

Gemischte lineare Modelle (Beispiele)

Beispiel 1.7:

Wir messen den Blutdruck von m Individuen $i = 1, \dots, m$ jeweils an verschiedenen Tagen t_{i1}, \dots, t_{in_i} , d.h. wir haben n_i Beobachtungen für Individuum i und insgesamt $\sum_{i=1}^m n_i$ Messungen $(Y_{ij}, x_{1ij}, \dots, x_{pij})$, $j = 1, \dots, n_i$, $i = 1, \dots, m$.

Modell für die Messungen von Individuum $i \in \{1, \dots, m\}$:

$$Y_{ij} = \beta_i + E_{ij}, \quad E_{ij} \stackrel{u.i.v.}{\sim} N(0, \sigma^2), \quad j = 1, \dots, n_i,$$

(Modell 1.1 mit $p = 1$).

Beispiel 1.10:

Nunmehr messen wir in 15 Kliniken ($k = 1, \dots, 15$) den Blutdruck Y von je 5 Patienten ($i = 1, \dots, 5$) nach jeweils 4 Behandlungen ($j = 1, \dots, 4$) (Placebo und drei Medikamente).

Modell:

$$Y_{kij} = \mu + \beta_j + a_k + b_i + c_{jk},$$

wobei die $a_k \sim N(0, \sigma_a^2)$, $b_i \sim N(0, \sigma_b^2)$, $c_{jk} \sim N(0, \sigma_c^2)$ und $e_{kij} \sim N(0, \sigma_e^2)$ alle unabhängig seien.

Die β_j seien hingegen genau wie μ feste Effekte (und unser Hauptinteresse).

Die Kliniken und Patienten betrachten wir hingegen als zufällig ausgewählt; jede Klinik stellt eine Art Cluster dar.

Beispiel 1.11:

Wir wollen den Einfluss von Faktoren wie Budget, durchschnittlicher Ticketpreis, Tordifferenz, Größe des Einzugsgebiets (Population) und Entropie der Punktezahlen (je Saison) auf die durchschnittlichen Zuschaueranzahlen im Profifußball bestimmen. Hierfür liegen uns diese Werte für die spanische Primera División der neun Spielzeiten 1992/1993 bis 2000/2001 vor.

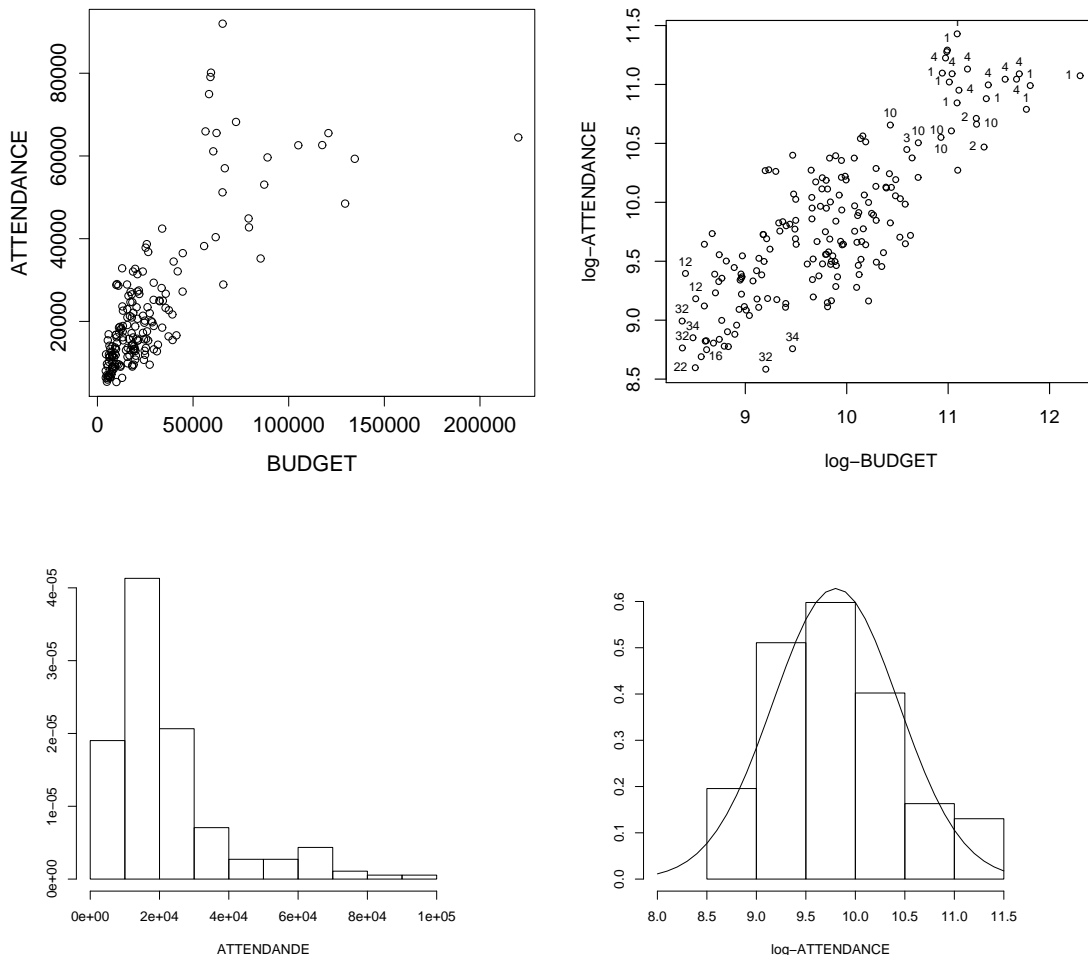


Abbildung 1: ATTENDANCE vs. BUDGET (oben links), log-ATTENDANCE vs. log-BUDGET (oben rechts), Histogramme von ATTENDANCE (unten links) und log-ATTENDANCE (unten rechts).

Basierend auf explorativen Analysen stellen wir als Modell für die logarithmierte durchschnittliche Zuschaueranzahl (Y) in Abhängigkeit von den Regressoren log-Ticketpreis (x_1), log-Budget (x_2), log-Population (x_3), Tordifferenz (x_4) und Entropie (x_5) ein lineares Regressionsmodell auf

$$\text{Modell A: } Y_{ij} = x_{1ij}\beta_1 + \dots + x_{5ij}\beta_5 + \beta_6 + e_{ij},$$

wobei $e_{ij} \stackrel{u.i.v.}{\sim} N(0, \sigma_e^2)$, $i \in \{1, \dots, 37\}$ der Index des Individuums (Teams) und $j \in \{1, \dots, t_i\}$ der Index der Saison ist.

Tabelle 1: Resultate für das Regressionsmodell A

coefficient	estimate	Std. error	t statistic	p-value
Intercept	-8.0026	5.7418	-1.394	0.1651
log-PRICE	0.1186	0.0756	1.568	0.1186
log-BUDGET	0.2458	0.0419	5.864	2.15E-08 ***
log-POPULATION	0.4571	0.0459	9.965	< 2E-16 ***
GOLAVERAGE	0.0039	0.0013	2.976	0.0033 **
ENTROPY	12.0927	5.8880	2.054	0.0415 *
loglike	$\hat{\sigma}_e^2$			
-20.91	0.0760			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Wir erweitern das Modell um Teameffekte b_i , $i = 1, \dots, 37$:

$$\text{Modell B: } Y_{ij} = x_{1ij}\beta_1 + \dots + x_{5ij}\beta_5 + \beta_6 + b_i + e_{ij},$$

wobei $b_i \stackrel{u.i.v.}{\sim} N(0, \sigma_b^2)$ unabhängig von $e_{ij} \stackrel{u.i.v.}{\sim} N(0, \sigma_e^2)$, $i \in \{1, \dots, 37\}$ der Index des Individuums (Teams) und $j \in \{1, \dots, t_i\}$ der Index der Saison ist.

Tabelle 2: Resultate für Modell B

Coefficient	Estimate	Std. Error	t statistic	p-value
Intercept	-10.0715	3.7012	-2.721	0.0073 **
log-PRICE	-0.1968	0.0716	-2.749	0.0067 **
log-BUDGET	0.1313	0.0319	4.112	0.0001 ***
log-POPULATION	0.2621	0.0585	4.479	< 2E-16 ***
GOLAVERAGE	0.0023	0.0010	2.399	0.0177 **
ENTROPY	17.5073	3.8118	4.593	< 2E-16 ***
loglike	$\hat{\sigma}_e^2$	$\hat{\sigma}_b^2$		
21.68	0.0267	0.1085		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

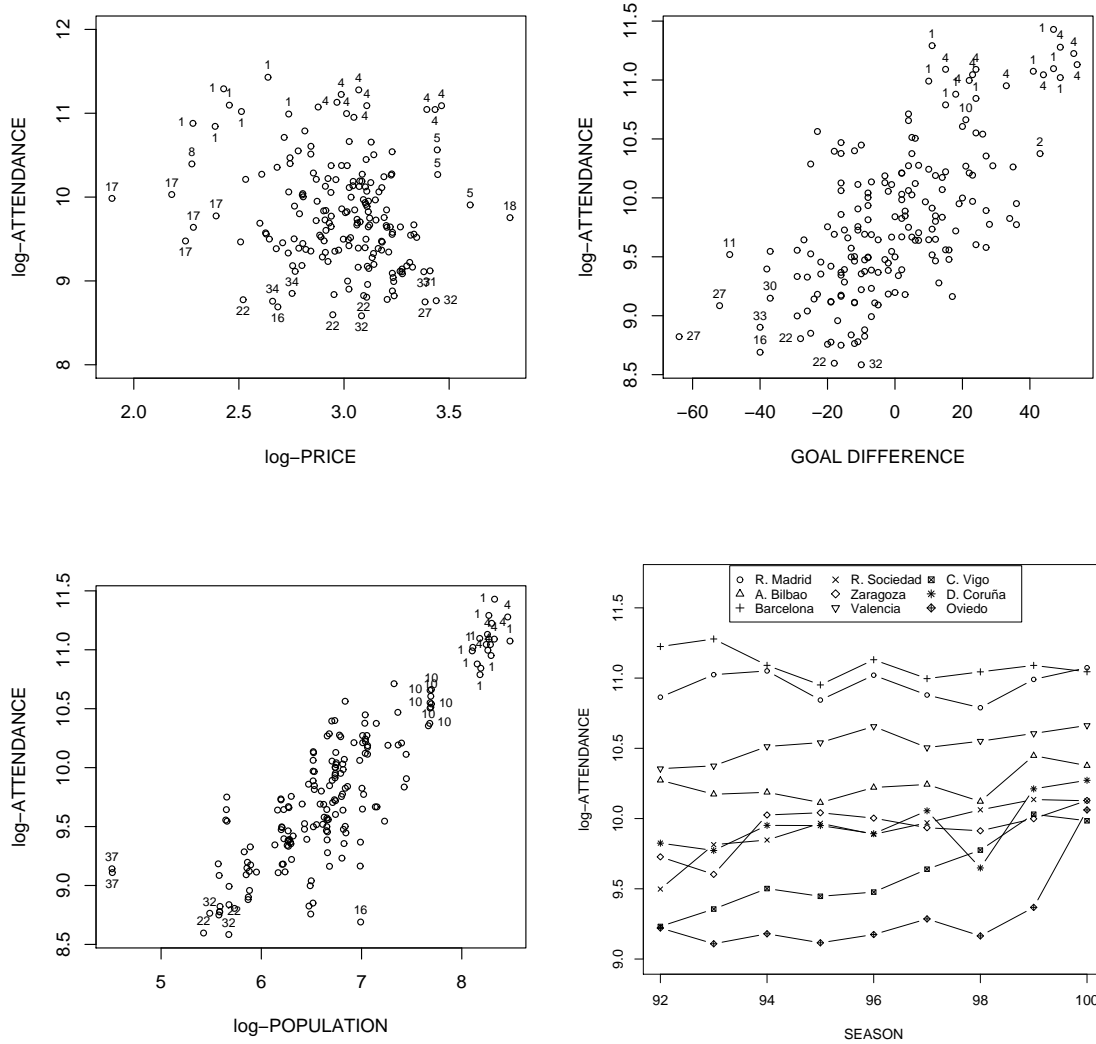


Abbildung 2: log-ATTENDANCE vs. log-PRICE (oben links) und log-ATTENDANCE vs. GOAL DIFFERENCE (oben rechts), log-ATTENDANCE vs. log-POPULATION (unten links) und log-ATTENDANCE vs. SEASON (unten rechts) für die neun Teams, welche die Primera División nie verlassen.

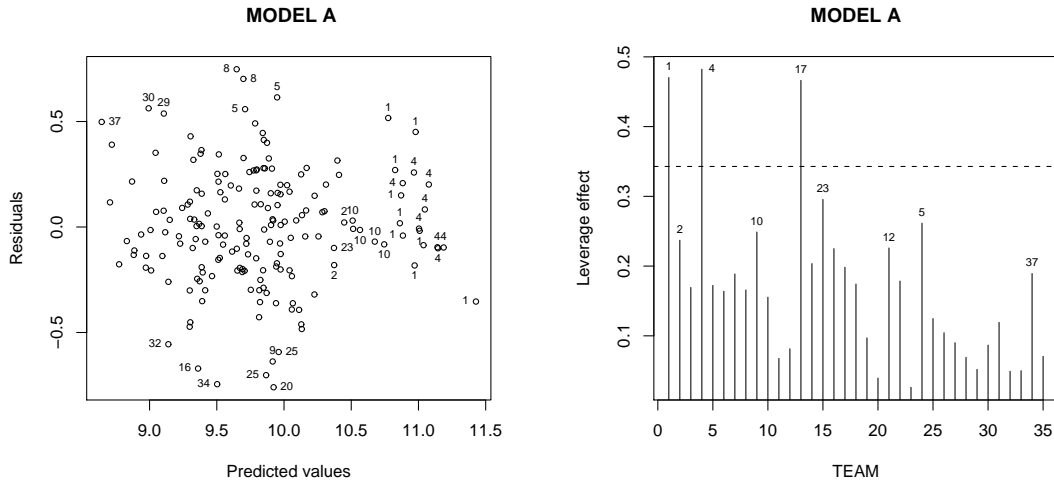


Abbildung 3: Residuen vs. Prognosen (links) und Leverage für jedes Team (rechts).

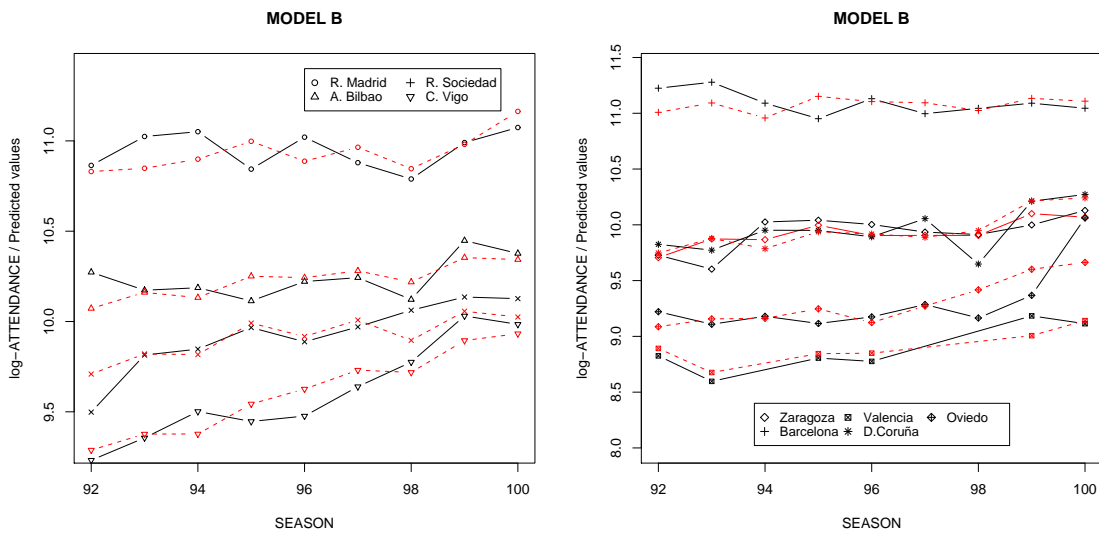


Abbildung 4: log-ATTENDANCE (durchgezogen) und predicted log-ATTENDANCE (gestrichelt) gegen Saison für die neun Teams, die in allen Saisons in der Primera División waren.