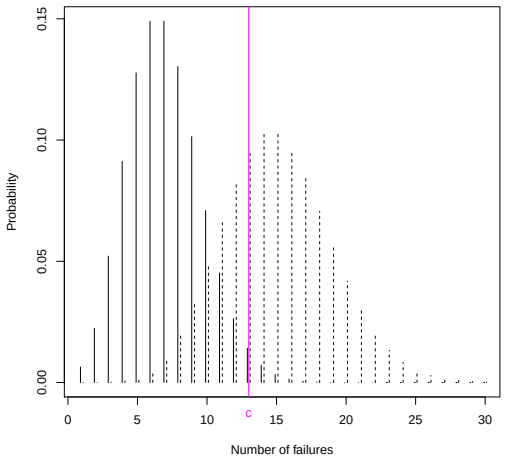
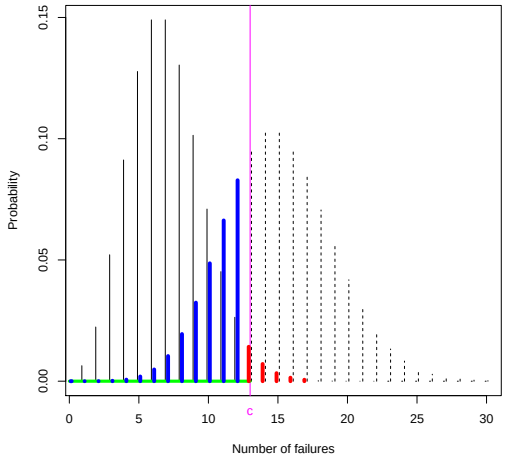
Exercise 1

With the random variable X, we denote the number of failures of some machine in a certain time range. We model X with a Poisson distribution: $X \sim \text{Pois}(\lambda)$. Before an examination of the failure rate, a null and an alternative hypothesis have been formulated. In the plot below, you find the distribution of X for the null H_0 (solid lines) and the alternative hypothesis H_A (dotted lines), respectively. The critical value of the test is $c = 13$ (Significance level: $\alpha = 0.05$; one-sided test; c belongs to the region of rejection).



(a) Draw the region of acceptance of the null hypothesis into the plot (in green), and indicate how the probability of a type I error (in red) or a type II error (in blue) can be read off the plot, respectively.



(b) Indicate the null and the alternative hypothesis (approximately) by looking at the plot.

```
# We have a poisson model, i.e. the parameter of interest is
# lambda (the mean of the distributions). Therefore, the null
# hypothesis is lambda=7, the alternative is lambda=14.
```

(c) Decide if the following statements are true or false.

☐ The power of the test decreases when the significance level of the test increases.

```
# False, if the significance level is increased, the critical
# point is shifted to the left and therefore the power increases.
```

☐ In the example in the plot, the alternative $\lambda = 20$ has a larger type II error than the alternative $\lambda = 15$.

```
# False, in this case the mean of the alternative would
# be shifted to the right and therefore the type II error
# decreases.
```

☐ By increasing the critical value to $c = 14$ in the plot, the power of the test decreases.

```
# True, if the critical value is shifted to the right, power decreases.
```

Exercise 2 (Sample size calculation)

A study is to be carried out in a rural area of Africa to investigate whether food supplementation during pregancy increases birth weight. Women attending the clinic are randomly assigned to either receiving food supplmenetation or not. It was decided to use a two sided t-test in order to measure the mean difference in birth weight. A clinically relevant increase would be 0.25 kg. Past data suggest that the common standard deviation of birth weight is 0.4 kg. Within the study we would like to have a power of 80% and a type one error rate of 5%.

(a) Calculate the number of samples you will need to find an effect (**R-Hint**: `power.t.test(...)`).

```
power.t.test(power=0.8, delta=0.25, sd=0.4 , alternative="two.sided",
             sig.level=0.05)

##
##      Two-sample t test power calculation
##
##               n = 41.16902
```

```
##           delta = 0.25
##              sd = 0.4
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group

# In order to find a significant result we need at least n=42 people
# within each group.
```

(b) Assume that we additionally want to split the data in three subgroups. Each subgroup is again evaluated using a t-test testing for a difference in birth weight. How many women do we have to recruit in order to reject the null at a 5% level?
(**Hint**: Use a Bonferroni correction in `power.t.test(...)`)

```
# Since we do multiple testing, we have to apply bonferroni correction.
# we therefore divide the p-value by 3 (3 subgroups).
power.t.test(power=0.8, delta=0.25, sd=0.4 , alternative="two.sided",
             sig.level=(0.05/3))

##
##      Two-sample t test power calculation
##
##               n = 55.05602
##           delta = 0.25
##              sd = 0.4
##       sig.level = 0.01666667
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group

# Instead of n=42 people, we have to recruit n=55 patients
# within each group.
```

Exercise 3 (Multiple testing)

The data from the breast cancer gene expression study of Hedenfalk et al. (2001) were obtained and analyzed. A comparison was made between 3,226 genes of two mutation types,

BRCA1 (7 arrays) and BRCA2 (8 arrays). The data included here are p-values, test-statistics, and permutation null test-statistics obtained from a two-sample t-test analysis on a set of 3170 genes, as described in Storey and Tibshirani (2003).

In order to analyse the data (`hedenfalk`), you have to install the `qvalue` package from bioconductor (open source software for bioinformatics). Use the following code to install thepackage and load the data:

```
# Install qvalue package from bioconductor
source("https://bioconductor.org/biocLite.R")
biocLite("qvalue")
```

```
# load the package and the data
library(qvalue)
data(hedenfalk)
```

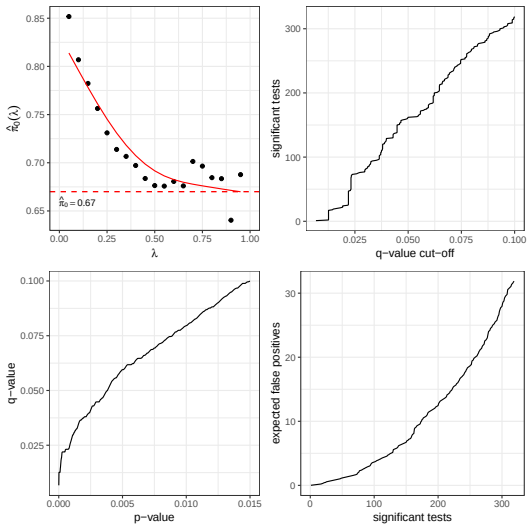
(a) Estimate the false discovery rates (q-values) using the function `qvalue`.

```
q <- qvalue(p = hedenfalk$p)
str(q)
```

```
## List of 8
##  $ call      : language qvalue(p = hedenfalk$p)
##  $ pi0       : num 0.67
##  $ qvalues   : num [1:3170] 0.0882 0.2094 0.668 0.1616 0.6325 ...
##  $ pvalues   : num [1:3170] 0.0121 0.075 0.9949 0.0418 0.8458 ...
##  $ lfdr      : num [1:3170] 0.168 0.413 1 0.309 1 ...
##  $ pi0.lambda: num [1:19] 0.852 0.807 0.782 0.756 0.731 ...
##  $ lambda    : num [1:19] 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 ...
##  $ pi0.smooth: num [1:19] 0.814 0.796 0.778 0.761 0.745 ...
##  - attr(*, "class")= chr "qvalue"
```

(b) Plot the values using `plot(qvalue(...))`. You should see four plots. Interpret them.

```
plot(q)
```




```
# Top left: The points show how often a specific p-value
# was observed. The dashed line shows the estimated cut off
# to decide if the a significant test is truly positive (above the line)
# or falsely positive (below the line).

# Top right: q-value vs. significant tests
# It shows the number of significant tests with varying q-value
# cut offs. The higher the q-value cut off, the more significant
# tests are expected.

# Bottom left: p-value vs. q-value
# The plot allows to decide which significance niveau can be chosen
# in order to increase/decrease the FDR. The higher the niveau,
# the higher the number of false positive signifant tests.

# Bottom right: significance tests vs. expected FP
# The higher the number of significant tests, the higher the number
# of expected false positive tests. For instance, if there are 300
# significant tests, we expect about 28 to be false positives.
```