Exam exercise: Berkely admission data

You may use the combined lecture notes for this module available at https://asta.math.aau.dk to guide you to the relevant methods and R commands for this exam.

Remember to load the `mosaic` package first:

```
library(mosaic)
```

The following table shows the total number of admitted and rejected applicants to the six largest departments in Berkeley in 1973.

	Admitted	Rejected
Male	1198	1493
Female	557	1278

Use a χ^2-test to check whether the admission statistics for Berkeley show any sign of gender discrimination. To enter the table in R you can do:

```
admit <- matrix(c(1198, 557, 1493, 1278), 2, 2)
rownames(admit) <- c("Male", "Female")
colnames(admit) <- c("Admitted", "Rejected")
admit <- as.table(admit)
```

Your analysis should as a minimum contain:

- Statement of hypotheses
- Calculation of expected frequencies
- Calculation of test statistic
- Calculation and interpretation of p-value.

A more detailed data set with the admissions for each department is available on the course web page. The variables are:

- `Admit` (admitted/rejected)
- `Gender` (male/female)
- `Dept` (department A, B, C, D, E, F)
- `Freq` (freqency of each combination)

Load the data into RStudio:

```
admission <- read.delim("http://asta.math.aau.dk/dan/static/datasets?file=admission.txt")
```

Using this data set you have to:

- Make the saturated model in RStudio
- Use `drop1` to test whether the model can be simplified.
- Even if the model cannot be simplified; make the simpler model `Dept*Gender + Dept*Admit`.
- Calculate the expected table and residuals for this simpler model.
- Find out which department deviates most from this model.

- Which kind of gender discrimination appears to be present?

Now we remove department A by removing the rows 1,2,3,4 (assuming the data is in a `data.frame` called `admission` and that department A is in the first four rows):

```
noA <- admission[-(1:4),]
```

Use this reduced dataset to conduct a new analysis which as a minimum contains

- Succesive removal of model terms using `drop1` and `update` starting from the saturated model.
- A graphical representation and interpretation of the final model for this dataset.