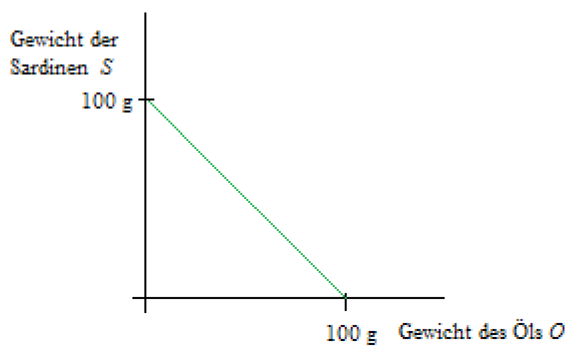
Orientierungshilfe Hausaufgabenblatt 3

Aufgabe 14:

a) Das Nettogewicht des Füllgutes N und das Verpackungsgewicht T sind unabhängig und damit unkorreliert. Offenbar kann zwischen N und B kein negativer Zusammenhang bestehen.
Wenn das Nettogewicht des Füllgutes steigt, muss B mindestens so groß sein wie N, da es kein negatives Verpackungsgewicht T geben kann, und in der Regel auch steigen; also positiv korreliert.
Man kann allerdings nicht davon ausgehen, dass, wenn T und N unkorreliert sind, auch B und N unkorreliert sein müssen.
An dieser Stelle sei darauf hingewiesen, dass T nicht als Konstante betrachtet werden kann. T hat eine eigene Verteilung(!) und auch Einfluss auf B.
Man kann dies auch mathematisch über $Cov(N, B) = Cov(N, N+T)$ zeigen; darauf wird jetzt aber verzichtet und war auch nicht relevant zur Beantwortung der Aufgabe.

b) Das Gewicht der Sardinen S und das Gewicht des Öls O sind negativ korreliert. Grund dafür ist die Beschränkung durch die Dose; es gilt: je mehr Sardinen man in der Dose hat, desto wenig Öl kann nachgefüllt werden.
Die Korrelation zwischen S und O beträgt -1.



Aufgabe 16:

Die Berechnung der empirischen Korrelationskoeffizienten erfolgt analog zu Aufgabe H10; siehe dazu die Orientierungshilfe zum 2. Hausaufgabenblatt. Es werden daher hier nur die wichtigsten Zwischenergebnisse aufgelistet.

a)

$\bar{F}=7203{,}7$	$s(F)=4351{,}322$	$cov(F,B)=-79965{,}45$	$r(F,B)=-0{,}1534980$
$\bar{B}=382{,}5$	$s(B)=119{,}7232$	$cov(F,S)=-10461{,}23$	$r(F,S)=-0{,}05014646$
$\bar{S}=132{,}9$	$s(S)=47{,}94257$	$cov(B,S)=-293{,}85$	$r(B,S)=-0{,}05119482$
$\bar{W}=174{,}1$	$s(W)=43{,}12876$		

b)

$\bar{S}_F=2{,}796$	$s(S_F)=1{,}844799$	$cov(S_F,B_F)=11{,}01506$	$r(S_F,B_F)=0{,}8680565$
$\bar{B}_F=8{,}59$	$s(B_F)=6{,}878437$		

c)

In diesem Teil der Aufgabe ist nach der Bestimmung des empirischen Korrelationskoeffizienten $r(S_W, B_W)$ gefragt. Dazu muss zunächst die Werte von B_W durch den Quotienten von B_i und W_i multipliziert mit dem Faktor 100 berechnet werden. Es ergeben sich folgende Werte:

$(B_W)_i$	235.6688	140.0000	170.0000	165.4054	210.7345	310.6145	426.2295	194.8052	201.3889	206.4748

Somit ergeben sich die Werte:

$$\bar{S}_W = 79{,}3 \qquad s(S_W) = 33{,}68694 \qquad cov(S_W, B_W) = 56{,}02105 \qquad r(S_W, B_W) = 0.02082947$$
$$\bar{B}_W = 226{,}1322 \qquad s(B_W) = 79{,}83833 \qquad cov(S_W, B_F) = -106{,}127 \qquad r(S_W, B_F) = -0{,}4580095$$

Es wird hier auch der empirische Korrelationskoeffizienten von S_W und B_F angeben, da es zu Missverständnissen in der Aufgabenstellung kam. Dies wurde auch als Lösung gewertet.

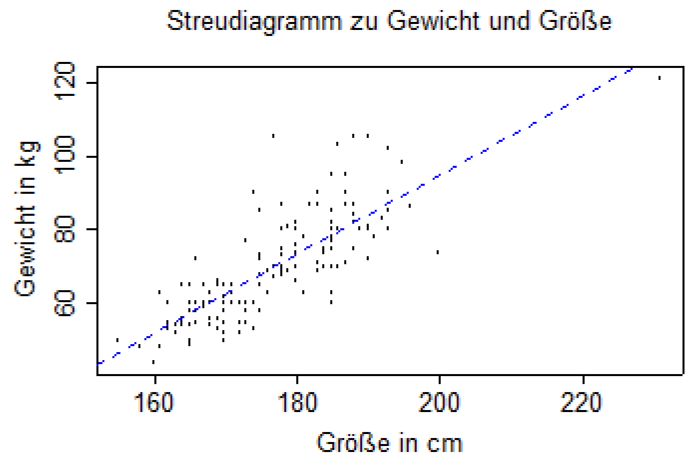
Interpretation: Nach Aufgabenteil a) ist ersichtlich das es keinerlei Korrelation zwischen der Anzahl der geborenen Babies, der Anzahl der Störche sowie der Fläche der Länder gibt. Dies ändert sich jedoch, wenn man die Anzahl der Babies und die Anzahl der Störche bezüglich der Fläche eines Landes betrachtet. Man erhält nun eine hohe positive Korrelation, d.h. je mehr Störche pro Fläche vorkommen, desto mehr Babies werden pro Fläche geboren. Bezieht man die Anzahl von Störchen wiederum auf die Wasserfläche so erhält man sogar einen leicht negativen Zusammenhang.
Offenbar widersprechen sich die Aussagen hier. Dies hängt damit zusammen, dass ein statistischer Zusammenhang nicht immer auf einen kausalen schließen lässt. In diesen Fall ist es weniger sinnvoll von der Anzahl an Störchen auf die Geburtenrate schließen zu wollen.

Aufgabe 17 und 18:

Im Folgenden wird eine Möglichkeit zur Lösung von Aufgabe 17 und 18 gegeben.

R-Code:	**Berechnung:**
covneu <- print(cov(GRO,GEW))	[1] 122.7345
vargro <- print(sd(GRO))	[1] 10.68069
vargew <- print(sd(GEW))	[1] 14.37881
print(covneu/(vargro*vargew))	[1] 0.7991802
print(cor(GRO,GEW))	[1] 0.7991802
print(mean(FACHSEM))	[1] 2.448529
print(mean(KLAUSUR))	[1] 4.595588
Laenge <-(length(KLAUSUR))	
print(sum(KLAUSUR>5))	[1] 14
print((sum(KLAUSUR>5)/Laenge)*100)	[1] 10.29412
print(cor(KLAUSUR,FACHSEM))	[1] 0.5297346

Zu Aufgabe 17
b) Offenbar korrelieren die beiden Variablen.
Dies ist auch sinnvoll; je größer man ist, desto schwerer sollte man auch sein.
c) Die Werte sind gleich.
e) Es ist im Streudiagramm ein linearer Zusammenhang erkennbar. Dieser wird durch den errechneten empirischen Korrelationskoeffizienten untermauert.



Zu Aufgabe 18

d) Die Ergebnisse decken sich. Betrachtet man die relative Häufigkeit für mehr als 5 bestandene Klausuren, so liegt diese bei ungefähr zwischen [0.9,1].

e) Es scheint eine positive Korrelation der beiden Werte geben. Dies ist auch sinnvoll, da man in der Regel mit zunehmenden Fachsemestern auch mehr Klausuren bestanden haben sollte.