Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Hausaufgaben
Data input and output

Erzeugen Sie ein lauffähiges und kommentiertes R-Skript (Dateiendung `.R`), in dem Sie die folgenden Aufgaben bearbeiten.

a) (5 Pkt.) Lesen Sie die Datei `lowbwt.sav` mit der Funktion `read.spss()` aus dem Paket `foreign` in einen Dataframe.

Übertragen Sie das Kodierungsschema in `lowbwt.cbk` auf den erzeugten Dataframe:

- vergeben Sie Variablennamen wie im Kodierungsschema angegeben;
- machen Sie Faktoren aus den Variablen `low` (Stufen `<= 2500g`, `> 2500g`, Anordnung beachten!), `race` (`White`, `Black`, `Other`), `smoke` (`No`, `Yes`), `ht` (`No`, `Yes`), `ui` (`No`, `Yes`);
- überprüfen Sie jeweils mit `table()` oder `xtabs()` die Zu- und Anordnung der Faktorstufen vor und nach der Umwandlung zum Faktor.

Der zu erstellende Dataframe soll folgende Struktur haben:

```
'data.frame':        189 obs. of  11 variables:
 $ id   : num  4 10 11 13 15 16 17 18 19 20 ...
 $ low  : Factor w/ 2 levels "<= 2500g","> 2500g": 1 1 1 1 1 1 1 1 1 1 ...
 $ age  : num  28 29 34 25 25 27 23 24 24 21 ...
 $ lwt  : num  120 130 187 105 85 150 97 128 132 165 ...
 $ race : Factor w/ 3 levels "White","Black",..: 3 1 2 3 3 3 3 2 3 1 ...
 $ smoke: Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1 1 1 2 ...
 $ ptl  : num  1 0 0 1 0 0 0 1 0 0 ...
 $ ht   : Factor w/ 2 levels "No","Yes": 1 1 2 2 1 1 1 1 2 2 ...
 $ ui   : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 1 2 1 1 1 ...
 $ ftv  : num  0 2 0 0 0 0 1 1 0 1 ...
 $ bwt  : num  709 1021 1135 1330 1474 ...
```

b) (1 Pkt.) Schreiben Sie den Dataframe mit der Funktion `write.table()` in eine tabstopp-getrennte Textdatei namens `lowbwt.txt`. Die erste Zeile soll die Variablennamen enthalten, Zeichenketten sollen *nicht* in Anführungszeichen stehen und die Datei soll *keine* Zeilennummern enthalten. Überprüfen Sie die erzeugte Datei mit dem Texteditor.

c) (2 Pkt.) Lesen Sie die Textdatei wieder in einen neuen Dataframe in R ein. Vergleichen Sie den Output von `str()` mit dem oben angegebenen. Bringen Sie die Faktorstufen wieder in die gewünschte Reihenfolge.

d) (2 Pkt.) Schätzen die Wahrscheinlichkeit für niedriges Geburtsgewicht bei rauchenden und nicht rauchenden Müttern.