

## Serie 12

1. In der folgenden Tabelle stehen die Weltrekorde der Männer über 13 verschiedene Laufdistanzen, Stand 1974.

Distanz (m)	100	200	400	800	1000	1500	2000
Zeit (s)	9.9	19.8	43.8	103.7	136.0	213.1	296.2
Distanz (m)	3000	5000	10000	20000	25000	30000	
Zeit (s)	457.6	793.0	1650.8	3464.4	4495.6	5490.4	

An diese Daten wurde folgendes Regressionsmodell angepasst:

$$\text{Zeit}_i = \beta_0 + \beta_1 \cdot \text{Distanz}_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Der Regressionsoutput und die Diagnoseplots sehen folgendermassen aus:

Call:

```
lm(formula = zeit ~ dist)
```

Residuals:

Min	1Q	Median	3Q	Max
-106.95	-24.90	15.77	33.71	102.08

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-62.59296	21.81098	-2.87	0.0152 *
dist	0.18170	0.00173	105.05	<2e-16 ***

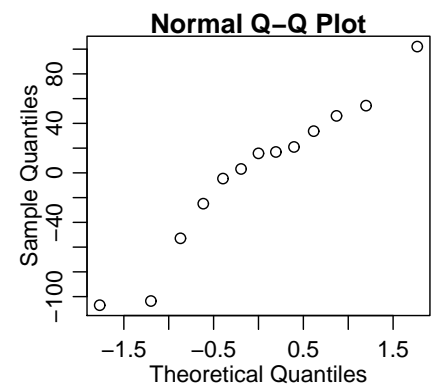
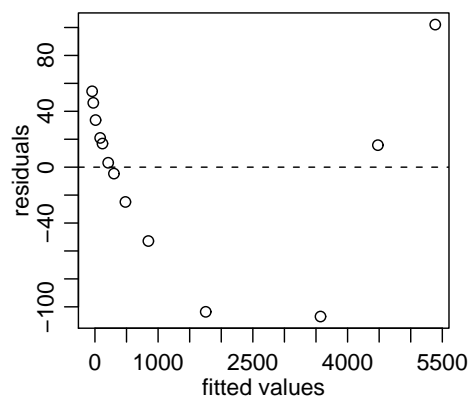
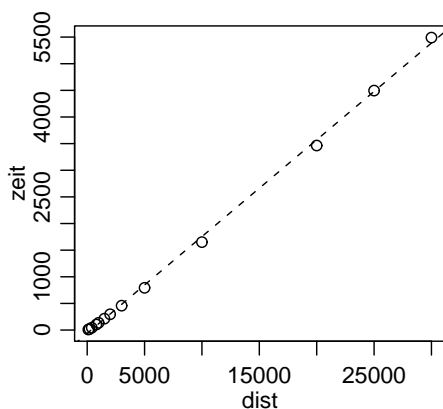
Residual standard error: 62.68 on 11 degrees of freedom

Multiple R-squared: 0.999, Adjusted R-squared: 0.9989

F-statistic: 1.103e+04 on 1 and 11 DF, p-value: < 2.2e-16

Residuals:

Min	1Q	Median	3Q	Max
-106.95	-24.90	15.77	33.71	102.08



- Gibt es einen signifikanten Zusammenhang zwischen Distanz und Zeit, d.h. ist  $\beta_1$  signifikant von 0 verschieden?
- Eines der folgenden 4 Intervalle ist das 95%-Vertrauensintervall für  $\beta_1$ . Welches?
  - [0.1800, 0.1834]
  - [0.1779, 0.1855]
  - [0.1765, 0.1869]
  - [0.1800, 0.1852]
- Wie gross ist das Residuum der 5. Beobachtung (1000m)?

- d) Dürfen wir die berechnete Regressionsgerade benutzen, um zu schliessen, dass 1974 der Weltrekord über 100km (100000m) ungefähr bei 18000s gelegen wäre?
- e) Wie gross ist die geschätzte Standardabweichung der Fehler  $E_i$ ? Was heisst das für die Brauchbarkeit des Modells?
- f) Was folgerst Du aus der Darstellung der Residuen gegen angepasste Werte?
- g) Formuliere ein Modell, das vermutlich besser zu diesen Daten passen würde.

**Besprechung:** Mittwoch, Mai 21.

**Abgabe:** Mittwoch, Mai 28.