

Umgang mit Extremwerten und Ausreißern

Michael Höfler

Hintergrund

- In der Psychologie wird das Problem oft ignoriert, und es herrscht Unsicherheit im Umgang damit:



Special Collection: [Methodological resources for social psychologists](#)

Research article

How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration

Authors: [Christophe Leys](#) ✉, [Marie Delacre](#), [Youri L. Mora](#), [Daniël Lakens](#), [Christophe Ley](#)

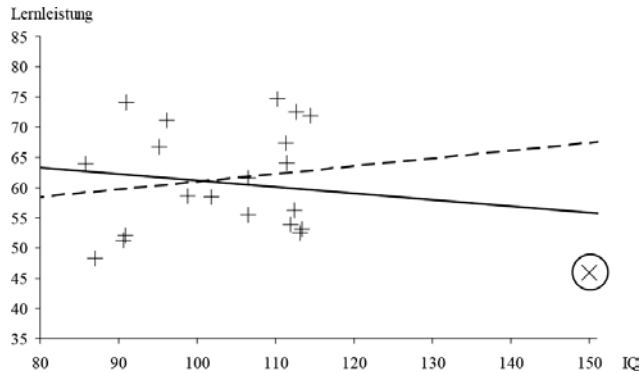
Ausreißer versus Extremwerte

- **Statistische Definition:** Ein **Ausreißer** stammt aus einer anderen Population als die anderen Werte
- Ein Ausreißer trägt somit **nicht** zu richtigen Schlüssen über eine **Zielpopulation** bei.
- Daher prinzipiell Ausreißer bei Auswertung **weglassen.**

- **Extremwerte** dagegen stammen aus derselben Population → **nicht pauschal weglassen.**
- Sind aber extrem groß bzw. klein

Auswirkungen von Extremwerten und Ausreißern

- Ergebnis kann stark vom **Ein- oder Ausschluss abhängen** („influential outliers“)
- Z.B. in linearer Regression durch Methode der kleinsten Quadrate



... führt dazu, dass **kein Datenpunkt** extrem weit von Regressionsgerade entfernt ist. Problem dabei: Durch Extremwerte sind womöglich **viele Punkte weit von Gerade entfernt**.

- → Ergebnisse mit **geringerer Anwendungsbreite** („little scope of analysis“): sagen nicht über viele Individuen etwas aus
(Beispiel für sinnlose Durchschnittsbildung: Ein Jäger schießt einmal zehn Meter links, einmal zehn Meter rechts am Hasen vorbei. Im Durchschnitt ist der Hase tot.)
- Durch Einschluss von Extremwerten wird oft **Varianz größer** (obwohl ausgewertete Stichprobe größer)

Immer primär durchführen:

Inhaltliche Ausreißeridentifikation

- Sind (mit demselben oder ähnlichem Messinstrument) bereits **ähnlich hohe Werte** (in ähnlicher Population) gefunden worden?
- Ist ein Wert aufgrund der Werte glaubwürdig, die die Person in anderen Variablen hat?

Wert nicht plausibel („error outlier“)

→ Fall mit dem Wert herauslassen (= auf „missing“ umkodieren)

→ oder durch **plausiblen Wert ersetzen**

falls Grund zur Annahme, dass Person tatsächlich sehr hohen, wenn auch nicht so hoch wie angegeben, Wert hat „**Winsorization**“: ersetze mit k-tem Perzentil; z.B. k = 1: Wert, wo 99% der Stichprobe drunter liegen; k wiederum **willkürlich**.

Wert plausibel

→ Wert prinzipiell belassen

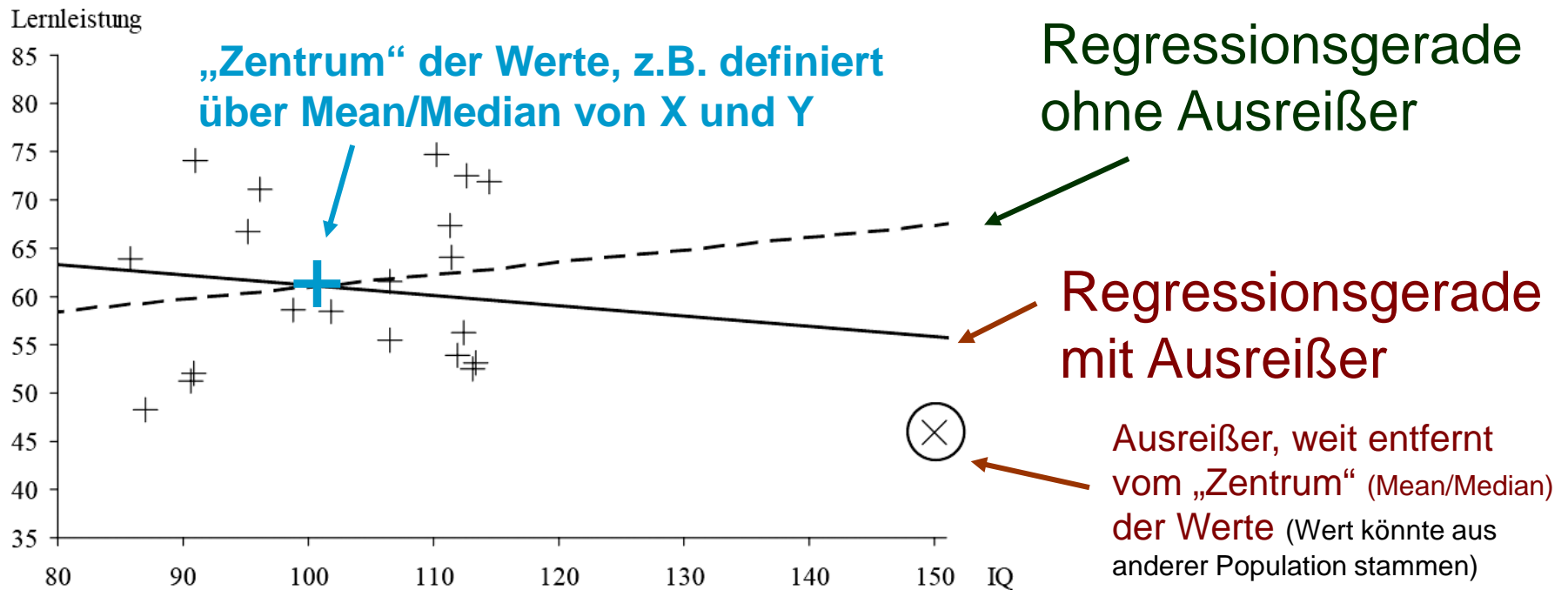
→ Aber bei der Auswertung beachten, s.u.

Univariate Ausreißer = Ausreißer hins. einer Variable

- Einer oder mehrere Werte einer Variable kommen einem ungewöhnlich hoch (niedrig) vor.
- Nur Werte dieser einen Variable bei der Frage herangezogen, ob es sich um einen Ausreißer handelt.

Multivariate Ausreißer

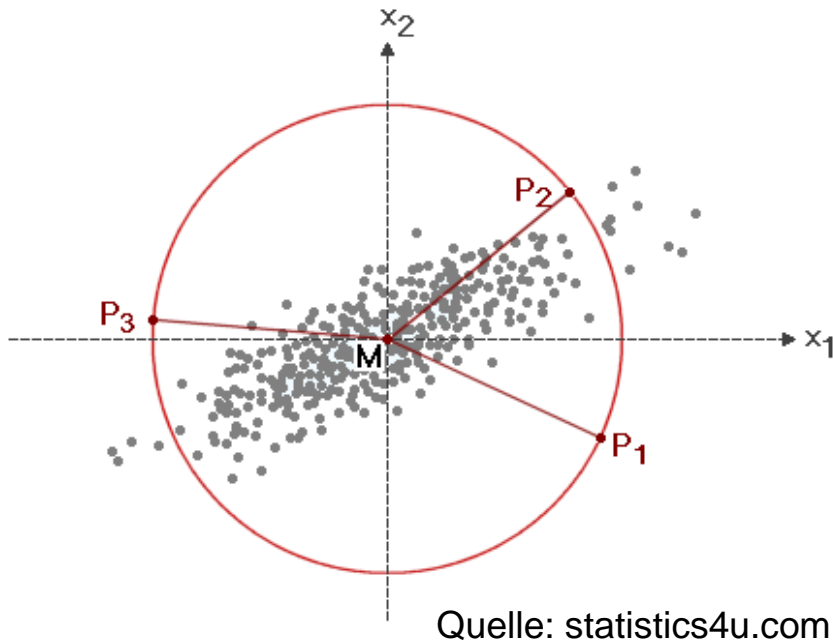
- Individuen, die bzgl. ihrer **Merkmalskombination** aus anderer Population stammen
- Fiktives Beispiel: Lernleistung und IQ



Quelle: <http://www.elearning-psychologie.de>

Mahalanobis-Abstand

Zwei Dimensionen (x_1, x_2), drei Datenpunkte (P_1, P_2, P_3)



Problem:

Varianzen und Kovarianzen sind bereits **sehr anfällig** gegenüber Extremwerten und Ausreißern

Euklidische Distanz: Summe der quadratischen Abstände vom Zentrum der Werte (Mittelwertsvektor M auf den beiden Dimensionen)

Mahalanobis-Abstand: Gewichtet die Abstände durch die Varianzen von x_1, x_2 und deren Kovarianz, S (größere Varianz, kleineres Gewicht; x_1, x_2 sind i.a. nicht gleich skaliert)

Berechne dann für jeden Probanden:

$$D_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

Ein Kriterium dazu: lasse solange Probanden weg, bis die Verteilung von D_i^2 nicht mehr sign. von Chi²-Verteilung abweicht (sehr abhängig von Stichprobengröße).

Gute Übersicht über auch andere multivariate Verfahren zur Untersuchung von Ausreißern:

[Front Psychol.](#) 2012; 3: 211.

Published online 2012 Jul 5. Prepublished online 2012 Jan 17.

doi: [10.3389/fpsyg.2012.00211](https://doi.org/10.3389/fpsyg.2012.00211)

PMCID: PMC3389806

PMID: [22783214](https://pubmed.ncbi.nlm.nih.gov/22783214/)

Distribution of Variables by Method of Outlier Detection

[W. Holmes Finch](#)^{1,*}

▶ [Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [Disclaimer](#)

Robustheitsmaß breakdown point einer statistischen Größe: Welchen Anteil der Werte kann man durch ∞ ersetzen, ohne dass die Größe sich ändert?

Beispiele:

- **Mean:** 0 (jeder Wert hat Einfluss auf mean)
- **Median:** 0.5 (generell maximal möglicher breakdown point)

Multivariat (Wert jeder Variable durch ∞ ersetzen)

- Steigung einer Regressionsgerade* β : 0
- **Pearson-Korrelation:** 0
- **Spearman-Korrelation:** max. 0.5
- Mahalanobis-Abstand: 0

Alle Standardverfahren (Normalverteilungsannahme) verändern ihre Ergebnisse, wenn man einen einzigen Wert ändert.

* Falls, wie üblich, mit „ordinal least squares“ geschätzt.

Statistische Ausreißeridentifikation

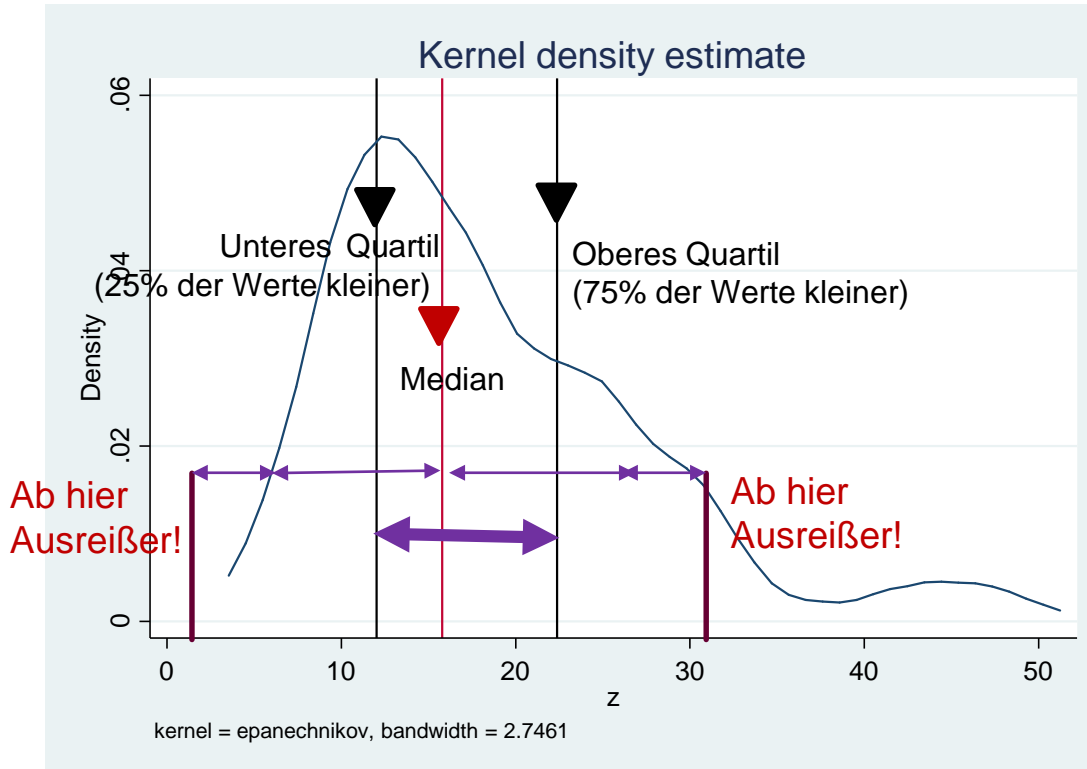
- Nützlich, um auffällige Werte zu entdecken
- sollte aber **nicht alleinige Grundlage** für Entscheidung sein, ob sie herauszulassen sind.

- benötigt immer ein **Referenzmodell**, das mittels Verteilungstyp die **Norm** festlegt.
- **Annahme**: Werte, die aufgrund dieser Verteilung äußerst unwahrscheinlich sind, stammen wahrscheinlich aus anderer Population
- Z.B. **Ausreißerkriterium bei Normalverteilung**: Werte, die größer (kleiner) als Mittelwert + (-) 3 * SD (Standardabweichung) sind. Wahrscheinlichkeit, dass ein Wert so groß ist, beträgt (unter Normalverteilungsannahme) 0.3%

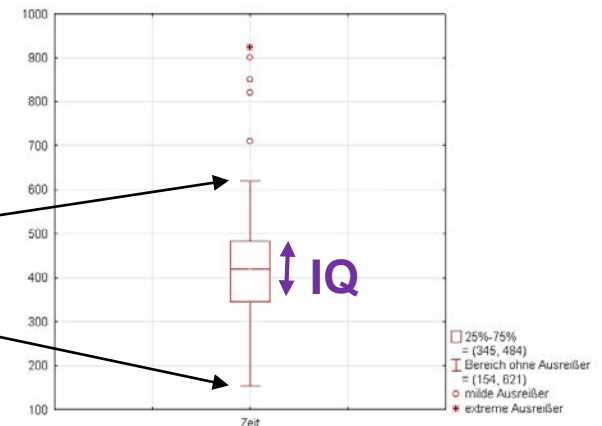
- ❖ Der Wert 3 ist **willkürlich!**
- ❖ Bei anderen, v.a. sehr schiefen, Verteilungen ist **SD als Streuungsmaß ungeeignet!**
- ❖ und ggf. selbst von Ausreißern stark beeinflusst!



Nonparametrische Alternative zur Standardabweichung: Interquartilsabstand (IQ)



Box-Plot, obere und untere Abschlusslinie („Whisker“)



Als **Ausreißer** gilt z.B. jeder Punkt außerhalb von
Median - 1.5 * IQ, Median + 1.5 * IQ
(1.5 ist wiederum **willkürlich**)

Replikationskrise/p-hacking

- **Variation in möglichen Ergebnissen** durch Umgang mit Extremwerten → für p-hacking missbrauchbar
- **Default-Vorgehen:** lasse Extremwerte drin!
- Rauslassen dann = Methode des p-hacking: kommt dadurch erwünschtes Ergebnis heraus?
- Dabei ist im Sinne der Anwendungsbreite das Rauslassen oft besser

Vorabregistrierung des Umgangs mit Extremwerten:

→ so genau, dass Variation in Ergebnissen durch Ein-/Ausschluss nicht missbraucht werden kann

→ Rohdaten und Syntax bereitstellen



Special Collection: Methodological resources for social psychologists

Research article

How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration

Authors: Christophe Leys, Marie Delacre, Youri L. Mora, Daniël Lakens, Christophe Ley

Aber welcher Umgang konkret?

- Beim **manuellen Herauslassen** kann man das willkürliche Kriterium selber bestimmen! (wenn man es nicht per Registrierung festgelegt hat)
- Lieber vorgegebenes Kriterium einer Software verwenden.

Vorgehen, das das Problem umgeht

- Neben Standardanalyse Verfahren verwenden, das **robust gegenüber Extremwerten** ist.
- Falls anderes Ergebnis (= Text der Interpretation ändert sich), ziehe Ergebnis des robusten Verfahrens vor, egal wie es aussieht!
- Begründung: robustes Verfahren macht schwächere Annahmen

Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers



Andy P. Field ^{a, *}, Rand R. Wilcox ^b

^a School of Psychology, University of Sussex, Falmer, Brighton, BN1 9QH, UK

^b Department of Psychology, University of Southern California, 618 Seeley Mudd Building, University Park Campus, Los Angeles, CA 90089-1061, USA

Alternativverfahren zur linearen Regression

Alternativverfahren	Robust gegenüber Ausreißern/verändert Punktschätzung	Nachteile
Verfahren mit Rangstatistiken (U-Test, Rangkorrelation u.a.)	Ja	Nicht möglich in komplexen Stichproben (z.B. gewichtete/geclusterte Daten)
Sandwich-Methode zur Schätzung der Standardfehler	Nein	Nur in großen Stichproben möglich
Bootstrapping	Nein	-
Robuste lineare Regression	Ja	Nur bei linearen Zusammenhängen möglich
Generalisierte lineare Modelle	Nur sehr bedingt	Nur in großen Stichproben möglich
Box-Cox-Transformation von Y ($g(Y) = (Y^L - 1)/L$ L aus Daten geschätzt; z.B. $\ln(Y)$, \sqrt{Y})	Nur bedingt	Größe von Zusammenhängen auf Y-Skala schwer zu beurteilen, da Transformation kompliziert

Beispiel robuste (lineare) Regression

Voraussetzung

- Zusammenhang zwischen **Y** und **X** ist **linear**
(evtl. vorher **Y** mittels „Box-Cox-Transformation“ Richtung Normalverteilung transformieren)
- **X** und ϵ sind unabhängig

Funktionsweise

- Ausreißer (extrem große **Residuen**) automatisch erkannt und **weggelassen** (z.B. Cook's distance > 1: Um wie viel ändern sich Regressionsergebnisse, wenn man ein Individuum weglässt?)
- Der Beitrag anderer Individuen mit besonders hohen Residuen wird **heruntergewichtet**
- Daten werden damit so „getrimmt“, dass sonst. Voraussetzungen der linearen Regression auch erfüllt (normalverteilte Residuen mit gleichen Varianzen)
- Ergebnisse hängen kaum von einzelnen Beobachtungen ab → robust und viel breiter interpretierbar
- Methode funktioniert oft auch in kleinen Stichproben
- Aber nicht in komplexen Stichproben (gewichtete, korrelierte Beobachtungen u.a.)

Beispiel

. regress y x **Lineare Regression („ordinary least squares“)**

Source	SS	df	MS	Number of obs	=	21
Model	10.329275	1	10.329275	F(1, 19)	=	2.57
Residual	76.2651898	19	4.01395736	Prob > F	=	0.1252
Total	86.5944648	20	4.32972324	R-squared	=	0.1193
				Adj R-squared	=	0.0729
				Root MSE	=	2.0035

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.2526831	.1575171	1.60	0.125	-.0770039 .5823702
_cons	-.4827175	.8045221	-0.60	0.556	-2.166602 1.201167

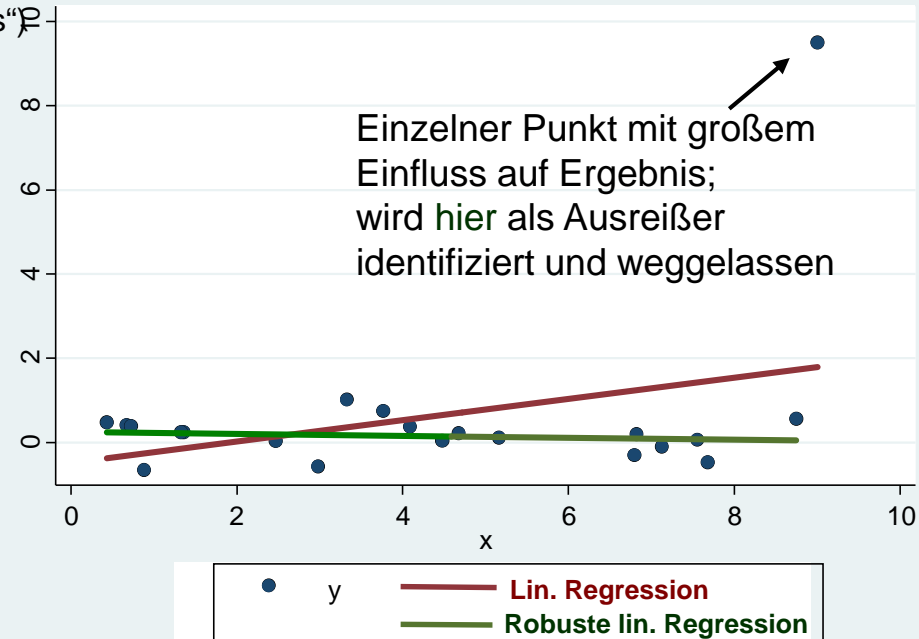
. rreg y x **Robuste lineare Regression**

Huber iteration 1: maximum difference in weights = .56662541
 Huber iteration 2: maximum difference in weights = .15080309
 Huber iteration 3: maximum difference in weights = .06817685
 Huber iteration 4: maximum difference in weights = .02367172
 Biweight iteration 5: maximum difference in weights = .29817371
 Biweight iteration 6: maximum difference in weights = .08324226
 Biweight iteration 7: maximum difference in weights = .04053151
 Biweight iteration 8: maximum difference in weights = .00902548

Robust regression

Number of obs	=	20
F(1, 18)	=	3.50
Prob > F	=	0.0776

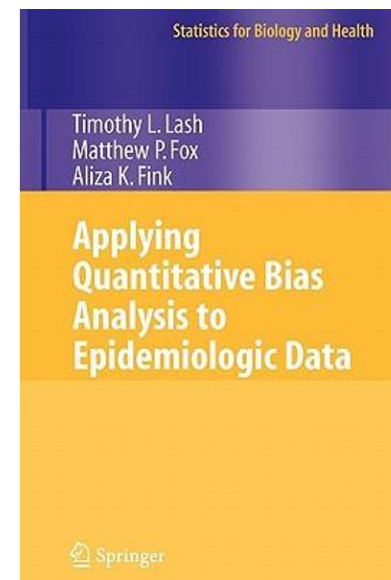
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	-.0574964	.0307186	-1.87	0.078	-.1220338 .007041
_cons	.4152154	.1484097	2.80	0.012	.1034181 .7270127



Wahl des Ausreißerkriteriums (Cook's $d > 1$) spielt hier keine so große Rolle (weil Beob. knapp daneben auf fast 0 heruntergewichtet werden (es gibt aber einen anderen „tuning parameter“, der das Ausmaß der Gewichtung steuert).

Verwandtes, allgemeines Vorgehen: **Sensitivitätsanalyse**

- Man kann **nicht entscheiden**, welche Analyse besser ist
- Führe daher auch **alternative Analyse** durch, um zu sehen, wie **sensitiv** Analyse gegenüber der Entscheidung ist
- **Ähnliche Ergebnisse**: Entscheidung spielt keine Rolle (der Einfachheit halber bei Standardvorgehen bleiben, „Sparsamkeitsprinzip“, „occams razor“)
- **Widersprüchliche Ergebnisse**: berichte beide, damit Leser sieht, was die Entscheidung zur Folge hat → **Transparenz**, Leser kann selber entscheiden und entspr. Schluss ziehen



Weiterführende Literatur

- DeSimone JA, Harms PD, DeSimone AJ: Best practice recommendations for data screening. *Journal of Organizational Behavior* 36; 171–181.
- Hastie R, Tibshirani J, Freedman J. *The elements of statistical learning. Data mining, inference, and prediction*. Springer, 2009
- Hardle W, Muller M, Sperlich S, Werwatz A. *Nonparametric and semiparametric models*. Springer, 2004
- Huber, P. J. 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35: 73–101.
- Leys C et al. How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration. *International Review of Social Psychology* 2019, 32: 5, 1–10.
- Li, G. 1985. *Robust regression*. In Exploring Data Tables, Trends, and Shapes, ed. D. C. Hoaglin, C. F. Mosteller, and J. W. Tukey, 281–340. New York: Wiley.