# Vergleich verschiedener Vorzeichen-Tests

**Prof. Dr. Christine Müller, Dr. Kevin Leckey**

Nimmt man an, dass die Residuen in einem Modell eine Verteilung mit Median gleich 0 besitzen, ist der klassische Vorzeichen-Test ein sehr einfacher Test. Sei $y_n$ die $n$'te Beobachtung bei $x_n$, $z_n = (y_n, x_n^\top)^\top$ und res$(z_n, \theta)$ das Residuum dieser Beobachtung unter einem Modell mit Modellparameter $\theta$ für $n = 1, \ldots, N$. Setze $y = (y_1, \ldots, y_N)^\top \in I\!R^N$ und $x = (x_1, \ldots, x_N)^\top$. Die Teststatistik für $H_0 : \theta = \theta^0$ ist dann durch die relative Anteil der positiven (oder negativen) Residuen gegeben, d.h. durch

$$T_1(y, x) = \sqrt{N} \frac{|\frac{1}{N}\sum_{n=1}^N \mathbb{1}\{\text{res}(z_n, \theta^0) > 0\} - \frac{1}{2}|}{\frac{1}{2}} \tag{1}$$

gegeben, wobei $\mathbb{1}\{h(z) > 0\}$ die Indikator-Funktion $\mathbb{1}_A(z)$ mit $A = \{z;\ h(z) > 0\}$ bezeichnet. Die asymptotische Verteilung dieser Teststatistik liefert der Zentrale Grenzwertsatz. Man kann für kleine Stichprobenumfänge $N$ auch $\sum_{n=1}^N \mathbb{1}\{\text{res}(z_n, \theta)$ und dessen Binomialverteilung benutzen, um den kritischen Wert des Tests zu berechnen.

Dieser klassische Vorzeichen-Test hat allerdings für mehrdimensionale Modellparameter $\theta$ bei verschiedenen Alternativen eine extrem schlechte Güte. Das ist zum Beispiel für die lineare Regression mit Residuen der Form res$(z_n, \theta) = y_n - \theta_0 - \theta_1 x_n$, $\theta = (\theta_0, \theta_1)^\top$, in Abbildung 1 dargestellt. Beide Geraden $\theta = (1, 2)^\top$ (schwarze durchgezogenen Linie) und $\theta = (36, -2)^\top$ (rote gestrichelte Linie) ergeben die gleiche Teststatistik $T_1(y, x)$ mit minimal möglichen Wert von 0. Dabei liefert nur $\theta = (1, 2)^\top$ eine sinnvolle Anpassung.

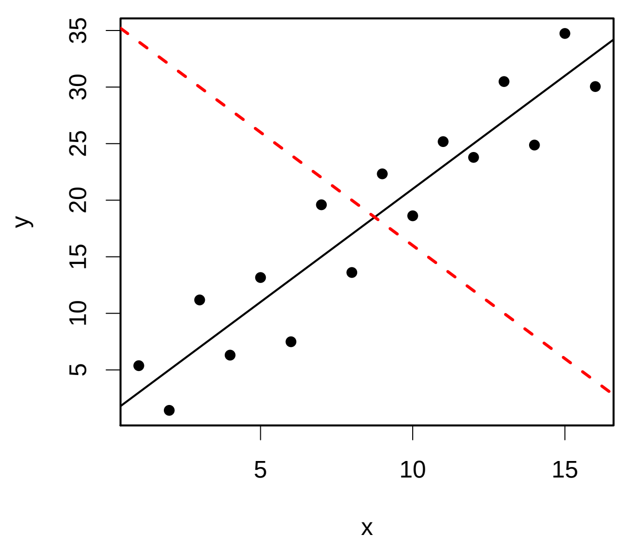


Figure 1: Vorzeichen-Test bei linearer Regression

Können die erklärenden Variablen (Regressoren) geordnet werden, d.h. gilt $x_1 < x_2 < \ldots < x_N$, kann stattdessen die Teststatistik eines verallgemeinerten Vorzeichen-Tests

$$\begin{aligned} T_2(y,x) &= \frac{N}{\binom{N}{3}} \Big( \sum\nolimits_{1 \le n_1 < n_2 < n_3 \le N} \big( \mathbb{1}\{\text{res}(z_{n_1},\theta^0) > 0, \text{res}(z_{n_2},\theta^0) < 0, \text{res}(z_{n_3},\theta^0) > 0\} \\ &\quad + \mathbb{1}\{\text{res}(z_{n_1},\theta^0) < 0, \text{res}(z_{n_2},\theta^0) > 0, \text{res}(z_{n_3},\theta^0) < 0\}\big) - \frac{1}{4} \Big), \end{aligned}$$

betrachtet werden. Hier ist es im Gegensatz zum einfachen Vorzeichen-Test (1) das Ziel einen möglichst hohen Wert der Teststatistik zu erreichen, d.h. möglichst viele alternierende Vorzeichen zu erhalten. Diese Teststatistik ergibt bei $\theta = (36, -2)^\top$ einen Wert gleich 0 und bei $\theta = (1, 2)^\top$ nimmt sie den maximal möglichen Wert an, d.h. $\theta = (36, -2)^\top$ würde als nicht passend verworfen werden, während $\theta = (1, 2)^\top$ als sehr passend erkannt wird. Bei dieser Teststatistik werden alle Mengen mit drei Datenpunkten daraufhin untersucht, ob die Residuen alternierende Vorzeichen haben. Die asymptotische Verteilung dieser Teststatistik wurde in Kustosz et al. (2016a) hergeleitet. Dabei wurde diese Teststatistik über Simplex-Datentiefe gewonnen (s. Kustosz et al., 2016b). Statt alle Mengen mit drei Datenpunkten zu betrachten, können auch nur Mengen mit drei aufeinander folgenden Datenpunkten betrachtet werden und wieder gezählt werden, wie oft drei aufeinander folgende Residuen alternierende Vorzeichen besitzen. Die Teststatistik wäre bei $\theta = (36, -2)^\top$ wieder 0 und bei $\theta = (1, 2)^\top$ maximal. Ist der Modellparameter $\theta$ $K$-dimensional, müssten $K + 1$ aufeinander folgenden Residuen daraufhin untersucht werden, ob alternierende Vorzeichen vorliegen. Das wurde in Kustosz et al. (2016b) vorgeschlagen und für zwei Beispiele mit $K = 2$ genauer untersucht. Es stellt sich aber die Frage, ob nicht $K$ oder $K - 1$ aufeinanderfolgende Residuen ausreichen bzw. ob die Güte des Tests besser wird, wenn mehr aufeinander folgende Residuen betrachtet werden. Weitere Fragestellungen sind die Anwendung zum Aufdecken von Änderungen (Strukturbrüche, Change-Points) in Zeitreihen und die Verallgemeinerung auf mehrdimensionale erklärende Variable (Regressoren) $x_1, \ldots, x_N$, wo keine natürliche Ordnung existiert. Das soll in den folgenden Abschlussarbeiten per Simulation untersucht werden, wobei das R-Paket dazu von Kustosz und Szugat (2016) genutzt werden kann:

**1. Vergleich von Vorzeichen-Tests für das Michaelis-Menten-Model:** Bachelorarbeit zu dem in der Enzym-Kinetik wichtigen Michaelis-Menten-Model, wobei die Untersuchungen in Swora (2015), Kustosz et al. (2016a) und Kustosz et al. (2016b) per Simulationen ausgebaut werden sollen.

**2. Vergleich von Vorzeichen-Tests für verschiedene AR-Modelle:** Masterarbeit, in der die Untersuchungen in Falkenau (2016), Kustosz et al. (2016a) und Kustosz et al. (2016b) mittels Simulationen zu einem nichtlinearen AR(1)-Modell ausgebaut sowie andere AR(1)-Modelle und AR(2)-Modelle untersucht werden sollen.

**3. Vergleich von Vorzeichen-Tests für die multiple Regression:** Masterarbeit, in der Verallgemeinerungen der Vorzeichen-Tests für die multiple Regression per Simulation und anhand von Datensätzen untersucht werden sollen. Insbesondere soll untersucht werden, wie sich die Tests verhalten, wenn zufällige Projektionen der erklärenden Variablen $x_1, \ldots, x_N$ zum Ordnen dieser Variablen genutzt werden.

**4. Anwendung der verallgemeinerten Vorzeichen-Test zum Aufdecken von Änderungen in Zeitreihen:** Masterarbeit, in der per Simulation und anhand realer Datensätze untersucht werden soll, wie gut die verallgemeinerten Vorzeichen-Tests zum Aufdecken von Änderungen (Strukturbrüche, Change-Points) im Datensatz geeignet sind. Als Datensätze stehen zwei große Datensätze aus den Ingenieurwissenschaften zur Verfügung.

**5. Vergleich der Güte von Vorzeichen-Tests bei Benutzung von Residuen-Paaren:** Masterarbeit, in der die Güte theoretisch und per Simulation für Paare von Residuen untersucht werden soll. Hier ist auch eine theoretische Untersuchung möglich, da in diesem Fall es sich um U-Statistiken mit symmetrischem Kern handelt.

## References


Falkenau, Ch.P. (2016). Depth based estimators and tests for autoregressive processes with application on crack growth and oil prices. *Dissertation*, TU Dortmund.

Kustosz, Ch.P., Leucht, A. und Müller, Ch.H. (2016). Tests based on simplicial depth for AR(1) models with explosion. *Journal of Time Series Analysis*, DOI 10.1111/jtsa.12186.

Kustosz, Ch.P., Müller, Ch.H. und Wendler, M. (2016). Simplified simplicial depth for regression and autoregressive growth processes. *Journal of Statistical Planning and Inference* 173, 125-146.

Kustosz, Ch.P. und Szugat, S. (2016). Simplicial Depth for Explosive Autoregressive Processes. `https://github.com/ChrisKust/rexpar/tree/sz` (der Titel des Paketes ist ein alter, zu spezieller Titel, der demnächst noch geändert wird, weil als Eingabe nur die Residuen eines beliebigen Models gebraucht werden).

Swora, M. (2015). Vereinfachte Simplex-Datentiefe in Regressionsmodellen. *Diplomarbeit*, TU Dortmund.