Orientierungshilfe zum 2. Hausaufgabenblatt

Auch zum zweiten Hausaufgabenblatt eine Orientierungshilfe.

Aufgabe 8)

Berechne zunächst den Mittelwert und die Standardabweichung:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i=\frac{276}{14}\approx 19.714286$$

$$s(x)=\sqrt{var(x)}=\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{1/2}=\left(\frac{2761}{49}\right)^{1/2}\approx 7.50646$$

Man sollte beachten, dass es immer besser ist mit Brüchen weiter zu arbeiten als mit auf eine bestimmte Nachkommastelle gerundeten Dezimalwerten, da diese genauer sind! Wenn man mit Dezimalwerten arbeitet sollte man auch genügend Nachkommastellen benutzen! So wird beispielsweise die Ergebnisse stark beeinträchtigt, wenn zu Anfang der Mittelwert 19.7 statt 19.714 genutzt wird. **FAUSTREGEL für's nächste Mal: mind. 3 Nachkommastellen!!!**
Mit Hilfe der Formel

$$x_i'=\frac{x_i-\bar{x}}{(var(x))^{1/2}}=\frac{x_i-\frac{276}{14}}{\left(\frac{2761}{49}\right)^{1/2}}$$

können nun die standardisierten Werte berechnet werden, die in der nachfolgenden Tabelle aufgeführt sind.

x_i	16	15	20	20	10	10	19	14	38	19	19	33	24	19
x_i'	-0.4948	-0.6280	0.0381	0.0381	-1.2941	-1.2941	-0.0952	-0.7612	2.436	-0.0952	-0.0952	1.7699	0.5709	-0.0952
$x_i'^3$	-0.1212	-0.2477	0.0001	0.0001	-2.1673	-2.1673	-0.0009	-0.4411	14.455	-0.0009	-0.0009	5.5443	0.1861	-0.0009
$x_i'^4$	0.0599	0.1556	0	0	2.8048	2.8048	0.0001	0.3358	35.2133	0.0001	0.0001	9.8129	0.1063	0.0001

Daraus ergibt sich für den Schiefeparameter:

$$p=\frac{1}{n}\sum_{i=1}^{n}(x_i')^3\approx 1.0741$$

und für den Wölbungsparameter:

$$w=\frac{1}{n}\sum_{i=1}^{n}(x_i')^4\approx 3.6638 \quad .$$

An dieser Stelle möchte ich nochmal darauf hinweisen, dass es einen fundamentalen Unterschied zwischen

$$\sum_{i=1}^{n} x_i'^3 \neq \left(\sum_{i=1}^{n} x_i'\right)^3$$

gibt!!! Das wurde reihenweise nicht beachtet. Man mache sich das an dem einfachen Beispiel

$$(a+b)^2\neq a^2+b^2$$

klar.

Aufgabe 10)

Bei dieser Aufgabe müssen nun r(Y,D), r(Y,D^2) und r(Y,D^3) berechnet werden.
Zunächst ist es wieder sinnvoll die Mittelwerte sowie die Varianzen bzw. Standardabweichung von Y, D, D^2 sowie D^3 zu bestimmen.

$$\bar{Y}=2.53 \qquad var(Y)=0.22522 \qquad s(Y)=0.4746$$
$$\bar{D}=5.5 \qquad var(D)=8.25 \qquad s(D)=2.8723$$
$$\bar{D}^2=38,5 \qquad var(D^2)=1051.05 \qquad s(D^2)=32.4199$$
$$\bar{D}^3=302.5 \qquad var(D^3)=106334.25 \qquad s(D^3)=326.089$$

So ergibt sich beispielsweise die Varianz von D^3 durch

$$var(D^3)=(\frac{1}{10}\sum_{i=1}^{10}(D_i^3)^2)-(\bar{D}^3)^2=(\frac{1}{10}\sum_{i=1}^{10}D_i^6)-(\bar{D}^3)^2 \quad .$$

Die anderen Ergebnisse werden analog berechnet. Für die Kovarianzen kann folgendes berechnet werden:

$$cov(D,Y)=0.372$$
$$cov(D^{2,}Y)=5.092$$
$$cov(D^{3,}Y)=62.172$$

Somit sind alle wichtigen Werte gefunden, um die empirischen Korrelationskoeffizienten zu bestimmen:

$$r(D,Y)=0.2729$$
$$r(D^{2,}Y)=0.33096$$
$$r(D^{3,}Y)=0.40175$$

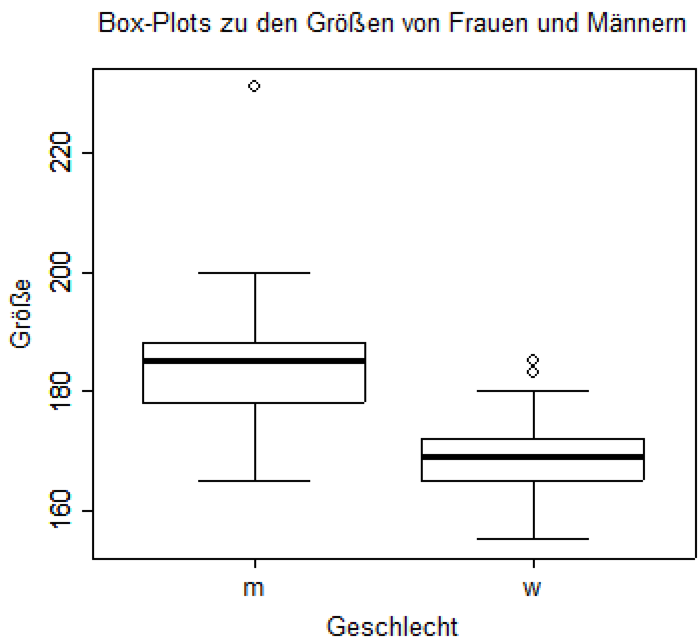
Alternativ hätte man auch die empirischen Korrelationskoeffizienten über die standardisierten Werte berechnen können!

Aufgabe 12)
Einige Anmerkungen zum Statistiklabor.
Wir hatten euch sowohl in der Vorlesung als auch in der Übung gezeigt, wie man einen Bericht erstellt, damit ihr eure Berechnungen ausdrucken und abgeben könnt. Hier nochmal schriftlich, in knapper Form.
Wenn ihr mit eurer Aufgabe fertig seid, stellt sicher, dass der R-Kalkulator auf 'run' steht, d.h. dass ihr im Berechnungsmodus seid (im R-Kalkulator das mittlere Knöpfchen drücken).
Um einen Bericht zu erstellen, müsst ihr auf 'Datei' → 'Bericht erstellen' gehen. Dann öffnet sich ein Fenster mit einer Liste eurer Objekte (Objektauswahl) und einer weißen Fläche.
Zieht den R-Kalkulator, alle eure Grafiken des Grafik Wizard und gegebenenfalls euren Texteditor auf die weiße Fläche. Der Datensatzimport und das Datensatzobjekt gehört **nicht** auf die weiße Fläche!!! Sonst druckt ihr nachher den ganzen Datensatz aus.
Bei den Symbolen des Grafik Wizards geht bitte mit einem Rechtsklick auf 'Einstellungen' und macht die Häckchen bei 'R-Code' und 'Einstellungen' weg, sodass nur ein Häckchen bei 'Bitmap' ist.
Geht dann auf 'Report/Bericht erzeugen'. Wenn ihr alles richtig gemacht habt, dann sollte sich ein

Texteditor (Standard müsste Word oder OpenOffice sein) und ihr bekommt die nachfolgende Ausgabe: (die Tabellenform wurde von mir erzeugt, um Platz zu sparen, ist aber nicht nötig.)

R-Kalkulator: 1	
R-Code:	**Berechnung:**
print(mean(GEBJAHR)) print(median(GEBJAHR))	[1] 88.55882 [1] 90
print(2012-(1900+ min(GEBJAHR))) print(2012-(1900+ max(GEBJAHR)))	[1] 38 [1] 14
w <- (GRO[GESCHL==2]) m <- (GRO[GESCHL==1]) mneu <- m[m<231]	
print(mean(w)) print(mean(m))	[1] 168.873 [1] 183.6712
print(var(w)) print(var(m)) print(var(mneu))	[1] 34.49974 [1] 81.33486 [1] 50.49276
print(max(w)-min(w)) print(max(m)-min(m)) print(max(mneu)-min(mneu))	[1] 30 [1] 66 [1] 35

R-Grafik Wizard:



Texteditor:

R-Code:

Aufgabe H12
6. Es ist ersichtlich ,dass Männer eher größer sind als Frauen. Anhand der Boxplots ist erkennbar, dass auch die Größen der Männer stärker streuen als die der Frauen (die Box der Männer ist fast doppelt so groß wie die der Frauen). Auffällig ist zudem ein Ausreißer bei den Männer der verdächtig groß ist (siehe dazu max(m)).

8. Anhand des Boxplots ist ersichtlich, dass die Größen der Männer stärker streuen als die der Frauen, dies spiegelt sich in der Varianz wider. Zusätzlich hat der Ausreißer, der eine Größe von 231m hat, einen starken Einfluss sowohl auf die Varianz als auch auf die Spannweite.

9. Man definiere sich eine neue Variable 'mneu <- m[m<231]'. Diese enthält nicht mehr den Ausreißer bei den Männern. Anhand der Ergebnisse mit der neuen Variable relativieren sich auch die Varianz und die Spannweite der Daten der Männen, sodass sie bezüglich der Frauen besser zu interpretieren sind.