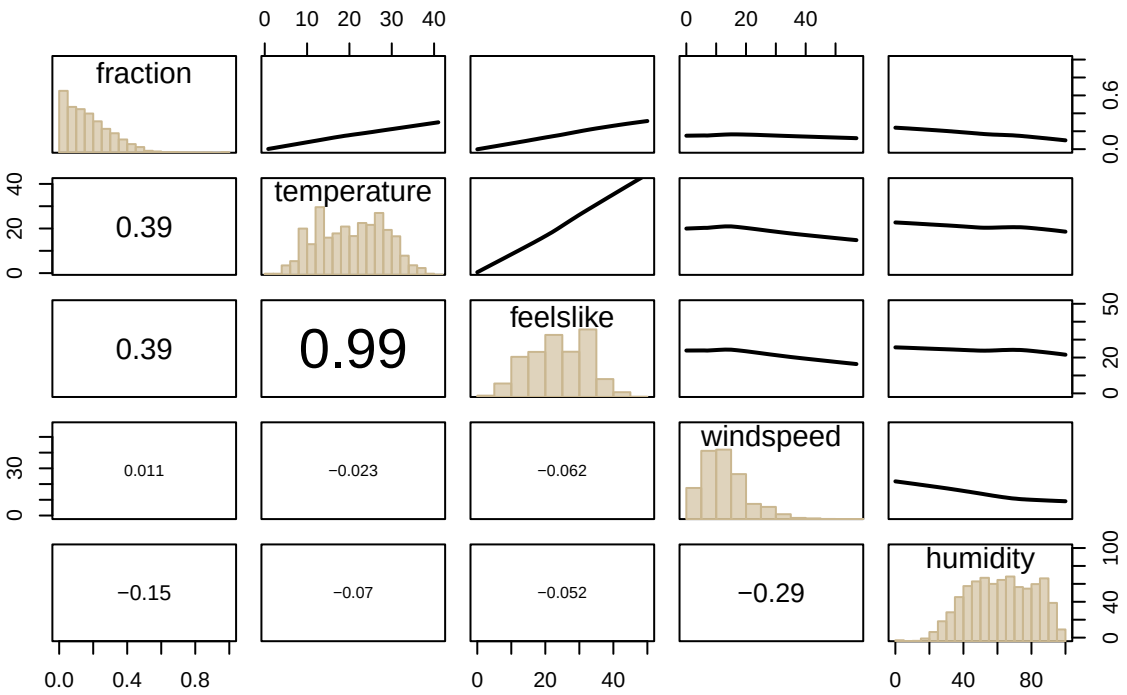
Sammenhængen mellem `fraction` og `hour` synes ikke-monoton.

Plots af `fraction` og vejrmæssige variable:

```
library(DescTools)
```

```
## Loading required package: stats
```

```
PlotPairs(rentalbikes[, c('fraction', 'temperature', 'feelslike',
                                    'windspeed','humidity')] )
```



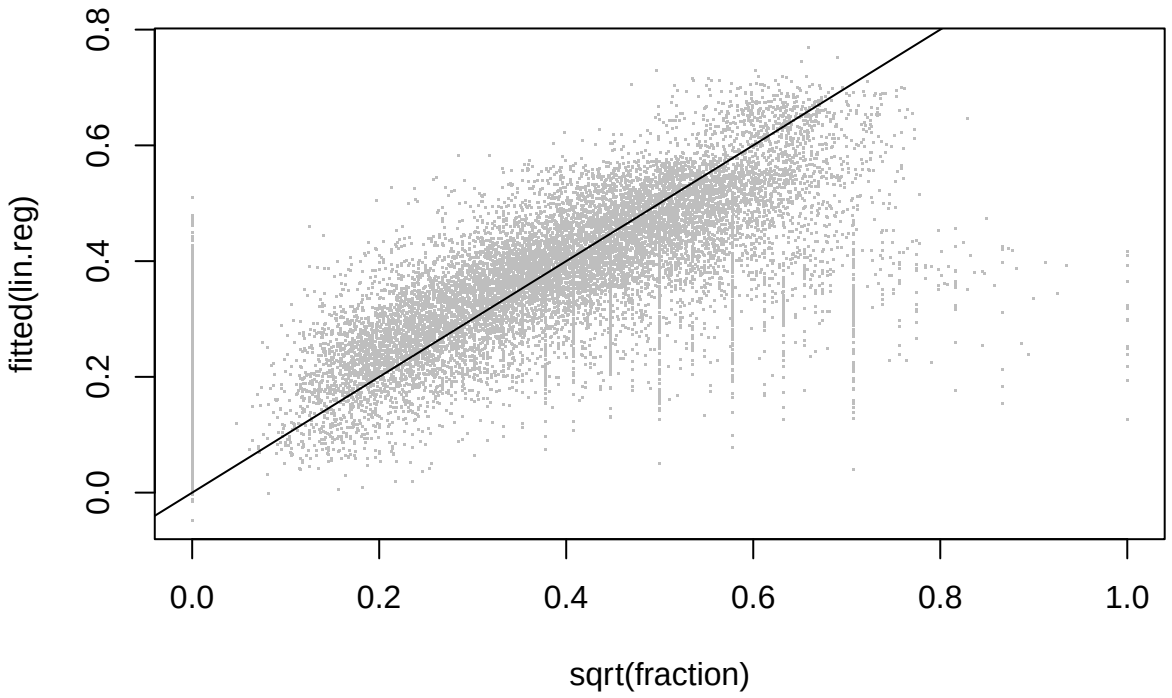
To simple modeller

Opdeling i træning og test:

```
set.seed(...)
index <- sample(1:nrow(rentalbikes),...)
train <- rentalbikes[index,]
test <- rentalbikes[-index,]
```

Lineær regression:

```
lin.reg <- lm(sqrt(fraction) ~ . - hour + factor(hour), data=train)
# plot af fittede værdier mod de observerede
plot(fitted(lin.reg)~sqrt(fraction), train, pch='.', col='grey')
# diagonalen; hvor punkterne ville ligge, hvis fittet var perfekt
abline(0,1)
```




```
# test data RMSE
DescTools::RMSE(predict(lin.reg, newdata=test), sqrt(test$fraction))
```

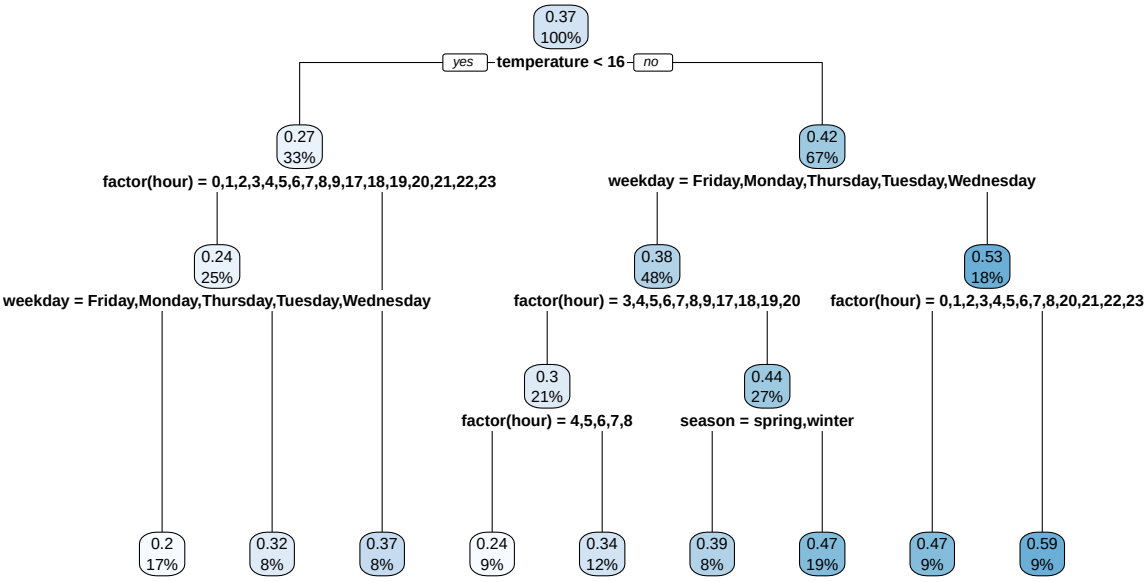
```
## [1] 0.134613
```

Regressionstræ:

```
library(rpart)
```

```
## Loading required package: grDevices
```

```
træ <- rpart(sqrt(fraction) ~ . - hour + factor(hour), data=train)
library(rpart.plot)
rpart.plot(træ)
```




```
# test data RMSE
DescTools::RMSE(predict(træ, newdata=test),sqrt(test$fraction))
```

```
## [1] 0.1443292
```