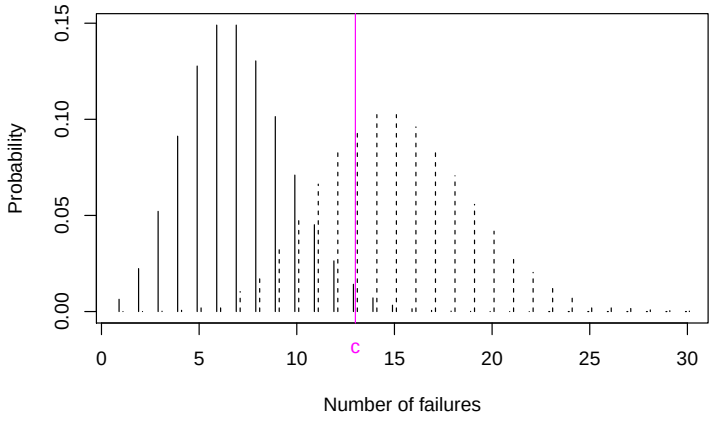
Biostatistics: Exercise 05

Beate Sick, Lisa Herzog

13.10.2020

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).



- 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.
- Indicate the null and the alternative hypothesis (approximately) by looking at the plot.
- Decide if the following statements are true or false.
  - The power of the test decreases when the significance level of the test increases.
  - In the example in the plot, the alternative $\lambda = 20$ has a larger type II error than the alternative $\lambda = 15$.
  - By increasing the critical value to $c = 14$ in the plot, the power of the test 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 pregnancy increases birth weight. Women attending the clinic are randomly assigned to either receiving food supplementation 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%.

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

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 3226 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 the package and load the data:

```
# Install qvalue package from bioconductor
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install(version = "3.11")
BiocManager::install("qvalue")
```

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

- Estimate the false discovery rates (q-values) using the function `qvalue()`.
- Plot the values using `plot(qvalue(...))`. You should see four plots. Interpret them.