Übungsblatt 3

1. Gegeben seien folgende Daten:

i:	1	2	3	4	5
y:	6	4	5	3	2
x:	1	2	3	4	5

Um Ihnen Arbeit zu sparen haben wir folgende Tabelle für Sie vorbereitet:

Sei $\ddot{y} := (y_i - \bar{y})$ etc.

	y	x	$\ddot{y}$	$\ddot{y}^2$	$\ddot{x}$	$\ddot{x}^2$	$\ddot{x}\ddot{y}$	$\widehat{y}$	$\hat{\ddot{y}}$	$\hat{\ddot{y}}^2$	$\ddot{y}\hat{\ddot{y}}$	$\hat{\varepsilon}$	$\hat{\varepsilon}^2$
	6	1	2	4	-2	4	-4	5.8	1.8	3.24	3.6	0.2	0.04
	4	2	0	0	-1	1	0	4.9	0.9	0.81	0	-0.9	0.81
	5	3	1	1	0	0	0	4	0	0	0	1.0	1.00
	3	4	-1	1	1	1	-1	3.1	-0.9	0.81	0.9	-0.1	0.01
	2	5	-2	4	2	4	-4	2.2	-1.8	3.24	3.6	-0.2	0.04
$\sum$	20	15	0	10	0	10	-9	20	0	8.1	8.1	0.0	1.9
MW	4	3						4					

(a) Berechnen Sie das Bestimmtheitsmaß R^2 der Regression $y_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_i + \hat{\varepsilon}_i$.

(b) Berechnen Sie den Korrelationskoeffizienten zwischen den gefitteten Werten $\widehat{y}$ und den beobachteten Werten y und vergleichen Sie das *Quadrat* dieses Korrelationskoeffizienten mit dem Bestimmtheitsmaß R^2 der Regression.

(c) Versuchen Sie allgemein zu zeigen, dass das Bestimmtheitsmaß R^2 das Quadrat des (Pearsonschen) Korrelationskoeffizienten zwischen den beobachteten Werten y und den gefitteten Werten $\widehat{y}$ ist, d.h. $R^2 = \text{corr}^2(y, \widehat{y})$.
Hinweis: Der Pearsonsche Korrelationskoeffizient ist definiert als

$$\text{corr}(y, \widehat{y}) = \frac{\text{cov}(y, \widehat{y})}{\sqrt{\text{var}(y)\,\text{var}(\widehat{y})}}$$

Berücksichtigen Sie, dass $y = \widehat{y} + \hat{\varepsilon}$ und und die Varianzrechenregeln $\text{cov}[x, (y + z)] = \text{cov}(x, y) + \text{cov}(x, z)$. Außerdem ist bekannt, dass

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{var}(\widehat{y})}{\text{var}(y)}$$

und dass in Regressionen mit Interzept $\text{cov}(\widehat{y}, \hat{\varepsilon}) = 0$ (warum eigentlich?).

(d) *Nur für TüftlerInnen:* Zeigen Sie, dass in einer bivariaten Regression das Bestimmtheitsmaß auch gleich dem Quadrat eines Korrelationskoeffizienten zwischen y und x ist (Achtung: dies gilt nur für bivariate Regressionen).

$$R^2 = \text{corr}^2(y, \widehat{y}) = \frac{\text{cov}^2(y, \widehat{y})}{\text{var}(y)\,\text{var}(\widehat{y})} = \frac{\text{cov}^2(y, x)}{\text{var}(y)\,\text{var}(x)} = r^2_{y,x}$$

2. Gegeben sei folgende Wahrscheinlichkeitsfunktion zweier diskreter Zufallsvariablen X und Y:

		Werte von Y		
		1	3	9
Werte	2	3/24	1/24	2/24
von	4	6/24	6/24	0
X	6	3/24	1/24	2/24

   (a) Handelt es sich dabei um eine Wahrscheinlichkeitsfunktion?

   (b) Wie groß ist die Wahrscheinlichkeit $\Pr(X = 2 \text{ und } Y = 9)$?

   (c) Wie groß ist die Wahrscheinlichkeit $\Pr(X = 2|Y = 9)$?
   Hinweis: $\Pr(X = 4|Y = 3) = 3/4$.

   (d) Berechnen Sie die Randverteilungen von X und Y (d.h. die *marginal probability density functions*).

   (e) Berechnen Sie die (unbedingten) Erwartungswerte von X und Y, d.h. $\mathrm{E}(X)$ und $\mathrm{E}(Y)$. (*Lösung:* $E(Y) = 3$)

   (f) Berechnen Sie die bedingte Wahrscheinlichkeitsfunktion von Y für $X = 2$ (d.h. die Wahrscheinlichkeit für jede Ausprägung von Y, wenn bekannt ist, dass $X = 2$).

   (g) Berechnen Sie die Varianz von X.
   Hinweis: $\mathrm{var}(X) := \mathrm{E}[X - \mathrm{E}(X)]^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$

   (h) Berechnen Sie die Kovarianz zwischen X und Y.
   Hinweis: $\mathrm{var}(X) := \mathrm{E}[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)] = \mathrm{E}(XY) - [\mathrm{E}(X)E(Y)]$

   (i) Sind X und Y stochastisch unabhängig?

   (j) Berechnen Sie die bedingten Erwartungswerte von Y für $X = 2$, $X = 4$ und $X = 6$, d.h. die bedingte Erwartungswertfunktion.
   Der bedingte Erwartungswert von Y, gegeben $X = x$, ist $\mathrm{E}(Y|X) = \sum_i y_i f(y_i|X = x)$.

   (k) Berechnen Sie die bedingte Varianz $\mathrm{var}(Y|X = 2)$ sowie die bedingte Varianzfunktion von Y (d.h. die bedingten Varianzen von Y für alle X).

3. Erfüllt die Funktion

$$f(x) = \begin{cases} 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst.} \end{cases}$$

   die Eigenschaften einer Dichtefunktion?

   Skizzieren Sie die Dichtefunktion graphisch und berechnen Sie die Wahrscheinlichkeit, dass x zwischen 0.5 und 0.8 liegt.

   Berechnen und skizzieren Sie die Verteilungsfunktion.

4. Zeigen Sie unter Verwendung der Definition für Varianz $\mathrm{var}(X) = \mathrm{E}[X - \mathrm{E}(X)]^2$ und Kovarianz $\mathrm{E}[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)]$, dass

   (a) $\mathrm{var}(c + 3X - Y) = 9\,\mathrm{var}(X) + \mathrm{var}(Y) - 6\,\mathrm{cov}(X, Y)$

   (b) $\mathrm{cov}[X, (Y + Z)] = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$