DSUR Errata

Wilcox Functions

Wilcox's website address has changed since the book was published. Latest versions of the functions for robust analysis by Wilcox are available by executing:

```
source("http://dornsife.usc.edu/assets/sites/239/docs/Rallfun-v26.txt")
```

Code changes/Package Updates.

- Chapter 4 (ggplot2): After the book was published Hadley Wickham updated ggplot2, and some of the syntax changed considerably (see http://docs.ggplot2.org/current/). Please let me know of anything that doesn't work, but here are a few problems that I know about already.
- **Line graphs not working** This is a bug introduced in ggplot2 0.9.3 There will likely be a fix soon (a version 0.9.3.1). In the meantime, a temporary fix can be found by executing (I didn't write this fix and it could create other problems)[1]. See https://github.com/hadley/ggplot2/issues/732

  ```
  install.packages("devtools")

  library(devtools)

  source_gist("https://gist.github.com/4578531")
  ```

- Page 155 (R´s Souls´ Tip 4.3): scale_fill_manual ("Gender", c("Female" = "Blue", "Male"="Green")) should be scale_fill_manual ("Gender", **values= c**("Female" = "Blue", "Male"="Green")). [Thanks Steffen Wild].
- **The opts() function is depreciated and has been replaced by the theme() function.** This has implications for anything in the chapter that uses opts(). There is a very good transition guide to help you transfer from opts() to theme() here. Needless to say I will have to update the chapter/code at some point. If you correct any code then please email it to me if you feel so inclined☺ To get rid of the legend use theme(legend.position = "none") instead of opts().
- **P. 156:** the factor() function has changed, so you'll get an error using:

  ```
  hiccups$Intervention_Factor<-factor(hiccups$Intervention, levels = hiccups$Intervention)
  ```

  Instead, you need to execute this (to order the levels as they are in the book rather than alphabetic):

  ```
  hiccups$Intervention_Factor<-factor(hiccups$Intervention, levels(hiccups$Intervention)[c(1, 4, 2, 3)])
  ```

- **P. 199** (R's Souls' Tip 5.4): the final command:

  ```
  dlf$meanHygiene<-ifelse(dlf$daysMissing < 2, NA, rowMeans(cbind(dlf$day1, dlf$day2, dlf$day3), na.rm = TRUE))
  ```

  should be (note the position of NA – it has moved to the end of the command):

  ```
  dlf$meanHygiene<-ifelse(dlf$daysMissing < 2, rowMeans(cbind(dlf$day1, dlf$day2, dlf$day3), na.rm = TRUE), NA)
  ```

[1] I have used this patch on three different machines (Macs) and had no issues at all. However, Isaac van Patten emailed to say that the patch had messed up his system. He said he:

" … had to delete R 2.15.2 altogether and reinstall it. What will work is to remove ggplot2 0.9.3 from the library and then go to the archives and load ggplot2 0.9.1 from the source code … using the older version it draws the graphs as needed."

Like I said, it's not my patch, so use it at your own risk. It works fine for me, but it can cause problems. See

- P. 216 (the cor() function): If you throw the examData dataframe into cor() you'll get an error saying that x must be numeric. The problem is the gender variable, which is a non-numeric factor (Male and Female). The way around this, is to either select only the first 4 variables of the dataframe (the numeric variables):

  ```
  cor(examData[,1:4])
  ```

  You could also convert Gender to a 0, 1 dummy coded variable, then run cor() on the whole dataframe (in which case correlations involving gender will be the point biserial correlations). In the code below as.numeric() converts the Gender variable to numbers, but R will use 1 and 2 by default, so the the minus 1 changes these values to 0 and 1 as per dummy coding:

  ```
  > examData$Gender<-as.numeric(examData$Gender)-1
  > cor(examData)
  ```

- P. 226 (Section 6.5.7): bootTau<- function(liarData,i) cor(liarData … won't run without a space before cor, there is a space in the book but because of the typesetting that isn't necessarily clear. It's safer to bracket the function {}, so you could write this function as (Thanks Jan Dittrich):

  ```
  bootTau<- function(liarData,i){cor(liarData … etc. )}
  ```

- P. 235 (section 6.6.2): a required dependency for the ***ggm*** package is no longer supported by CRAN – the ***graph*** package is no longer available. It is being maintained at Bioconductor.org but requires individual download and installation. It also requires some other dependencies from Biocoductor, ***BiocGenerics*** & ***RBGL***, to be downloaded and installed in your library folder. To do this execute:

  ```
  source("http://bioconductor.org/biocLite.R")

  biocLite(c("BiocGenerics", "RBGL"))

  install.packages("ggm")

  library(ggm)
  ```

  Once this was done, the material in Section 6.6.2 will work. Without it you cannot load the ***ggm*** package. One other thing that is not obvious is a data frame must just be the variables included in the partial correlation for the var() argument (e.g. – it'll choke if you forget to strip out the subject numbers!). [Thanks, Isaac T. Van Patten, Radford University and Jeff P.]

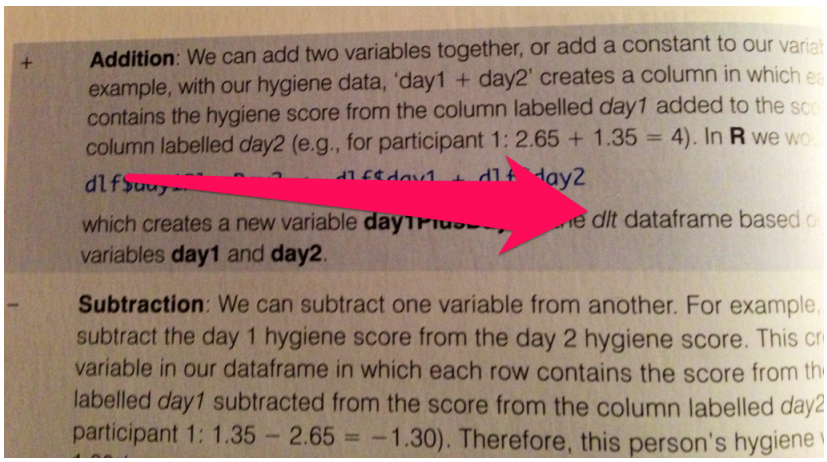
- P. 299: bootReg <- function (formula, data, indices)
  - Indices should be I to match the data[i,] two lines below. The code sample is correct, just the book that's wrong.
- P. 895: Growth models. If you use the file **Honeymoon Period Restructured.sav** everything will be fine. However, if you use the **Honeymoon Period.dat** file and restructure the data in R (using melt()) then you will get an error message resulting rom the fact that the variable **Time** is treated as a factor rather than a numeric variable. IN my code sample the data are prepared as follows:

```
satisfactionData = read.delim("Honeymoon Period.dat",  header = TRUE)

restructuredData<-melt(satisfactionData, id = c("Person", "Gender"), measured =
c("Satisfaction_Base", "Satisfaction_6_Months", "Satisfaction_12_Months", "Satisfaction_18_Months"))

names(restructuredData)<-c("Person", "Gender", "Time", "Life_Satisfaction")

restructuredData$Time<-as.numeric(restructuredData$Time)-1
```

However, in the book, I don't talk about this in detail (because of space) and I really should have flagged the need for the final line because it converts time into a numeric variable. In fact, I also subtract 1 from the numeric values because the as.numeric() function will convert the Time factor into values of 1, 2, 3, 4 and I want them to be 0, 1, 2, 3 (because the baseline value of time is a meaningful zero point).

Typos

- Page 14 line 11. 'j14the' >>>'the'
- - page 58, subsection *Statistical power*: "...as long as we know three of these properties" - shouldn't this mean "...two of these properties.."?
- Page 194: dlt should read dlf (thanks Bastian Wimmer):



- Page 212 (third variable problem): Reference to Jane Superbrain Box 1.1 should be 1.4.
- - page 218: within the two last cor.test() functions there is a bracket too much after "less"
- - page 291: The parentheses within the formula calculating the average leverage is wrong, it should be (k+1)/n rather than k+ 1/n.
- Page 299, line 3 and 5 from the bottom. advert >>> adverts time. >>> time).
- Page 224, line 3 and 6 from the top, miss-typing. liarData = >>> liarData <-
- Page 329: Variable name **Cured** should say **Intervention**.
- Page 379, line 12 from the top. statistics). ------> statistics.
- Page 382, line 4 from paragraph 2. -40 and 47 ------> 40 and 47
- Page 388: Equation's equal sign is omitted.
- page 415, line 6: There should not be a double dot
- page 428, heading: "hoc" should be also in italic letters
- page 455, below: *calculates.es* does not exist, the name is *compute.es* ;-)
- page 472, "...robust version of ANOVA,..." should be rather "...robust version of ANCOVA,..."
- page 474, R's Souls' Tip: "...to the effects in ther overall ANOVA..." should be also "...ANCOVA..."
- page 475, Jane Superbrain: First sentence: As far as I see type IV sums of squares has not been introduced
- page 476, last paragraph of Jane Superbrain: "... main choice in ANOVA designs is between Type II and Type III sums..." This contradicts Gramming Sam's tips on page 491 where in the third point it is written "you need to decide whether to use Type I or Type III sums of squares"
- page 482, R-code: should be "plot(viagraModel)" instead of "plots(viagraModel)"
- page 488, last paragraph: What is the small "x"? Should this be the capital "X" appearing in the sentence before?
- page 493, fourth R-code "mes(5.988117,...)". You have taken the wrong values here, these are not the mean and the adjusted mean but the values of the 95% confidence interval shown in output 11.4
- page 537, last R-code: This works (at least at my PC) only in case we have additionally specified "est=mom". Otherwise, only NA's are shown.
- page 538, Output 12.8: the right hand side of the output is completely missing
- page 543: "calculate.es" should rather be "compute.es"
- page 556, Figure 13.2: on the second level, SS_B respectively SS_W is one time above, one time below
- page 562 "General procedure...", point 4: "Depending on what you find in the previous step.." it is ton the previous step but the step before that
- page 566, R-code (last line): Should this be in blue, or is it rather a part of output 13.1?
- page 579: I am not sure if the "hat Psi" symbol has been introduced yet
- page 595, first R-code: This should be named "drinkModel", the "baseline" has been already defined before
- page 661, output 15.2: I think we need the package "car" to perform the Levene's Test. This package has not been mentioned at the begin of the chapter

- page 678, second last paragraph: "Output 15.8 shows that the Kruskal-Wallis test... " should be rather "... Shapiro-Wilk test..."
- page 689, last paragraph: "Friedman's ANOVA is significant..." should be replaced by "The Shapiro-Wilk Test is significant..."
- page 727, Figure 16.6: In the graph on the left hand side below: Should be the outlier (26) in blue?
- page 768, fourth last line: The name of the file is "raq.dat" instead of "RAQ.dat"
- page 778, "R's Souls' Tip": You should change the names "pc2" to "pc1", since these models are the same compared to the pc1 models above on this page. pc2 in contrast is defined ob page 781 as the rerun of pc1 using only relevant factors.
- page 783, R-code after last sentence on this page: Should be in blue color and separated from the output.
- page 788, in the middle of the page: R-command "pc2" should be changed to "pc3"
- - page 818, assumption 2 for the chi-square test: There are rules regarding frequencies > 5 or < 5 in the two first sentence, this excludes the case = 5. So what happens if all frequencies equal to 5?
- P. 818, line 3 from bottom, begins with "catData". Subsequently, when I refer to this data-frame (e.g., on p. 821, line 7 from bottom), I call it "catsData". I meant to call it "catsData" throughout. [Ronald Wyllys]
- page 839, R-codes: Here you suddenly use "=" instead of "<-" to define objects. A comment would be nice if "=" works always analogously to "<-"
- page 845, Figure 18.5: The title within the figure is wrong, it should be "Cats: Expected values"

Thanks to everyone spotting mistakes, especially @jongminbag and Dr. Moritz Mercker