Exercise 1 (Poisson distribution)

During measurements through Positron emission tomography (PET) a detector produces "clicks". These clicks are counted for each volumetric pixel (voxel). The counts of different voxels (v1,v2) from different PET machines (a,b,c) are stored in the dataset. The data can be downloaded from the website (**R-Hint**: `read.table(..., sep=",", header=TRUE)`).

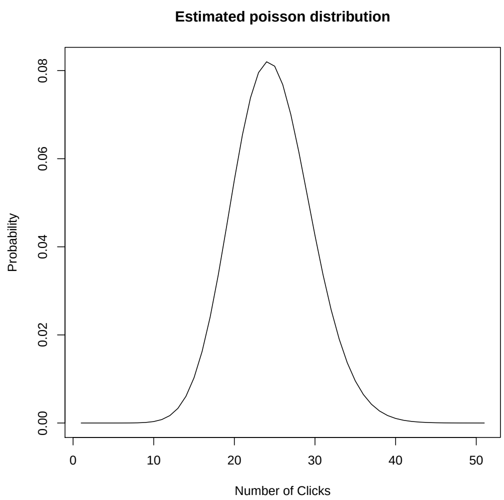
(a) Since the data contains count data (number of "clicks" per voxel), we fit a poisson model. The poisson model is specified by the value of the parameter λ (Lambda).

- Estimate the parameter λ by calculating the mean over the counts of `a.v1`.

```
# Read data
dat <- read.table(paste0(dir,'data/pet_counts.csv'), header=TRUE, sep=',')
# I defined "dir" earlier which containes the complete path to the
# folder "data" where the file "pet_counts.csv" is stored
# (dir <- "C:/.../"). Paste0() combines dir and "data/pet_counts.csv"
# to "C:/.../data/pet_counts.csv".

# Estimate lambda as the mean over the counts
lambda <- mean(dat$a.v1)
```

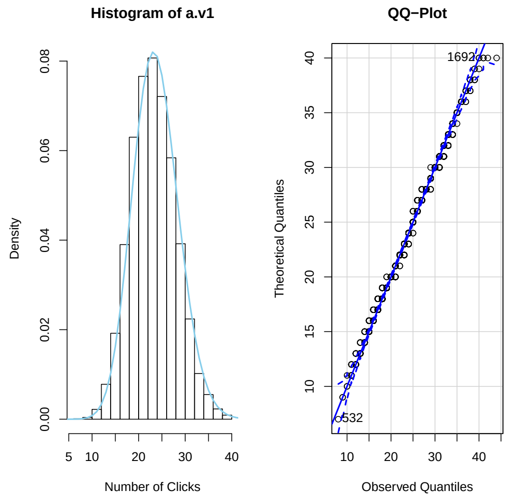
- Plot the distribution of the counts, that you would expect from the poisson model (**R-Hint**: `plot(dpois(0:50,lambda=...),...)`)

```
# dpois calculates the poisson distribution with the parameter lambda
poi <- dpois(0:50, lambda=lambda)

plot(poi, type = "l", ylab = "Probability",
     xlab = "Number of Clicks", main = "Estimated poisson distribution")
```



(b) Find out how well the poisson model describes the data. Plot the distribution of the observed data `a.v1` as histogram together with the poisson distribution. Additionally use a QQ-Plot to compare the expected to the observes quantiles of the counts. (**R-Hint**: `hist(..., probability=TRUE), lines(dpois(...))`. For the qqplot use the function `qqPlot()` from library `car`. If the library is not installed used `install.packages("car")` and `library(car)`. If the library is installed, you can load it directly with `library(car)`)

```
par(mfrow=c(1,2))

# plot the distbrution of a.v1 with a histogram and add the expected
# poisson distribution
hist(dat$a.v1, probability = TRUE, main = "Histogram of a.v1",
     xlab = "Number of Clicks")
lines(0:50, dpois(0:50, lambda = lambda), col = "skyblue", lwd = 2)

# Use a QQ-plot:
# The QQ-Plot shows the observed vs. the theoretical quantiles (which are
# presented on the axes). The linear line represents the optimal case,
# in which the observed match the theoretical quantiles.
# The dashed line shows the 95% confidence band, i.e. if we would sample
# from the poisson distr, only 5% of the data should lie outside that band.
# That is, most of the points should lie within that band if the poisson
# is appropriate.
library(car)

## Loading required package:  carData

qqPlot(dat$a.v1, dist = "pois", lambda = lambda, main = "QQ-Plot",
       ylab = "Theoretical Quantiles", xlab = "Observed Quantiles")
```

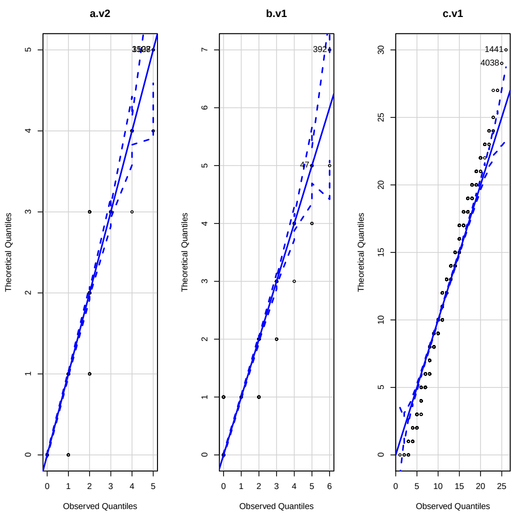



```
## [1]  532 1692

# The distribution plot and the QQ-Plot show that the poisson model
# is appropriate to model the variable a.v1.
```

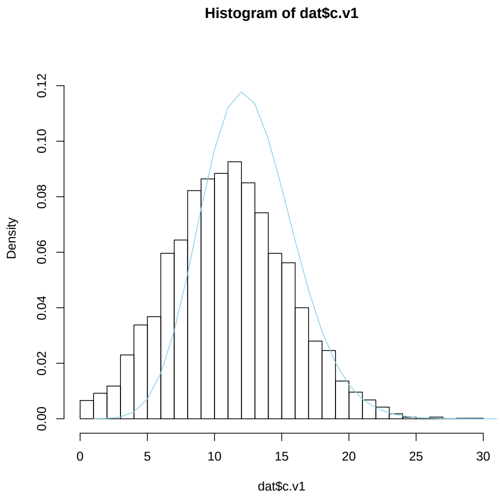
(c) The literature on PET states that a simple poisson model is not always well suited due to overdispersion, i.e. the variance in the data is higher than assumed in the poisson model. Consider the mean and the variance for the other three variables (`a.v2, b.v1, c.v1`). Do you think simple poisson models are appropriate? Verify your assumptions with QQ-plots.

```
# calculate the mean and variance for the three variables
c("a.v2" = mean(dat$a.v2), "b.v1" = mean(dat$b.v1), "c.v1" = mean(dat$c.v1))

##    a.v2    b.v1    c.v1
##  0.6946  0.9020 11.5546

c("a.v2" = var(dat$a.v2), "b.v1" = var(dat$b.v1), "c.v1" = var(dat$c.v1))

##       a.v2       b.v1       c.v1
##  0.6990707  0.8869734 19.1536496

# The poisson model assumes equal mean and variance. Therefore,
# a.v2 and b.v1 can be modelled with a simple poisson model.
# For c.v1, the variance is much higher than the mean and we expect
# a worse fit.
```

```
# We use QQ-Plots to verify the assumptions:
par(mfrow=c(1,3))
qqPlot(dat$a.v2, dist = "pois", lambda = mean(dat$a.v2), main = "a.v2",
       ylab = "Theoretical Quantiles", xlab = "Observed Quantiles")

## [1] 1107 3528

qqPlot(dat$b.v1, dist = "pois", lambda = mean(dat$b.v1), main = "b.v1",
       ylab = "Theoretical Quantiles", xlab = "Observed Quantiles")

## [1] 392  47

qqPlot(dat$c.v1, dist = "pois", lambda = mean(dat$c.v1), main = "c.v1",
       ylab = "Theoretical Quantiles", xlab = "Observed Quantiles")
```




```
## [1] 1441 4038

# As expected, the poisson model seems to be appropriate for the variables
# a.v2 and b.v1. The third plot shows, that most of the points lie outside
# the 95% confidence band. A simple poisson model would therefore be
# inappropriate to model the variable c.v1. You can additionally look at
# the density in order to understand what the QQ-plot shows:
par(mfrow=c(1,1))
hist(dat$c.v1, probability = TRUE, ylim=c(0,0.13), breaks=40)
lines(dpois(min(dat$c.v1):max(dat$c.v1),lambda=mean(dat$c.v1)), col='skyblue')
```



Exercise 2 (Binomial distribution)

Imagine that there is a new treatment offering a relief for 70% of the patients. This treatment is given to 100 new patients.

(a) What is the probability that the symptoms disappear in exactly 60 patients (**R-Hint**: `dbinom()`)?

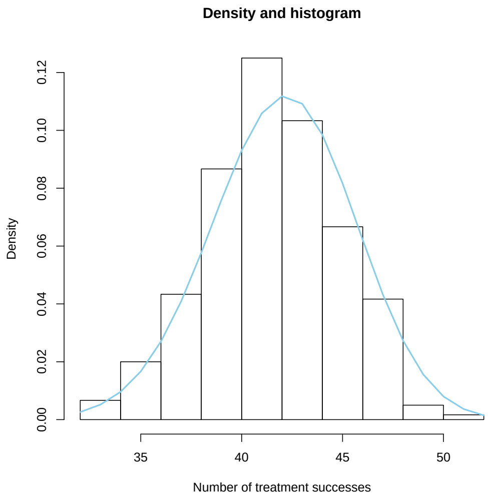
```
dbinom(60, 100, 0.7)

## [1] 0.008490169
```

(b) Simulate 300 data points for the situation described above (**R-Hint**: `rbinom()`)?

```
sim <- rbinom(300, 60, 0.7)
```

(c) Visualize both the simulated data and the underlying theoretical distribution.

```
par(mfrow=c(1,1))
hist(sim, probability = TRUE, xlab = "Number of treatment successes",
     main = "Density and histogram")
x <- seq(min(sim), max(sim))
lines(x, dbinom(x, 60, 0.7),col = "skyblue", lwd = 2)
```


Density and histogram
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
35
40
45
50
Number of treatment successes