

**Aufgabenblatt 4 (1.6.2010)**

1. Im Skript (S. 134f.) gibt es zwei Varianten von Daten, die zu den gleichen bedingten Verteilungen (Tabelle 7.3) führen. Finden Sie ein drittes Beispiel. Was bedeutet die Tatsache, dass unterschiedliche Daten die gleichen bedingten Verteilungen implizieren können, für die Modellbildung?
2. Es gibt folgende Daten für drei statistische Variablen:  $X$  (Bildungsniveau),  $Y$  (Höhe des Arbeitseinkommens) und  $Z$  (Indikator für Gruppe).

$X$	$Y$	$Z$
2	2000	1
3	3000	1
3	3200	1
2	2500	1
2	2800	1
4	4000	1
4	1000	2
5	2000	2
5	2200	2
4	1500	2
4	1700	2
6	2900	2

- a) Zeichnen Sie die Daten in ein Streuungsdiagramm ein.
- b) Berechnen sie (mit OLS) lineare Regressionsgraden: erstens für alle Daten, dann gesondert für die Gruppen ( $Z = 1$  und  $Z = 2$ ).
- c) Zeichnen Sie die Regressionsgraden in das Streuungsdiagramm ein.
- d) Interpretieren Sie die Regressionsgraden als Modelle für bedingte Mittelwerte:  $M(Y|X = x, Z = z)$ .
- e) Berechnen und interpretieren Sie exemplarisch:  $M(Y|X = 2)$ ,  $M(Y|X = 2, Z = 1)$ ,  $M(Y|X = 2, Z = 2)$  und  $M(Y|X = 3)$ ,  $M(Y|X = 3, Z = 1)$ ,  $M(Y|X = 3, Z = 2)$ .
- f) Erläutern Sie, inwiefern es sich um ein Beispiel für Simpsons Paradox handelt.