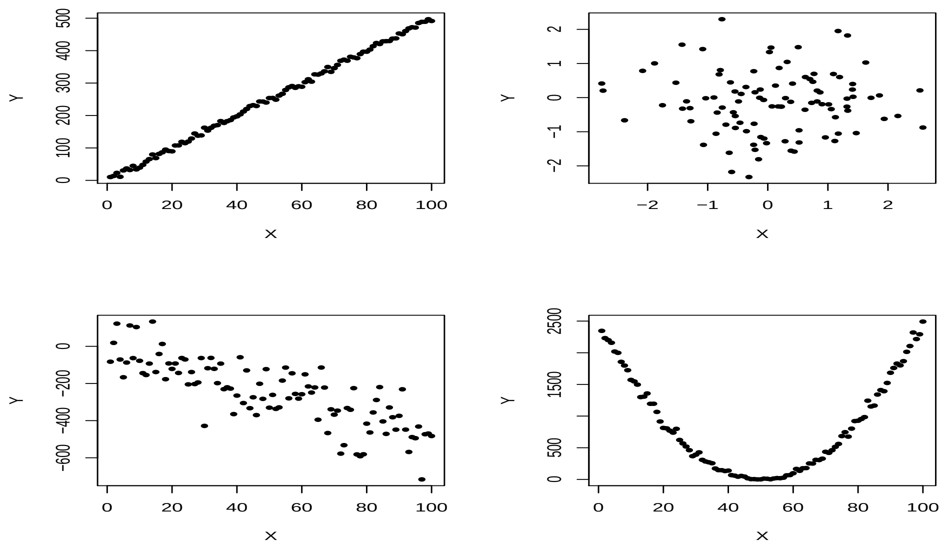
Grundbegriffe der medizinischen Biometrie und Epidemiologie (SoSe 2010)
Übungsblatt 8 am 17.6.2010

1 Aufgabe: Allgemeines zur Korrelation

(a) Plots und Korrelation



Die obigen Plots beschreiben Realisierungen von Zufallsvariablen X und Y. Geben Sie für jeden der Plots an, ob die Zufallsvariablen positiv, negativ korreliert oder unkorreliert sind. Welcher Plot deutet auf eine besonders starke Korrelation zwischen X und Y hin?

(b) Korrelation zwischen X und X^2
Gegeben sei eine beliebige Zufallsgröße mit $E(X) = 0$. Berechnen Sie die Korrelation zwischen X und X^2.

(c) Von den Zufallsvariablen X und Y ist bekannt, dass $Var(X) = a$, $Var(Y) = b$ und $Var(dX + eY) = c$, wobei $a, b, c, d, e > 0$. Wie groß ist dann der Korrelationskoeffizient zwischen X und Y?

2 Aufgabe: Kleinste Quadrate Schätzer

Der gängigste Schätzer für die Koeffizienten β_0 und β_1 einer linearen Regression ist gegeben durch den sogenannten "Kleinste Quadrate (KQ) Schätzer". Hat man Daten $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ gegeben, ergibt sich der KQ-Schätzer durch Minimierung von

$$\text{KQ} := \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

Berechnen Sie den KQ-Schätzer von β_0 und β_1.

3 Aufgabe: Regressionskoeffizient und Korrelation

1. Leiten Sie die Beziehung zwischen dem Regressionskoeffizient und Korrelation für die Variablen X und Y her.

2. Zeigen Sie, dass der Regressionskoeffizient von X auf Y unterschiedlich sein kann zu dem Regressionskoeffizient von Y auf X.

4 Aufgabe: Regressionsmodell-Output

In diesem Datensatz wird die Wegstrecke untersucht, die ein Spielzeugauto zurückgelegt hat, nachdem man es in unter- schiedlichen Winkeln eine Rampe herunterfahren ließ.

Der Datensatz enthält folgende Variablen

- *distance*: Gibt an, wie weit ein Auto von einer Rampe herab gefahren ist.
- *angle*: Bezeichnet den Winkel der Rampe.

Angle	Distance
1.3	0.37
4.0	0.92
2.7	0.64
2.2	0.70
3.6	0.89
4.9	1.30
0.9	0.38
1.1	0.43
3.1	0.69

Das Statistikprogramm `R` liefert folgende Ergebnisse:
`Coefficients:`

	Wert	Standardfehler	t-Wert	Pr(>|t|)	
β_0	0.14811	0.06503			
β_1	0.20954	0.02203			

(a) Testen Sie, ob die Regressionskoeffizienten signifikant von 0 verschieden sind. Interpretieren Sie diese Ergebnisse.

(b) Berechnen Sie auch das R^2 und interpretieren Sie es.

Zusatzaufgaben zum Vorrechnen

5 Aufgabe: Multivariate Verteilung

Gegeben ist die Funktion

$$f(x,y) = \begin{cases} cx^2 + y^2 & \text{für } 0 \leq x \leq 1 \text{ und } 0 \leq y \leq 1 \text{ ;} \\ 0 & \text{sonst.} \end{cases}$$

1. Bestimmen Sie die Konstante c so, dass $f(x,y)$ eine Dichtefunktion ist.
2. Berechnen Sie die Randdichten und Randverteilungsfunktionen von den Zufallsvariablen X und Y.
3. Sind X und Y voneinander unabhängig?
4. Bestimmen Sie die Verteilungsfunktion $F(x,y)$.

6 Aufgabe: Maximum Likelihood Schätzung

Eine gleichverteilte Zufallsvariable $X \sim Unif(0,\theta)$ mit $0 < \theta$ besitzt die folgende Dichte:

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{falls } 0 < x < \theta \\ 0 & \text{sonst} \end{cases}$$

Nun werden insgesamt n unabhängige und identisch verteilte (i.i.d.) Stichprobenelemente $x_i, i \in \{1, ...n\}$, beobachtet.

1. Man bestimme die Likelihood-Funktion $f(x;\theta)$ des unbekannten Parameter θ für die Beobachtungen $x_1, ..., x_n$.
2. Man bestimme mittels der ersten Teilaufgabe den Maximum Likelihood Schätzer für den unbekannten Parameter θ (Tip: Auch graphische Argumentation bzw. Darstellung möglich).

Übungsleiter:
Bernd Klaus (Dipl. Wi-Math) Mail: *bernd.klaus@uni-leipzig.de*
Verena Zuber (M.Sc.) Mail: *vzuber@uni-leipzig.de*