Exam H

You can download the combined lecture notes for this module at: https://asta.math.aau.dk/course/asta/ 2018-2/cph-bd/lecture/5-exam?file=H/module-H.pdf

Remember to load the `mosaic` package first:

```
library(mosaic)
```

Data collection

- Describe your data collection strategy and considerations (including weaknesses) in your project (or a previous/future/others'/...). This probably includes the following:

    - Very briefly what your project is about.
    - The ideal population/target group/market segment (the ones you ideally would like to make inference about).
    - Describing and discussing the sampling scheme, both advantages and disadvantages (e.g. practical as time/money).
    - Discussed the measured quantities and their status (for controlling, are there any potential lurky or confounding variables).
    - Describe the generalisability of your sample. Is it a true, random sample from the population of interest?

If you do not have a project with data collection, you can use this case: https://plast.dk/wp-content/uploads/ 2017/07/Plast-og-kemi-survey-af-danskernes-holdning-2017.pdf.

Berkeley admission

Six largest departments

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 and explanation of expected frequencies

- Calculation and explanation of test statistic
- Calculation and interpretation of p-value
- If relevant, interpretation of the standardized residuals

For each department

In the following, results are given for when each department is analyzed individually. Explain what is happening (also in relation to the previous problem with the overall numbers for all six departments combined).

Department A

	Male	Female
Admitted	512	89
Rejected	313	19

p-value $= 5.2 \times 10^{-5}$.

Standardized residuals:

	Male	Female
Admitted	-4.153073	4.153073
Rejected	4.153073	-4.153073

Department B

	Male	Female
Admitted	353	17
Rejected	207	8

p-value $= 0.77$.

Department C

	Male	Female
Admitted	120	202
Rejected	205	391

p-value $= 0.43$.

Department D

	Male	Female
Admitted	138	131
Rejected	279	244

p-value = 0.64.

Department E

	Male	Female
Admitted	53	94
Rejected	138	299

p-value = 0.37.

Department F

	Male	Female
Admitted	22	24
Rejected	351	317

p-value = 0.64.