Notes e-book

**Content overall:**

* Generic chapters (all blocks, needs to be done now)
* Data analysis methods theory (all blocks, some needs to be done now, some can wait e.g. PLS)
* MST chapters (block 3, missing)
* F&MCR chapters (block 4, missing)
* TFIH chapters (block 1, needs to be done now)

General formatting: Whenever in narrative mentioning packages put them in bold: \*\*devtools\*\*

**To do list** is found after the Index...

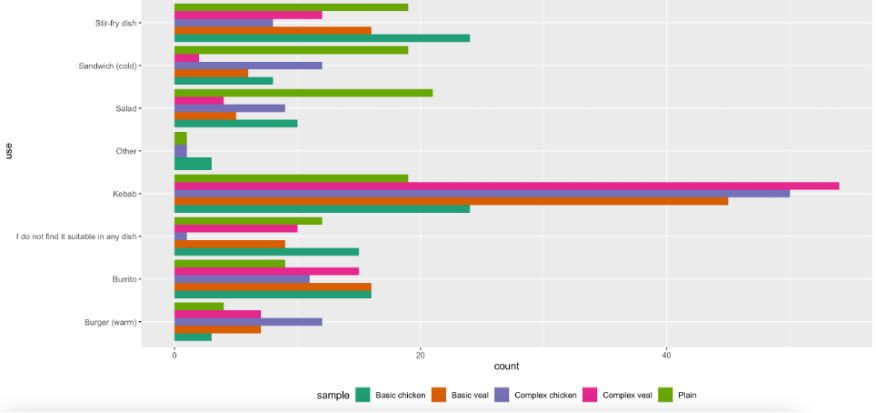
**Index** **(work in progress):**

1. Introduction to the book
2. (#, 01) Introduction to R – generic DONE
   1. (##) How to get started - understanding R (and RStudio) DONE
      1. (###, 01) Organise and save scripts DONE
   2. (##) How to import data DONE (there is no text, so easy 😊)
      1. (###) Import data from R-package DONE
      2. (###) Importing a csv file DONE
      3. (###) Importing an Excel file/sheet DONE
      4. (###) Clipboard import DONE
      5. (###) Looking at the imported elements DONE
      6. (###) Numbers and factors - changing categorisation DONE
   3. (##, 01) How to edit and merge datasets
      1. (###, 01) Import and merge in R (DONE)
         1. (####, 01) Edit using Tidyverse DONE
         2. (####, 01) Merging datasets DONE
            1. (#####, 01) Adding survey to buffet
            2. (#####, 01) Adding buffet to survey (DONE)
   4. (##, 01) How to save the data (missing: needs more elaboration / text)
   5. (##, 01) How to export data / results DONE
   6. How to load your RData DONE
   7. How to clear your environment DONE
   8. (##, 01) Ready for analysis DONE
3. (#, 02) Libraries DONE
4. (#, 03) Descriptive statistics
   1. ## Descriptives for a continuous variable DONE
      1. ### Mean/ median DONE
      2. ### Variance DONE
      3. ### Standard deviation DONE
      4. ### Calculations DONE
   2. ## Distributions of count data (MANGLER: skal vi overveje at lave dette med CATA data i stedet for? jeg er ikke helt vild med legumes datasættet nemlig.... men kan også bare blive stående)
   3. ## Aggregate DONE
   4. ## Tidyverse DONE
5. (#, 05) Inferential statistics DONE
   1. Intro DONE
   2. Hypothesis testing DONE
   3. Confidence intervals
   4. T-test DONE
   5. F-test DONE
   6. ANOVA - DONE
   7. Introduction to linear and mixed models DONE
   8. Normal and Mixed models DONE
6. (#, 06) Plotting data DONE
   1. ## Histograms and boxplots semi- DONE – check if text is sufficient
   2. ## Scatter plots (semi- DONE – check if text is sufficient
   3. ## How to export plots DONE
7. (#, 07) Introduction to PCA and multivariate data (ved ikke om der mangler noget???)
   1. ## Interpreting model output (MANGLER MORTEN: Eksempel på PCA output? Jeg foreslår PCA på sensorik datasæt lagt ind I Excel filen: BeefSensoryProfile. Jeg har beskrevet datasættet her)
8. (#, 09) Buffet and survey data
   1. Buffet data BODIL SKAL LÆSE DET HELE
      1. Introduction to buffet data BODIL SKAL LÆSE DET HELE
      2. Plotting buffet data BODIL SKAL LÆSE DET HELE
      3. Mixed model for buffet data BODIL SKAL LÆSE DET HELE
   2. Survey data BODIL SKAL LÆSE DET HELE
      1. Plotting survey data BODIL SKAL LÆSE DET HELE
      2. Linear model for buffet data BODIL SKAL LÆSE DET HELE
      3. Post-hoc test for survey data BODIL SKAL LÆSE DET HELE
   3. Combining consumption and survey data BODIL SKAL LÆSE DET HELE
9. MST Exercises – mangler, laves til 2024 med 2023 data!
10. FMCR COURSE HERE med intro først, at det er Large survey data. DA kan gøres på flere forsk måder. I dette kurser laver vi cluster analyse.
    1. Table one (på fake survey 2018-2022)
    2. Descript stats (fake survey) – Der mangler lidt forklaringer på nogle plots + andre koder, og så skal der tages stilling til, om noget af introduktionen skal fjernes/flyttes til den generelle intro i bogen
    3. Cluster analysis (EUqpc = pork, Julius rette arket i Dropbox til, således ikke relevante data slettes – data simplificeres) – Ikke så vigtigt, vi bruger bare variable vi kender, og snakker ikke om resten
       1. K means – er forklaret meget fint, men måske at nogle af plottene skal forklares bedre?
       2. Choosing segments by attributes – Hvis der her menes “Table 1”, så er der skrevet noget, men koden skal forklares bedre, og der skal kigges generelt på den røde tråd i hele det kapitel.
    4. Cluster profiling
       1. Logistic regression – Der er vildt meget god information + kode i både ”ProfilingSegments” og ”LogisticRegression”, men det der mangler, er dels noget mere forklaring på koderne, og dels en rød tråd igennem profilingen. Der står en smule mere inde i alle de individuelle .rmd-filer (for alle emnerne til kurset), men det der står her i word-dokumentet er det overordnede.
11. FMCR exerises on big data EU + reipurt data (denne hedder ”mimic data tables”, måske vi skal skrive forslag til kommandoer eller sørege for de kommer igennem de vigtigste steps med lidt ledende spørgsmål? Fx at finde hvilke variable der skal laves clustre efter. Sørge for de også laver en Table one. Stille spørgmål i øvelserne) – Dette er lige nu inkluderet i ”LogsiticRegression”, og er også en del af det med den røde tråd. For skal koden ”bare” stå der, så de kan køre den og bare lave det table, eller skal de selv lidt mere på banen?
12. # CATA data (Check-all-that-apply)
    1. ## Importing and looking at the beer data (MANGLER: checke om data er korrekt)
    2. ## Two versions of the data (MANGLER: punktopstilling. put in picture. Is my interpretation of the functions correct? Forklare de to do outputs. Indsætte kapitel link)
    3. ## Cochran’s Q test
       1. ### Post hoc contrasts (MANGLER: Morten: explain the code in words. Skal man bruge p.value eller p.adjust?)
       2. ### For all attributes in one run (nice to know) (MANGLER: Morten explain the code in words)
    4. ## PCA on CATA data (MANGLER: explain the code in words, BOM: interpretation of the PCA plot – send by email. Bodil tilføjer når det kommer)
13. # Hedonic rating scores (DONE)
    1. (##, 10) Plotting liking scores (MANGLER: fortolkning af output, Bom)
    2. (##, 10) Simple mixed models (MANGLER: fortolkning af output, Bom)
       1. (###, 10) Post hoc test (MANGLER: stor note er slettet, det skal udbygges med tiden?)
    3. (##, 10) Multivariable models (DONE)
       1. ### Additive models (DONE)
       2. ###Effect modification and Interactions (DONE)
14. # CATA and hedonics (DONE)
    1. ## Individual attributes and liking (DONE)
       1. ### An example with Refreshing (DONE)
       2. ### All attributes (MANGLER: explanation for models, is this a “nice to know”??)
       3. ### A beer centric model (MANGLER: explanation for models)
    2. ## PCA on CATA and Liking (MANGLER: explanation for models)
15. # Projective mapping (MANGLER: ALT, men vi ser ikke på det i år 2022)
    1. PCA on projective mapping data
    2. (MFA on projective mapping data)
16. # TFIH exercises (MANGLER: ALT, der er skrevet et skelet dog)
17. # Preference mapping (først til 2023: PLS on CATA and liking data)

**TO DO LIST:**

**EDITS to version 0.0 – missing in present version:**

1. **Chapter 9.1 Plotting liking scores:** add explanation for model
2. **Chapter 8.3.1 CATA, post hoc test**: can we find a solution to the letter based representation here too? (as with linear models)
3. **Add chilli and pasta datasets** to Excel sheet ”DatasetRbook” in Dropbox. – is this done?
4. Hvordan **ændrer man rækkefælge på sine samples i plots** ( se vedhæftet kode + nedenstående plot) dette giver samples i vilkårlig rækkefælge, kan man ændre denne, sådan at plain eks kommer først?



Morten svarer:

prøv at factor(,levels = c('lev1','lev2',...))

hvor i smider lev1 ogg lev2,.. til den relevante rækkefølge

og i størrelser, så se

theme(axis.text.y = element\_text())

også prøv at googgle theme() detaljer

1. **Chapter 10.1.2 All attributes:** add explanation to model, by now we do not know where all the different names and variables come from.
2. **Chapter 9.2.1** Last text bit: "The resulst indicate Brown Ale and Ravnsborg Red is the most likable beers". - results spelled wrong.
3. **Chapter 9.2.1**. Same last text bit: "These three at the bottom is also not significantly different." - is =are
4. **Chapter 9.3.1** This text: "In the beer dataset we would like to know which of the explanatory variables that is most related to the liking." - "is" should be "are" I guess?
5. **Chapter 9.3.2**. Last sentence: "You cannot remove a single variable if an interaction including it is significant." - including should be included? then I understand it
6. **Chapter 10.1.2**.: after the barplots "For each attribute a interaction model will be used to qualify further analysis. - a = an.

**After TFIH course editions, we need to do this:**

1. Code for exporting model output as an Excel table (make headline so it is easy to find)
2. Rette Rmd numrene (titler) så de passer med index
3. Tilføje dette til descriptive analysis: mean(chili$LikingAppearance, na.rm = TRUE) #remove NA values in dataset (lige nu ch 4.1.4)
4. Add relevel to mixed model in chapter 5.8, in doing the lm with summary.
5. Add (if relevant for data) the paste function or other funky way, to combine two columns. E.g. in MST adding gender + treatment as a unquie column – Chapter 8, or 5.7

**Bodil**

* Edit Word file to delete the stuff in the list, which has been fixed

**Morten**

* Sæt beef m PCA som intro to PCA, see text in INDEX.
* Unwrap liking long and liking wide in chp 8.
* Change headings in chapter 10 to analysis of CATA + hedonic rating – subchapters ## PCA and ## PLS.
* Check “plotting data” om text er sufficient?

**Helene**

* Intro – Chose what to do.
* Initiate basic plots (outliere and distributions)
  + Histogram
  + Boxplots
  + Point plots med liner til dommer
* Kontakte følgende ift. nye datasæt (se IFRO møde noter):
  + Thomas Bøker Lund, fødevareusikkerhed som hustandssurvey?
  + Sinne Smed, Indkøbsdata via GFK
  + Jørgen Dejgaard, kombinering af 4-5 forskellige databaser.
  + Lotte Holm & Thomas Bøker Lund: Mad og spisning i de nordiske lande. Forløbsdata. Tilbage fra 1990’erne.
  + COOP data, spørger til Aftagerpanelmøde
  + DTU kostundersøgelser
  + Varefakta

**Notes from meetings**

**Notes from meeting 3/10-2022 (Hennnnrik, Bodil, Helene, Morten)**

Make introduction chapter with

* Definition of variables (ordinal, conti,…)
* How to choose statistical model (pointing towards the chapters, what is a multivariate problem? What is a univariate problem? )

Make chapter on sensory = “Aroma summer course”

MST course (mostly just descriptives and plots and **adding linear models**)

Food+Meal CR course (plotting /descrp. K.means, logit, profiling)

In the Master in Technology, we will use the material as go-to for self-brush-up on stats.

Can this chapter on PCA be used as general introduction to multivariate / PCA for foreing students? [**Ask Åsmund to review**].

**Other ideas for data sets**

* Analysis of IFRO-kind-data (see meeting notes)
* House-hold data
* Single person
* New data: twitter, recepies, fødevarestyrelsen’s database…
* Marianne Thomsen?
* Dairy data?
* FQMC data? Nils?
* SOL Helene
* COOP dataset?
* App data? Maybe Inge Tetens?
* DTU kostundersøgelser – Helene kan nogle navne.

**Notes from IFRO meeting with Kia 24/10-2022:**

* Kia: Introducing Chi2 test in Excel in Social Science Methods, mentions R will come later in block 3 (us, MST course). Bør forstå bivariate sammenhænge efter hendes kursus.
* Kia kontakter Arne Hemmingsen ift hans R bog.
* Thomas data, fødevareusikkerhed som hustandssurvey? Helene kontakter.
* Indkøbsdata – Sinne Smed lektor <https://ifro.ku.dk/medarbejdere/?pure=da/persons/227352> ? Data ejes af firma. Tutorial datasæt kan måske frikøbes. Specialer. POCS. Ellers Sigrid Denver (???). Hedder GFK data. Helene kontakter.
* Jørgen Dejgaard data? Kombinering af 4-5 forskellige databaser. Helene kontakter.
* Lotte Holm & Thomas: Mad og spisning i de nordiske lande. Forløbs data. Tilbage fra 1990’erne. Helene kontakter.