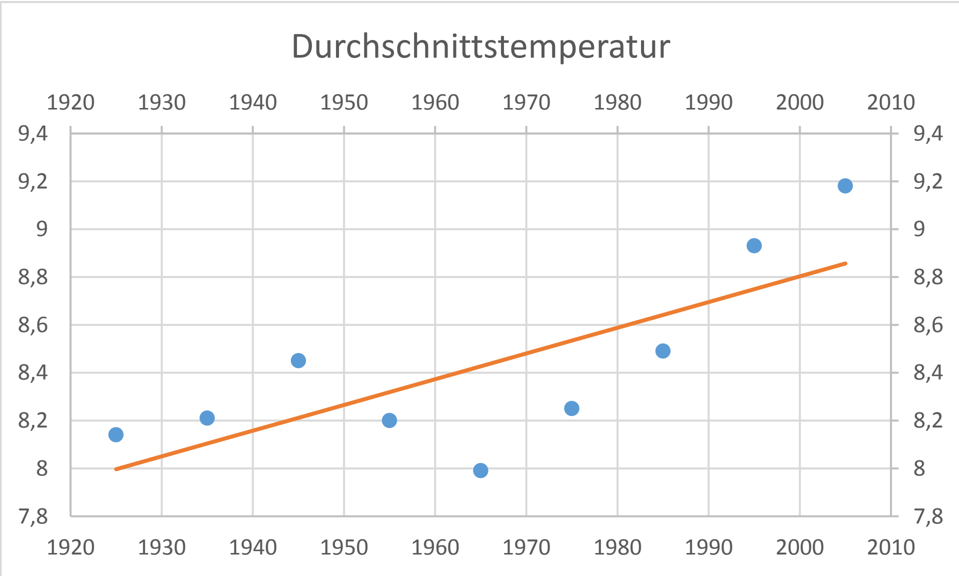
Ergebnisse 7.1

a)



b)

Mittelwert(Jahreszahl) = 1965

Mittelwert(Temperatur) = 8.43

$s_{Jahreszahl} \approx$ 25.82 (oder *korrigierte* Stichproben-Std-Abw: 27.39)

$s_{Temperatur} \approx$ 0.369 (oder *korrigierte* Stichproben-Std-Abw: 0.392)

$s_{Jahreszahl,Temperatur}$ = 7.17 (oder *korrigierte* Stichproben-Std-Abw: 8.06)

$r_{Jahreszahl,Temperatur} \approx$ **0.751** (Stichproben-Korrelation)

c)

Mit x = Jahreszahl ist die Gleichung der optimalen Regressionsgeraden

also $f(x) \approx 0.0108 \cdot x - 12.7$

d) Pro Jahrzehnt wird es im Mittel um **0.108 Grad** wärmer

e) i: **8.96 Grad**

ii: 9.29 Grad (unzuverlässig!)

iii: ii ist weniger zuverlässig, da weiter in die Zukunft extrapoliert wird.

Der Schätzwert wird dadurch stärker beeinflusst

- von der Annahme, dass der Temperaturanstieg linear erfolgt
- von den Schätzfehlern beim Parameter *k*

Ergebnis 7.2

	X = -1	X = +1	**Summe**
Y = -1	10%	0%	**10%**
Y = +1	80%	10%	**90%**
Summe	**90%**	**10%**	**100%**

$\bar{x} = -0.8$

$\bar{y} = 0.8$

$Cov(X,Y) = \sigma_{xy} = 0.04$

$\sigma_x \approx 0.6$

$\sigma_y \approx 0.6$

Korrelation:

$\boldsymbol{\varrho_{x,y} \approx \quad 0.11}$

Ergebnis 7.3 ohne Begründungen

a)

i. Nicht sicher richtig:
ii. Sicher richtig,
iii. Nicht sicher richtig:
iv. Sicher falsch,
v. Nicht sicher richtig:.
vi. Nicht sicher richtig:

b) .

i. Sicher falsch.

ii. Sicher falsch.
iii. Sicher richtig.
iv. Sicher richtig.
v. Sicher falsch.
vi. Sicher richtig,

Lösung 7.4

Hinweis:

- Skalenunterschiede beachten: Die Y-Werte bei iii und iv haben einen größeren Wertebereich, dadurch sind auch die Fehlerquadrate und Steigungen größer.
- Die Korrelation gibt nicht die Steigung der Regressionsgeraden an, sondern ist daran gekoppelt, um wie viel weniger die Y-Werte um die Regressionsgerade streuen als um ihren Mittelwert, d.h. in diesem Sinn „wie gut“ sich das Muster durch eine Gerade beschreiben lässt.
- Manche Vergleiche sind mit Augenmaß schwer; dann akzeptiere ich im Zweifel mehrere Antworten.

a) (ii) - (i) ≈ (iii) – (iv)
((i) und (iii) könnte man auch vertauschen – ist mit Augenmaß schwer zu entscheiden)
b) (iii) – (ii) – (i) – (iv) (Skala beachten)
c) (ii) – (i) - (iv) – (iii) (Skala beachten)
(die Fehler sind bei (iv) größer als bei (i), da y anders skaliert ist)

d) (iii) ist am kleinsten, (iv) = (i) , Rest ist schwierig zu entscheiden
(Je größer die Summe der Fehlerquadrate der Regressionsgeraden in Relation zur Varianz der y-Werte, desto näher ist die Kovarianz an Null)
((i) und (iv) haben identische Korrelation, da sie sich nur in der Skalierung unterschieden, und dies bei der Korrelation keine Rolle spielt)
(ii) ist schwer abzuschätzen, da Fehlerquadrate klein, aber Varianz der y-Werte ebenfalls.
e) (ii) - (iii) – (i) – (iv) (Skala beachten)

Lösung 7.5

a) 0.3 b) 0.5 (1.0 gäbe noch Teilpunkte)
c) $P(0 < X < 0.1) \approx 0.7 \cdot 0.1 = \underline{7\%}$ und $P(1 < X < 1.1) \approx 0.5 \cdot 0.1 = \underline{5\%}$

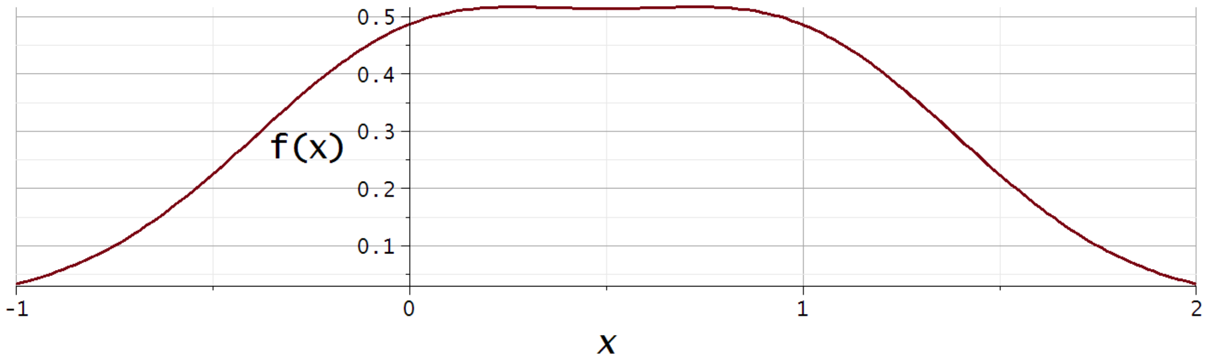
d) Nicht plausibel: D4 (da auch Y-Werte um 2 vorkommen, die Dichte dort aber fast Null ist)
Grenzwertig: D3 (Y-Werte sehr nahe an 2 oder 0 kommen vor)
Plausibel: D1 und D2

e) Nein, da bei kleinen X-Werten nur große Y-Werte vorkommen und bei großen X-Werten nur kleine Y-Werte

f) Ja, da die Verteilung der Y-Werte unabhängig vom X-Wert immer ähnlich ist.
g) 50% (Von den Punkten mit Y<2 liegt ca. die Hälfte rechts von der Y-Achse)

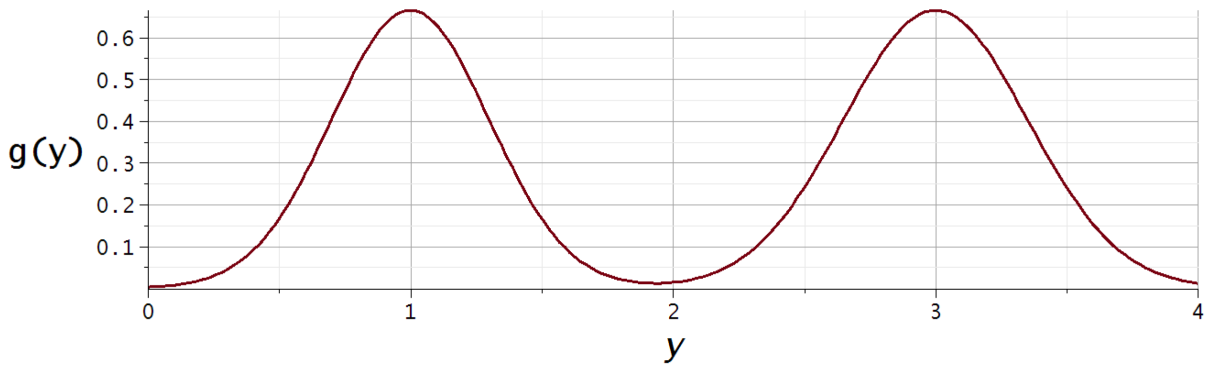
Lösung 7.6

a) Die Dichte von X, die tatsächlich beider Erzeugung des Datensatzes verwendet wurde, ist:



Hinweis: Der Verlauf muss nur grob stimmen, d.h.im X-Bereich zwischen 0 und 1 hoch sein und dann bis $x = -1$ bzw. $x = 2$ auf Null fallen. Sowohl ein flacher Verlauf zwischen 0 und 1 als auch ein Peak bei 0.5 als auch eine Delle bei 0.5 wären akzeptabel; das ist schwer per Augenmaß abzulesen.

b) Die Dichte von Y, die tatsächlich beider Erzeugung des Datensatzes verwendet wurde, ist:



Hinweis: Wesentlich sind zwei etwa gleich hohe Peaks bei y=1 und y=3, und dazwischen und zu den Rändern bis auf (fast) Null abfallen. Die Form des Graphen (geschwungen, dreiecksförmig) darf auch anders sein; das ist per Augenmaß nicht zu beurteilen.

c) Nein, denn z.B. ist die Wahrscheinlichkeit, dass ein X-Werte unter 0 liegt, für Y-Werte über 2 sehr viel geringer als für Y-Werte unterhalb von 2.

d) ca. 50%

e) ca. 25%

f) ca. 1/3

Lösung 7.7

y \ x:	-1	1	2	**Summe**
3	0	20%	30%	**50%**
1	25%	25%	0	**50%**
Summe	**25%**	**45%**	**30%**	

Man beachte die Analogie des folgenden Lösungsweges zu dem in 7.6

a) Nein, denn beispielsweise ist P(X=-1 | Y=3) = 0, aber P(X=-1) = 25%

b) P(X>0 | Y<2) = $\frac{P(X>0 \cap Y<2)}{P(Y<2)} = \frac{P(X=1 \cap Y=1)}{P(X=-1 \cap Y=1)+P(X=1 \cap Y=1)} = \frac{0.25}{0.25+0.25} =$ **50%**

c) P(X>0 ∩ Y<2) = $P(X = 1 \ \cap \ Y = 1)$ = **25%**

d) P(Y<2 | X>0) = $\frac{P(X>0 \cap Y<2)}{P(X>0)} = \frac{P(X=1 \cap Y=1)}{P(X=1 \cap Y=1)+P(X=1 \cap Y=3)+P(X=2 \cap Y=3)} = \frac{0.25}{0.25+0.20+0.30} =$ **1/3**

e) Einige der für diese Rechnung benötigten Formeln müssen Sie auswendig beherrschen; sie stehen nicht auf dem Formelblatt.
E(X) = -1·0.25 + 1·0.45 + 2·0.3 = 0.8
E(Y) = 1·0.5 + 3·0.5 = 2
Cov(X,Y) = (-1 - 0.75)(3 - 2)·0 + (1 - 0.75)(3 - 2)·0.20 + (2 - 0.75)(3 - 2)·0.30 +
(-1 - 0.75)(1 - 2)·0.25 + (1 - 0.75)(1 - 2)·0.25 + (2 - 0.75)(1 - 2)·0 = 0.8
V(X) = $(-1 - 0.75)^2 \cdot 0.25 + (1 - 0.75)^2 \cdot 0.45 + (2 - 0.75)^2 \cdot 0.3 = 1.26$
V(Y) = $(3 - 2)^2 \cdot 0.5 + (1 - 2)^2 \cdot 0.5 = 1$
Korrelation: $\varrho(X,Y) = \frac{COV(X,Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}} \approx$ **0.71**