Ludwig-Maximilians-Universität München
Institut für Statistik

Seminararbeit

Überblick über Messfehler und ihre Auswirkungen in der linearen Regression

Seminar "Statistische Herausforderungen im Umgang mit fehlenden bzw. fehlerbehafteten Daten "

Leitung:
Prof. Dr. Augustin
Betreuerin:
Eva Endres

Autor:
Hanna Marshalava

15. März 2015

Inhaltsverzeichnis

1 Einleitung

Ungenauigkeiten sind bei keiner Art von Messung zu vermeiden. Dabei ist es wichtig, alle erdenklichen Fehlerarten und deren Auswirkung zu kennen, um für jedes Messergebnis die entsprechende Genauigkeit angeben zu können. Das ist insbesondere dann wichtig, wenn die gesuchten Größen nicht beobachtet werden können.

Nimmt man eine physikalische oder chemische Messung vor, so ist diese niemals exakt. Das stellt man fest, wenn die Messung wiederholt wird oder wenn verschiedene Personen das Gleiche messen. Es treten dann Schwankungen oder Ungenauigkeiten in den erhaltenen Messwerten auf. (Höpcke, 1980, S. 41)

Zur Verdeutlichung der Problematik betrachten wir die Messergebnisse der Astronomischen Einheit (AE). Unter einer Astronomischen Einheit versteht man die Maßeinheit der großen Halbachse der Erdbahnellipse, d.h. den mittleren Abstand zwischen Sonne und Erde.

Tabelle 1: Messergebnisse der Astronomischen Einheit (Hartung et al., 2009, S. 321)

Ort bzw. Durchführende und Jahr der Messung	AE in Millionen Meilen	Schätzung der Schwankung durch Experimentator
Necomb, 1895	93.28	93.20 - 93.35
Hinks, 1901	92.83	92.79 - 92.87
Noteboom, 1921	92.91	92.90 - 92.92
Spencer Jones, 1928	92.87	92.82 - 92.91
Spencer Jones, 1931	93.99	92.99 - 93.01
Witt, 1933	92.91	92.90 - 92.92
Adams, 1941	92.84	92.77 - 92.92
Brouwer, 1950	92.977	92.945 - 93.008
Rabe, 1950	92.9148	92.9107 - 92.9190
Millstone Holl, 1958	92.874	92.873 - 92.875
Jodrell Bank, 1959	92.876	92.871 - 92.882
S. T.L., 1960	92.9251	92.9166 - 92.9335
Jodrell Bank, 1961	92.960	92.958 - 92.962
Cal.Tech, 1961	92.956	92.955 - 92.957
Soviets, 1961	92.813	92.819 - 92.816

Man weiß, dass die Astronomische Einheit im Beobachtungszeitraum konstant ist, so dass die unterschiedlichen Ergebnisse nur durch Messfehler entstanden sein können.

Abbildung 1 der Ergebnisse zeigt die Schwankungen der Messwerte und die wahren Werte der AE

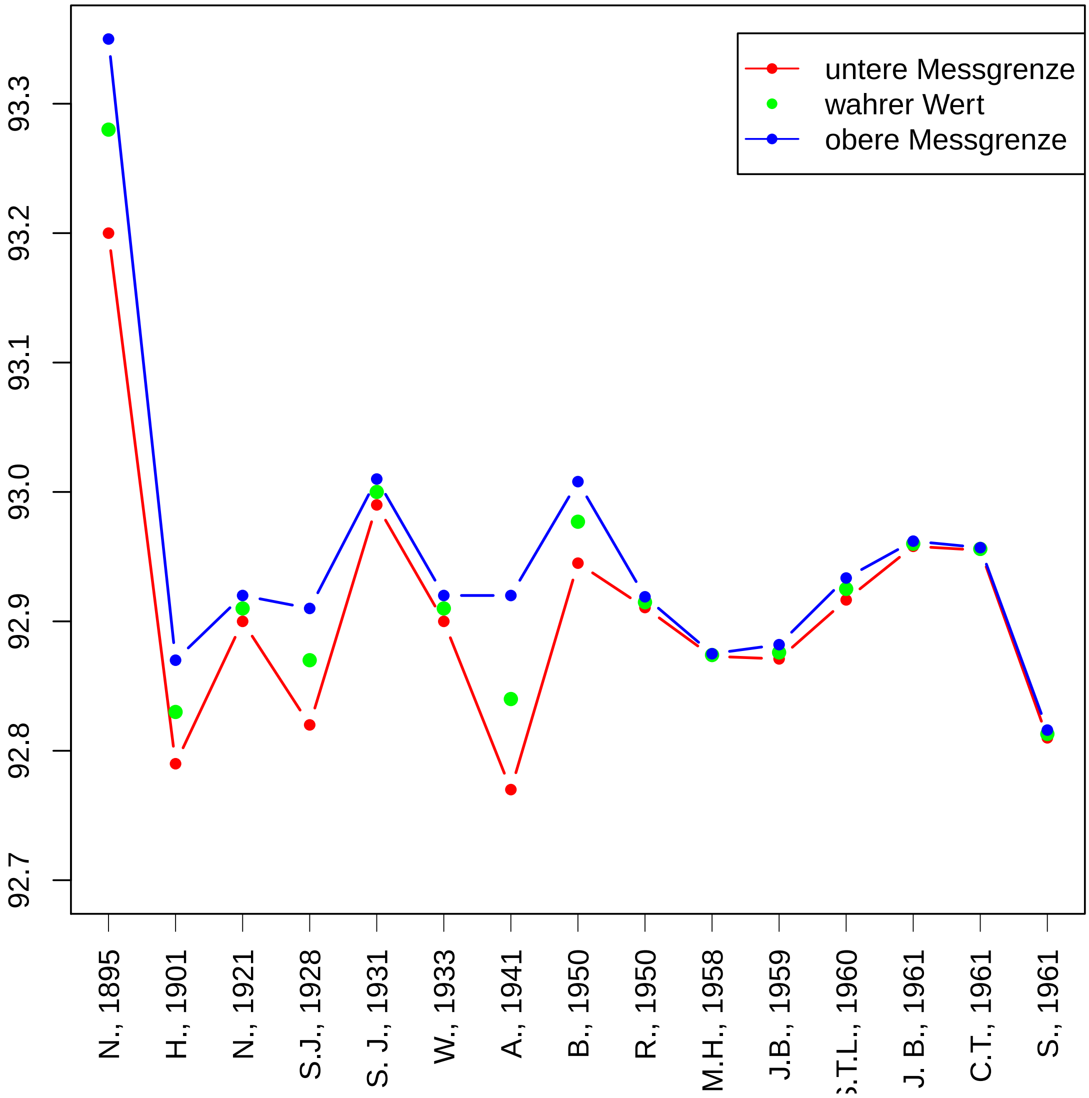

Abbildung 1: Messergebnisse der Astronomischen Einheit über die Jahre

Es ist erkennbar, dass die wahren Werte immer innerhalb der Intervalle der beobachteten Werte liegen und wurden nie exakt gemessen.

Messfehler sind jeweils einer einzelnen Beobachtung zugeordnet. Sie werden nach ihren Ursachen und charakteristischen Eigenschaften unterschieden.

In dieser Arbeit werden insbesondere Messfehler in den Einflussvariablen betrachtet. In der Literatur gibt es unterschiedliche Meinungen zu den Auswirkungen von Messfehlern in der Zielvariablen. Diese Problematik wird ebenso thematisiert und in der Seminararbeit „Fehler in der abhängigen Variable“ Markovic (2014) untersucht.

Messfehler in Einflussvariablen haben folgende Auswirkungen:

- sie verursachen Abweichungen der Parameterschätzer in den linearen Modellen
- sie führen zu dem (manchmal hochgradigen) Verlust des Ausmaßes der Zusammenhänge zwischen Variablen (loss of power)
- sie verdecken die Dateneigenschaften, was die graphische Darstellung der Datenanalyse erschwert.

Die ersten zwei Punkte werden "doppelseitiger Schlag"(double whammy) des Messfehlers genannt. Wenn man den dritten Punkt hinzunimmt, werden sie "dreifacher Whammy"genannt.(Carrol et al., 2006, S. 1)

In der Literatur werden meistens sogenannte klassische Messfehler beschrieben. Dabei wird der wahre Wert mit einem additiven Fehler und meistens einer konstanten Varianz gemessen. In dem folgenden Kapitel werden weitere Arten von Messfehlern und entsprechende Messfehlermodelle vorgestellt.

Im dritten Kapitel werden die Auswirkungen von Messfehlern in den Kovariablen auf die Parameterschätzungen in der linearen Regression gezeigt. Außerdem werden in dieser Arbeit Ansätze von zwei Methoden zur Messfehlerkorrektur vorgestellt.

2 Übersicht Messfehler

2.1 Grundlagen Messfehlertheorie

Bei der Informationsgewinnung kommt es darauf an, dass die bestimmten Größen genau erfasst und durch einen Zahlenwert oder eine Grafik dargestellt werden können. Das Messergebnis ist der Schätzwert der Messgröße, der aus der Auswertung einer Messung resultiert. Es stellt sich dabei die Frage nach der Vertrauenswürdigkeit derartiger Zahlenwerte, d.h. nach der Messunsicherheit und der Messungenauigkeit. Die Unsicherheit bzw. Ungenauigkeit des Messwertes wird von Fehlern erhöht. Man kann dabei Fehler nach der Art ihres Auftretens in systematische und zufällige (statistische) Fehler, aber auch nach der Art ihrer Entstehung in subjektive und objektive Fehler einteilen.(Gräber, 2009, S. 2, Kapitel 6.1)

Abbildung 2 liefert eine Übersicht der "natürlichen" statistischen Fehler:

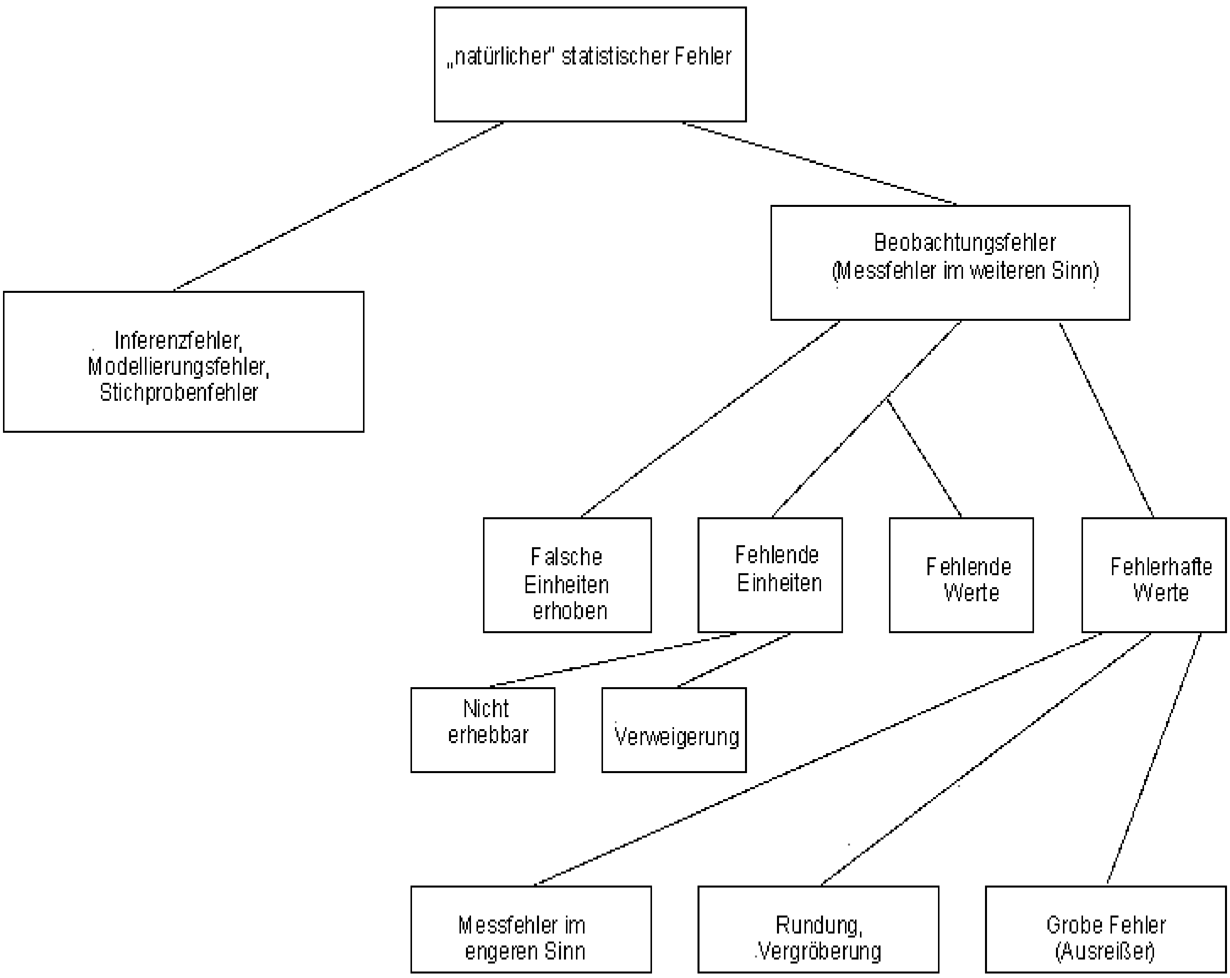

Abbildung 2: Art der statistischen Fehler (Augustin and Wiencierz, 2013)

Ursachen für einen systematischen Messfehler können personenspezifisch, itemspezifisch, itemkontextbezogen oder erhebungskontextbezogen sein. Beispielsweise sind personenspezifische Effekte insbesondere dann zu erwarten, wenn die unabhängigen und die abhängigen Variablen bei denselben Probanden gemessen werden. (Rathmann, 2014, S. 127)

Bei jeder Untersuchung stellt sich die Frage nach der Qualität des Messvorgangs, die den Untersuchungserfolg und die Aussagefähigkeit der Ergebnisse entscheidend beeinflusst. Messfehler sind bei jedem Messvorgang, z.B. bei Persönlichkeitstests in der Psychologie oder Einstellungsfragebögen im Marketing, unvermeidbar. Daher muss es bei Messungen das Ziel sein, die Messqualität zu beurteilen und Messfehler zu minimieren. Traditionell werden in diesem Zusammenhang in der klassischen Testtheorie Haupt- und Nebengütekriterien betrachtet. (Albers et al., 2007, S. 375)

Der Schwerpunkt des Modells der klassischen Testtheorie liegt auf der Genauigkeit einer Messung bzw. auf der Größe des jeweiligen Messfehlers. Daher wird sie oft auch als Messfehlertheorie bezeichnet.

Es werden folgende Hauptgütekriterien eines Messinstruments unterschieden:

1. **Objektivität**
   Objektive Messergebnisse liegen vor, wenn verschiedene Personen, die die Messungen unabhängig voneinander vornehmen, zu den gleichen Messergebnissen gelangen. Die Objektivität wird über drei verschiedene Aspekte weiter differenziert:

   - Durchführungsobjektivität ist gegeben, wenn der Untersuchungsleiter die Probanden nicht durch seine eigenen Vorstellungen und sein Untersuchungsziel beeinflusst.
   - Auswertungsobjektivität ist dadurch gekennzeichnet, dass es bei der Auswertung der Messergebnisse keine Freiheitsgrade gibt.
   - Schließlich betrifft die Interpretationsobjektivität den Spielraum bei der Interpretation der Messergebnisse. Interpretationsobjektivität ist dann vorhanden, wenn aus gleichen Ergebnissen gleiche Schlussfolgerungen gezogen werden.

2. **Reliabilität (Zuverlässigkeit)**
   Die Reliabilität betrifft die Zuverlässigkeit und Stabilität eines Messinstruments. Das Kriterium bezieht sich auf die Frage, wie gemessen wird, und fordert, dass die Messergebnisse bei wiederholter Messung reproduzierbar sein sollten.

3. **Validität (Gültigkeit)**
   Die Validität bezieht sich auf die Gültigkeit und materielle Genauigkeit eines Messinstruments. Im Rahmen der Validitätsprüfung stellt sich die Frage, ob mit einem

Messinstrument genau das gemessen wird, was gemessen werden soll.

Die Reliabilität (Zuverlässigkeit) lässt sich anhand der Zerlegung des beobachteten Messwertes („observed score“, X_O) verdeutlichen :

$$X_O = X_T + X_S + X_R$$

Der beobachtete Messwert setzt sich demnach aus dem „wahren“ Wert („true score“, X_T), einem systematischen Fehler („systematic error“, X_S) und einem zufälligen Fehler zusammen („random error“, X_R).
Die Reliabilität bezieht sich auf den zufälligen, unsystematischen Fehler:
Bei einer vollkommen reliablen Messung gibt es keine Zufallsfehler ($X_R = 0$). Dieser Zufallsfehler ist ein variabler Fehler, der alle möglichen Einflussfaktoren enthält, die bei jeder Messung die Messergebnisse mit unterschiedlicher Stärke ohne erkennbare Systematik beeinflussen.(Albers et al., 2007, S. 376)
Alternativ zu dieser Betrachtungsweise wird in der klassischen Testtheorie die Reliabilität ausgedrückt als Verhältnis der Varianz der wahren Messwerte zur Varianz der beobachteten Messwerte. Da die Varianz der wahren Werte nicht beobachtbar ist, wird der Reliabilitätskoeffizient über Korrelationen geschätzt. Der zufällige Fehler, der für den Unterschied zwischen der Varianz der wahren und beobachteten Messwerte sorgt, kann auf verschiedene Ursachen zurückgeführt werden (z.B. Messfehler aufgrund unpräziser Fragestellungen, Einflüsse unterschiedlicher Interviewer, situative Unterschiede). Angesichts dieser Ursachen wurden unterschiedliche Methoden der Reliabilitätsprüfung entwickelt. Zu näheren Informationen über diese Methoden wird an dieser Stelle auf die Quelle (Albers et al., 2007, S. 377) verwiesen.

2.2 Stochastische und systematische Fehler

Führt man eine Messung durch, so wird der gemessene Wert fast immer vom wahren Wert abweichen, die Messung ist also mit einem Fehler behaftet. Dieser Fehler wird in den systematischen und in den zufälligen Fehler aufgeteilt. In Abbildung 3 ist die entsprechende Aufteilung der Messfehler dargestellt.

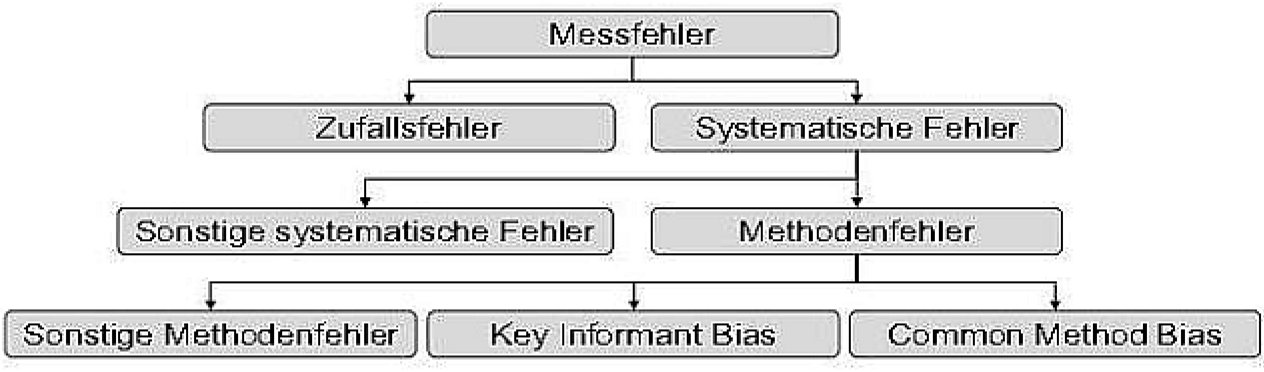

Abbildung 3: Systematisierung möglicher Messfehler (Albers et al., 2007, S. 136)

Zufällige bzw. stochastische Fehler werden von messtechnisch nicht erfassbaren Änderungen der Maßkörper, der Messgeräte, des Messgegenstandes, der Umwelt und der Beobachter hervorgerufen. Auch numerische Berechnungen, z.B. mit einem Computer, sind in der Regel nicht beliebig genau, sondern fehlerhaft. So führen Rundungen und endliche Tiefe jedes Iterationsverfahrens zu stochastischen Fehlern. Die zufälligen Fehler können bei einer Einzelmessung weder nach ihrem Betrag noch nach ihrem Vorzeichen bestimmt werden, sie sind daher nicht zu korrigieren und machen das Ergebnis unsicher. Wiederholt man eine Messung unter gleichen Bedingungen, so weichen die einzelnen Messwerte voneinander ab, sie „streuen". Sie können nach den Regeln der mathematischen Statistik in ihrer Gesamtheit zahlenmäßig erfasst und durch eine statistische Kenngröße charakterisiert werden.(Gräber, 2009, S. 2, Kapitel 6.1)

Der Zufallsfehler ist also die zufällige Abweichung einer Beobachtung vom theoretisch wahren Wert, durch die die Reliabilität einer Messung beeinflusst wird. Reliabilität kann als der Grad der Messgenauigkeit eines Instruments definiert werden. Sie ist umso höher, je kleiner der zu einem Messwert gehörende Fehleranteil ist. Perfekte Reliabilität würde bedeuten, dass ein Instrument in der Lage ist, den wahren Wert ohne jeden Messfehler zu erfassen. Somit müsste eine vollständig reliable Messung bei wiederholter Befragung der selben Respondenten, sofern sich der wahre Wert nicht verändert, immer die selben Ergebnisse liefern. Eine perfekte Korrelation der Messergebnisse beider Messreihen wäre die Folge. Es gibt verschiedene statistische Tests für Reliabilität (z.B. Retest-Reliabilität, Paralleltest-Reliabilität usw.), auf die an dieser Stelle aber nicht weiter eingegangen wird. Reliabilität ist eine notwendige, aber keine hinreichende Bedingung für die Validität einer Messung. (Albers et al., 2007, S. 135)
Der methodische Fehler beeinflusst die Validität einer Messung. Die Validität (Gültigkeit) ist das wichtigste Gütekriterium. Sie beschreibt, ob ein Test oder ein Konstrukt in der Lage ist, den Sachverhalt zu messen, der auch gemessen werden soll. Auch wenn eine hohe Reliabilität vorliegt, kann ein Test oder eine Befragung nutzlos sein, wenn nicht der richtige

Sachverhalt gemessen wird.(Albers et al., 2007, S. 136)

Systematische bzw. statistische Fehler werden hauptsächlich durch Unvollkommenheiten der Maßverkörperung der Messgeräte, der Messverfahren und des Messgegenstandes sowie von messtechnisch erfassbaren Einflüssen der Umwelt und persönlichen Einflüssen der Beobachter hervorgerufen. Sie haben:

- ein bestimmtes Vorzeichen (+ oder -)
- unter gleichen Bedingungen den gleichen Betrag, d.h. sie können durch Wiederholung der Messung nicht festgestellt werden, sondern nur durch ein anderes (genaueres) Messgerät oder Messverfahren

Der systematische Fehler besitzt bei gleichen Messvorgängen die gleiche Struktur, z.B. eine Uhr geht um 1 Promille vor, so ergibt dies bei jeder Messung von 10 Sekunden einen systematischen Fehler von 0.01 Sekunden.

Der statistische Fehler beeinflusst die Messergebnisse in unterschiedlicher, unvorhersehbarer Weise. Die Größe des Fehlers lässt sich jedoch abschätzen, wenn Ergebnisse mehrerer gleichartiger, etwa wiederholter Messungen, zur Verfügung stehen. So stellt sich der zufällige Fehler oft als Summe vieler Elementarfehler (z.B. Ablesefehler, Umgebungsveränderung) dar und kann damit - wegen der Gültigkeit des zentralen Grenzwertsatzes - als normalverteilt angenommen werden.(Gräber, 2009, S. 2, Kapitel 6.1)

Die Ergebnisse $x_1, ..., x_n$ wiederholter Messungen eines wahren Wertes μ können dann als Realisationen einer $N(\mu^*, \sigma^2)$-verteilten Zufallsgröße X aufgefasst werden. Hier ist der zufällige Fehler bei der i-ten Messung

$$\epsilon_i = x_i - \mu^*,$$

während der systematische Fehler, den man auch **Bias** nennt, gegeben ist durch

$$b = \mu^* - \mu$$

Damit ergibt sich nun

$$\begin{aligned} x_i &= \mu + b + \epsilon_i \\ &= \text{wahrer Wert} + \text{systematischer Fehler} + \text{zufälliger Fehler}. \end{aligned}$$

Mit den Messwerten $x_1, ...x_n$ lassen sich $\mu^* = \mu + b$ und σ z.B. durch

$$\overline{x} = \tfrac{1}{n} \sum x_i \text{ und } s = \sqrt{\tfrac{1}{n-1} \sum (x_i - \overline{x}^2)} \text{ für } i \in [1, n]$$

schätzen.(Hartung et al., 2009, S. 322)

Darüber hinaus können Konfidenzintervalle für μ^* und σ^2 angegeben werden. Dadurch lässt sich ein Eindruck von Größe und Struktur des zufälligen Fehlers gewinnen.

Über die Größe des systematischen Fehlers b hingegen lassen sich aus den Messwerten $x_1, ..., x_n$ allein keine Informationen ziehen, da dieser für alle n Werte gleich ist. Es muss also dafür gesorgt werden, dass ein systematischer Fehler weitgehend vermieden wird. Dabei hilft z.B. die Kalibration bzw. Justierung der verwandten Messgeräte. Kann man annehmen, dass kein systematischer Fehler vorliegt, so liefert $\overline{x}$ eine Schätzung für den wahren Wert μ, und ein Konfidenzintervall für μ entspricht dann einem Konfidenzintervall für den Parameter μ^*. (Hartung et al., 2009, S. 322)

2.3 Additive und multiplikative Fehler

Aus dem vorherigen Kapitel geht hervor, dass die systematischen Messfehler eine (deterministische) Funktion der systematischen Variablen darstellen. Das bedeutet:

$$U = f(X)$$

wobei U der systematische Messfehler und X der wahre (latente) Wert sind.
Im einfachsten Fall sind die Funktionswerte konstant.

Für die beobachteten Variablen gilt dann

$$W = X + U.$$

In diesem Fall wird über einen **additiven** oder **konstanten systematischen Fehler** gesprochen.

Eine andere Möglichkeit ist der Fall eines **multiplikativen** oder **proportionalen systematischen Fehlers**. Hier ist der Fehler proportional zur systematischen Variablen:

$$U = pX$$

Daraus ergibt sich, dass die beobachtete Variable W ebenfalls proportional zu der systematischen Variable X ist. Mit $a := 1 + p$ gilt

$$W = X + pX = aX$$

Beide Fälle werden verallgemeinernd als **linearer systematische Fehler** bezeichnet, wenn gilt :

$$W = aX + U$$

(Schneeweiß and Mittag, 1986, S. 21)

2.4 Differentieller und nicht-differentieller Fehler

In diesem Kapitel wird eine weitere Differenzierung von Messfehlern vorgestellt. Die Unterscheidung zwischen differentiellen und nicht-differentiellen Messfehlern ist sowohl bei der Datenerhebung als auch bei der Datenauswertung von großer Bedeutung, da die gesuchten Größen (z.B. Angst, Herzinfarktrisiko usw.) sehr oft nicht messbar oder nicht beobachtbar sind.

Wenn die unbekannte Variable X nicht beobachtbar oder wegen fehlender Erhebungen nicht vorhanden ist und eine ähnliche Variable W keine anderen Informationen über die Response Y als X und die fehlerfrei gemessene Variable Z enthält, ist W als mit nicht-differentiellem Fehler behaftete Variable äquivalent zu X. Formal hängt die Verteilung von $Y|(X, W, Z)$ nur von (X, Z) ab. In dem Fall wird W als Surrogat bezeichnet. Im anderen Fall ist der Messfehler differentiell.

Viele Probleme können bei nicht-differentiellen Messfehlern auftreten. Z.B. wenn die Erhebungen der wahren und beobachteten Variablen in einemm bestimmten Zeitpunkt und der Response erst später durchgeführt werden.

So ist in einer Brustkrebs-Studie die wahre Variable das Ernährungsverhalten vor der Diagnose. In den Fallkontrollstudien ist die berichtete Ernährung erst nach dem Befund beobachtbar. Eine Frau, bei der Brustkrebs diagnostiziert wurde, stellt möglicherweise ihre Ernährung um. Die Messung des berichteten Ernährungsverhaltens nach dem Befund ist mit den Krebsauswirkungen korreliert. Anders wäre es bei der Betrachtung der Ernährung vor dem Befund.(Carrol et al., 2006, S. 36)

Der Grund, warum nicht-differentielle Messfehler wichtig sind, ist:
Man kann den Response-Parameter bei der gegebenen wahren Einflussvariablen in einem Modell schätzen, selbst, wenn die wahre Variable X nicht beobachtbar ist. Anders ist es bei den differentiellen Messfehlern. Außer in ein paar speziellen Fällen, sollte man die wahre Variable beobachten können.

Differentielle Fehler können beispielsweise in biometrischen Untersuchungen relevant sein, wenn kranke Individuen ihr Verhalten (etwa Ernährungs- und Schlafgewohnheiten, Medikamentenkonsum, Rauch- und Suchtverhalten) in der Vergangenheit im Rückblick anders betrachten als gesunde Individuen. Bei selbst gemessenen oder selbst berichteten Variablen führt diese verschobene Wahrnehmung zu differentiellen Fehlern, die vom Krankheitsverlauf und damit von der Response abhängen. Das Vorliegen eines nicht-differentiellen Fehler ist folglich kritisch zu prüfen.(Carrol et al., 2006, S. 37)

Angenommen ein differentieller Messfehler beeinflusst die ZielVariable, d.h.

$$[Y|X,W] = [Y|X]$$

Für Y gibt es keine weitere Information in W oder X, wenn X bekannt ist. Dann kann das Fehler- und Haupt-Modell aufgesplittet werden:

$$[Y|W,X] = [Y|X]\,[W|X]\,[X] \mapsto f(Y|X)$$

Z.B. ist Y die Wahrscheinlichkeit einen Herzinfarkt zu bekommen, wenn ein bestimmter (bekannter!) Blutdruck X vorliegt.

Aus substanzieller Sicht:

- Messprozess und Zielvariable Y sind unabhängig.
- Blutdruck an einem bestimmten Tag ist irrelevant für das Herzinfarktrisiko, wenn bekannt ist, dass der Proband ein blutdrucksenkendes Medikament einnimmt.
- Durchschnittsexposition ist irrelevant, wenn die individuelle Exposition bekannt ist.
- Menschen können ihr Ernährungsverhalten anders betrachten, wenn sie bereits einen Herzinfarkt hatten.

Nicht-differentielle Fehler haben den Vorteil, dass sich die Untersuchung der einfachen linearen Regression vereinfacht und die Bestimmung der Regressionsparameter grundsätzlich auch mit der fehlerbehafteten Variable W möglich ist. Wegen der Gültigkeit des Transformationssatzes gilt

$$\begin{aligned} E(Y|W) &= E\{E(Y|X,W)|W\} \\ &= E\{E(Y|X)|W\} \\ &= E(\beta_o + \beta_x X|W) \\ &= \beta_0 + \beta_x E(X|W). \end{aligned}$$

(Carrol et al., 2006, S. 38)

2.5 Klassischer und Berkson Fehler

Zufällige Messfehler sind auf zwei unterschiedliche Arten darstellbar, welche trotz ihrer technischen Ähnlichkeiten unterschiedliche Auswirkungen auf die Regressionsanalyse haben und nachfolgend vorgestellt werden.

Im klassischen Fehlermodell ist W die fehlerbehaftete Messung der wahren latenten Variable X, modelliert durch den Zusammenhang

$$W = X + U \text{ mit } U_{ij}|X_i \sim N(0, \sigma^2),$$

wobei U und X stochastisch unabhängig sind.

In diesem Fall gilt $E(W|X) = X$ und W ist folglich eine unverzerrte Messung für X. (Carrol et al., 2006, S. 26)

Betrachtet man stattdessen

$$X = W + U \text{ mit } U_{ij}|X_i \sim N(0, \sigma^2),$$

wobei U und W stochastisch unabhängig sind, so gilt $E(X|W) = W$.

Man erhält trotz der Ähnlichkeit ein anderes Modell. Dieser Fehler wird Berkson-Fehler genannt und spiegelt eine andere Ausgangslage wieder.

Während beim klassischen Fehler der exakte Wert X durch zufällige Messfehler verdeckt wird, stellt sich beim Berkson-Fehler der wahre nicht beobachtbare Wert X erst als Ergebnis einer (kontrollierbaren) Variablen W ein. Berkson-Modelle finden bei epidemiologischen Studien ihre Anwendung.

Messfehler müssen mit erheblicher Vorsicht behandelt werden. Der Unterschied zwischen dem Berkson und dem klassischen Messfehler ist groß, insbesondere wenn man vorhat, a priori zu erforschen und vor allem wenn man versucht den Einfluss zu kalkulieren. Es gibt einige technische Ähnlichkeiten zwischen dem klassischen Fehler und dem Berkson Fehler, aber in den Einflussberechnungen treten unterschiedliche Aspekte auf. Wenn man sich überzeugen möchte, dass es trotz Messfehlern in Variablen einen großen statistischen Einfluss auf die gesuchte Größe gibt, muss angenommen werden, dass der Messfehler mit einer gegebenen Varianz Berkson und nicht klassisch ist.(Carrol et al., 2006, S. 27)

Wie unterscheidet sich der Berkson vom klassischen Fehler? Wenn man zwischen den beiden Messfehlern zu wählen hat, , vor allem wenn die Messung wiederholt werden kann, handelt es sich grundsätzlich um einen klassischen Fehler und die fehleranfällige Kovariable mit Hilfe von Messmitteln eindeutig bei einer Person gemessen wurde. Zum Beispiel, wenn Menschen Fragen zum Essverhalten beantworten oder bei der Messung ihres Blutdrucks, handelt es sich um einen klassischen Messfehler.

Anders verhält es sich, wenn alle Personen einer kleinen Gruppe oder einer sozialen Schicht Angaben zu einer fehleranfälligen Variable machen. Z. B. zeigen Textil- oder Bergarbeiter bei der gleichen Beschäftigungsdauer das gleiche Staubbelastungsbild. Das wahre Bild ist bei jedem einzelnen Individuum jedoch anders. In diesem Fall handelt es sich typischerweise um einen Berkson-Messfehler.(Carrol et al., 2006, S. 27)

Weiterer Unterschied zwischen Berkson und klassischem Fehler ist:

- $Var(W) \succ Var(X)$ für den klassischen Fehler und $Var(X) \succ Var(W)$ für den Berkson-Fehler.

In der Praxis besteht die Wahl nicht immer nur zwischen dem klassischen additiven Messfehler-Modell ($W = X + U$) und dem klassischen Berkson-Fehler-Modell ($X = W + U$). Für die Messungen, die auf den Angaben der Befragten basieren, sind komplexere Modelle, die solche Abweichungen berücksichtigen, notwendig. Die einzelnen Größen werden gemessen und die Messungen ebenfalls wiederholt, aber die Abweichungen müssen verdeutlicht und berücksichtigt werden. Dafür ist das allgemeine klassische Fehler-Modell ($W = \gamma_0 + \gamma_x^t X + \gamma_z^t Z + U$, $E(U|X_Z) = 0$) geeignet.

Anstatt des Berkson Modells wird für ad hoc Zwecke häufig die allgemeine Regressionskalibrierung angewendet.(Carrol et al., 2006, S. 26-28)

Grundlage der Regressionskalibrierung ist es, im Hauptmodell, in dem die Zielgröße Y auf die Einflussgrößen (X, Z) regressiert wird, die fehlerbehaftete KoVariable X durch die Regression von X auf (Z, X^*) zu ersetzen. Dabei soll Z ohne Fehler gemessen werden können und die Hilfsgröße X^* sollte in Beziehung zu der Variable X stehen. X^* kann z. B. die fehlerbehaftete Messung von X sein. Nach diesem Vorgang kann ohne Weiteres eine gewöhnliche Standardanalyse durchgeführt werden. Der theoretische Hintergrund und das genaue Vorgehen wurden in der Seminararbeit (Le, 2014) vorgestellt.

3 Messfehler in der linearen Regression

3.1 Übersicht über die linearen Regressionsmodelle

Das zentrale Problem in einem linearen Modell mit fehlerbehafteten Daten ist die Parameterschätzung. Wenn die fehlerfreie systematische Variable X nicht direkt beobachtbar ist, muss man auf die fehlerbehaftete beobachtbare Variable W zurückgreifen und mit ihrer Hilfe möglichst genaue Schätzungen konstruieren.

Man versucht zunächst eine Schätzung mit der Methode der kleinsten Quadrate (KQ-Methode) auf die Regression $(Y, X|W)$ anzuwenden. Die so gewonnene Kleinstquadrat-Schätzer (KQ-Schätzer) sind i.d.R. inkonsistent. (Schneeweiß and Mittag, 1986, S. 32)

In der Literatur wird oft über Messfehler in der linearen Regression geschrieben und der Schluss gefasst, dass deren Einfluss eine nur geringe Abweichung der Schätzung verursacht. Diese Schlussfolgerung muss mit Vorsicht genossen werden. Die Messung W, der wahre Wert X, und weitere mögliche Variablen in einem Regressionsmodell können abhängig sein. Genauer gesagt, der Effekt des Messfehlers in einem linearen Modell muss unter Berücksichtigung anderer Variablen und deren Verteilungen betrachtet werden.

Der Einfluss der Messfehler in der linearen Regression variiert in Abhängigkeit von den folgenden Merkmalen:

- ist das Regressionsmodell uni- oder multivariat
- ist die fehlerbehaftete Einflussgröße uni- oder multivariat
- ist die Messung fehlerhaft

Die Verteilung des Messfehlers ist für die Auswirkung auf ein lineares Modell entscheidend. Davon hängt das zu wählende Verfahren für die Messfehlerkorrektur ab. (Carrol et al., 2006, S. 41)

Die grundlegenden Parameter der Fehler-Modelle und der linearen Regression für $(Y, X|W)$ sind in Tabelle 2 aufgeführt. Für sämtliche Fälle wurde folgendes Regressionsmodell betrachtet:

$$Y = \beta_0 + \beta_x X + \epsilon$$

wobei X und ϵ unabhängig sind und $E(\epsilon) = 0$ und $Var(\epsilon) = \sigma^2$.

Error Modell	ρ^2_{xw}	**Intercept**	**Slope**	**residual variance**
Differential	ρ^2_{xw}	$\beta_0 + \beta_x\mu_x - \frac{\beta_x\sigma_{xw}+\sigma_{ew}}{\sigma^2_w}\mu_w$	$\beta_x(\frac{\sigma_{xw}}{\sigma^2_w} + \frac{\sigma_{ew}}{\sigma^2_w})$	$\sigma^2_e + \beta^2_x\sigma^2_x - \frac{(\sigma_{xw}\beta_x+\sigma^2_{ew})}{\sigma_w{}^2}$
Surrogate	ρ^2_{xw}	$\beta_0 + \beta_x\mu_x - \frac{\beta_x\sigma_{xw}}{\sigma^2_w}\mu_w$	$\beta_x(\frac{\sigma_{xw}}{\sigma^2_w})$	$\sigma^2_e + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
Classical	$\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_{ue}}$	$\beta_0 + \beta_x\mu_x(1-\rho^2_{xw})$	$\beta_x(\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_{ue}})$	$\sigma^2_e + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
B/C mixture	$\frac{\sigma^4_L(\sigma^2_L+\sigma^2_{u_b})^{-1}}{(\sigma^2_L+\sigma^2_{ue})}$	$\beta_0 + \beta_x\mu_x(1 - \frac{\sigma^2_L}{\sigma^2_l+\sigma^2_{ue}})$	$\beta_x(\frac{\sigma^2_L}{\sigma^2_l+\sigma^2_{ue}})$	$\sigma^2_e + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
Berkson	$\frac{\sigma^2_x-\sigma^2_{ub}}{\sigma^2_x}$	β_0	β_x	$\sigma^2_e + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
No error	1	β_0	β_x	σ^2_e

Tabelle 2: Übersicht zur Korrektur der Parameter: quadrierte Korrelation, Intercept, Steigung und ResidualVarianzen in einem differentiellen, Surrogat, klassischen, klassisch-Berkson-gemischten, Berkson und fehlerfreien Regressionsmodell (Carrol et al., 2006, S. 50)

3.2 Systematische Messfehler in der Regression

Wir betrachten eine einfache lineare Regression mit folgendem Aufbau:

$$Y = \beta_0 + \beta_x X + \epsilon, \text{ wobei } \epsilon \sim NV(0, \sigma_e^2)$$

Eine fehlerbehaftete Variable W ist über eine feste funktionale Abbildung $W = f(X)$ mit der fehlerfreien Variable X verknüpft. Wir beschränken uns hier nur auf die linearen Verknüpfungen und damit das Verhalten bei konstanten und proportionalen Fehlern.

Zuerst stellen wir eine fehlerbehaftete Variable $W = X + d$, die um einen konstanten Messfehler d von der wahren Variablen abweicht, dar. In diesem Fall verändert sich die angegebenen Modellgleichung, und es gilt

$$Y = \beta_0 + \beta_X^*(X + d) + \epsilon$$

Durch Umformung ergibt sich

$$Y = \beta_0^* + \beta_X^* X + \epsilon$$

Als Parameterschätzer erhalten wir einen verzerrten, inkonsistenten Wert $\beta_0^* = \beta_0 + \beta_X d$ sowie einen konsistent geschätzten Wert $\beta_X^* = \beta_X$. Die beobachteten Werte sind um einen konstanten Wert d nach rechts und in der Folge um den Wert $\beta_X d$ nach oben verschoben.

Nimmt man eine mit einem proportionalen Fehler behaftete Variable $W = cX$ an, dann sieht die Modellgleichung folgendermaßen aus:

$$Y = \beta_0^* + \beta_X^* cX + \epsilon$$

und die verzerrten KQ-Schätzer ergeben sich als β_0^* und $\beta_X^* = \frac{1}{c}\beta_X$.

In beiden Fällen erhalten wir durch die KQ-Schätzung verzerrte Schätzer, die um den bekannten konstanten Fehler d und den proportionalen Fehler c korrigiert werden können. Ist der systematische Messfehler bekannt, so können auch die gemessenen, fehlerbehafteten Werte W vor Durchführung der Regression korrigiert werden. (Schneeweiß and Mittag, 1986, S. 32-37)

3.3 Stochastische Messfehler in der Regression

3.3.1 Klassische Messfehler in der Regression

Wir beschäftigen uns weiter mit dem linearen Regressionsmodell, das folgende Struktur hat:

$$Y = \beta_0 + \beta_X X + \epsilon, \text{ wobei } \epsilon \sim NV(0, \sigma_\epsilon^2)$$

Wenn X fehlerbehaftet ist, also $W = X + U$ mit $U \sim NV(0, \sigma_U^2)$ der klassische Messfehler ist, gilt:

$$Y = \beta_0^* + \beta_W (X + U) + \epsilon.$$

(Carrol et al., 2006, S. 42-43)

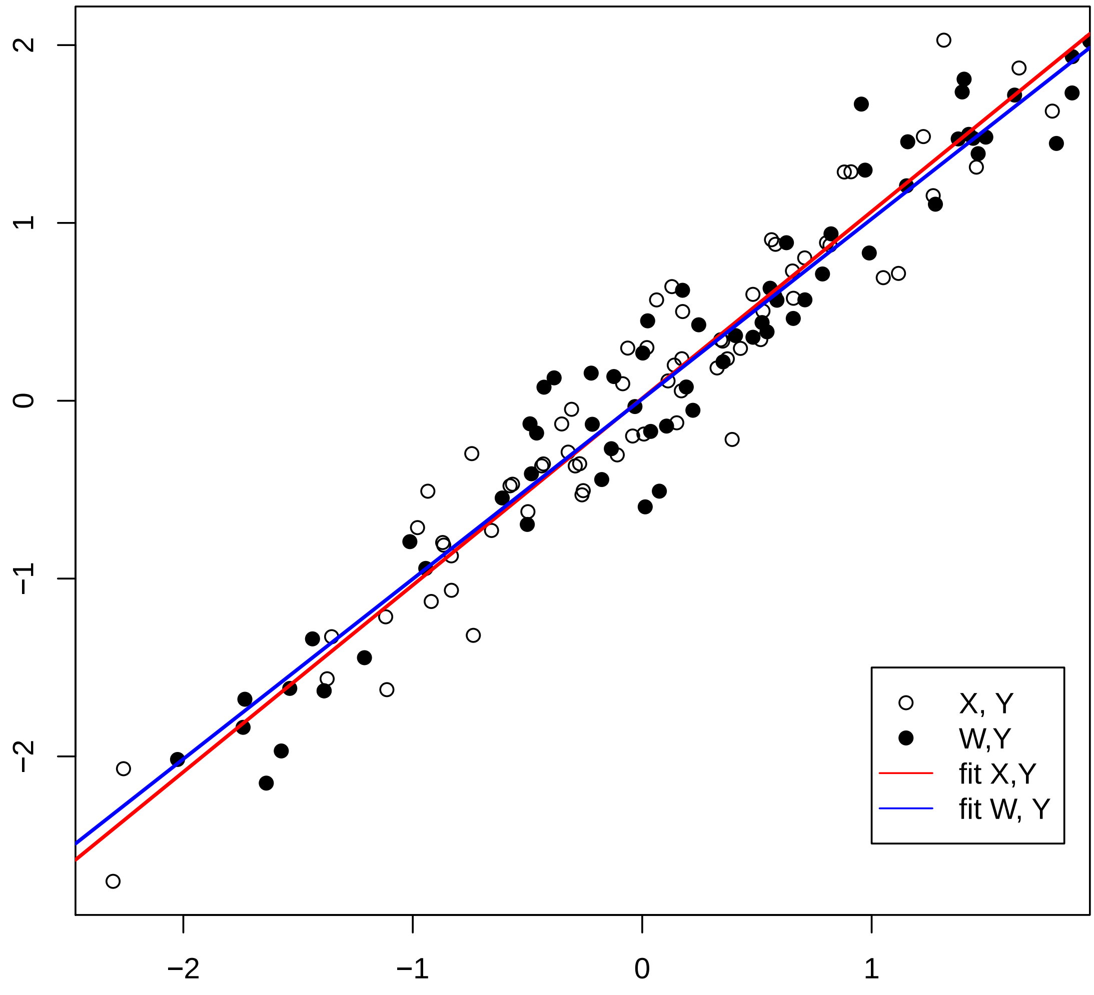

Abbildung 4: Einfache lineare Regression beim klassischen Fehler

Abbildung 4 zeigt die Folgen der klassischen Messfehler in der linearen Regression. Die rote Regressionsgerade gibt den wahren Zusammenhang zwischen Einflussvariable und Response an. Der klassische Messfehler ist hier um einen konstanten Term nach links bzw. rechts modelliert. Die fehlerbehaftete Regressionsgerade (blau) weicht von der wahren Regressionsgeraden ab. Es kommt zu einer systematischen Unterschätzung der Steigung.

Die gewöhnliche KQ-Schätzung von X auf Y ist kein konsistenter Schätzer von β_X. Anstatt β_x bekommt man $\beta_{x*} = \lambda\beta_x$, wobei

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} < 1,$$

σ_x^2 Varianz des wahren Wertes X und σ_u^2 Varianz des Messfehlers U sind.

In diesem Fall wird λ *Reliabilitätsfaktor* genannt. Da W eine fehlerbehaftete Beobachtung ist, besitzt sie einen schwächeren Einfluss auf die Response als X. Darüber hinaus vergrößert sich die Varianz der Beobachtungen. Anstelle von $Var(Y|X) = \sigma_X^2$ erhalten wir

$$Var(Y|W) = \sigma_\epsilon^2 + \frac{\beta_x^2\sigma_u^2\sigma_x^2}{\sigma_x^2+\sigma_u^2} = \sigma_\epsilon^2 + \lambda\beta_x^2\sigma_u^2 > Var(Y|X).$$

(Carrol et al., 2006, S. 44)

3.3.2 Berkson-Fehler in der Regression

Die Problematik der Auswirkung des Berkson-Fehlers in der linearen Regression ist von großer Bedeutung. Gegeben sei das folgende lineare Modell

$$Y = \beta_0 + \beta_X X + \epsilon,\ \epsilon \sim NV(0, \sigma_\epsilon^2)$$

mit $E(X|W) = W$ und $X = W + U_b$, wobei U_b ein Berkson-Fehler ist. In Abbildung 5 ist zu sehen, dass die wahre und die mit einem Berkson-Messfehler behaftete Regressionsgerade gleich sind.

Da W ein Surrogat, d.h. $E(X|W) = W$, ist, betrachten wir die Auswirkungen der wahren Werte X auf die dazu gehörenden Werte der Response.

Die Kleinste-Quadrate-Schätzungen sind beim beobachteten mit dem wahren Fall identisch, wenn wir wie im klassischen Fall annehmen, dass alle beobachteten Werte gleichermaßen realisiert werden. Es entsteht keine Verzerrung der Regressionskoeffizienten im Berkson-Modell. $X = W + U$ ergibt sich nach Festlegung von W als zufälliges Ergebnis. Betrachten wir folgendes Regressionsmodell:

$$Y = \beta_0^* + \beta_W W + \epsilon, \text{ wobei } \epsilon \sim NV(0, \sigma_\epsilon^2)$$

Da $E(X|W) = W$ (mit nicht-differentiellem Fehler behaftete Variable), $E(Y|W) = \beta_0 + \beta_X E(X|W) = \beta_0 + \beta_x W$. Daraus folgt, dass die Koeffizientenschätzer für β_0 und $\beta_X = \beta_W$ unverzerrt sind. Statt $Var(Y|X)$ erhalten wir

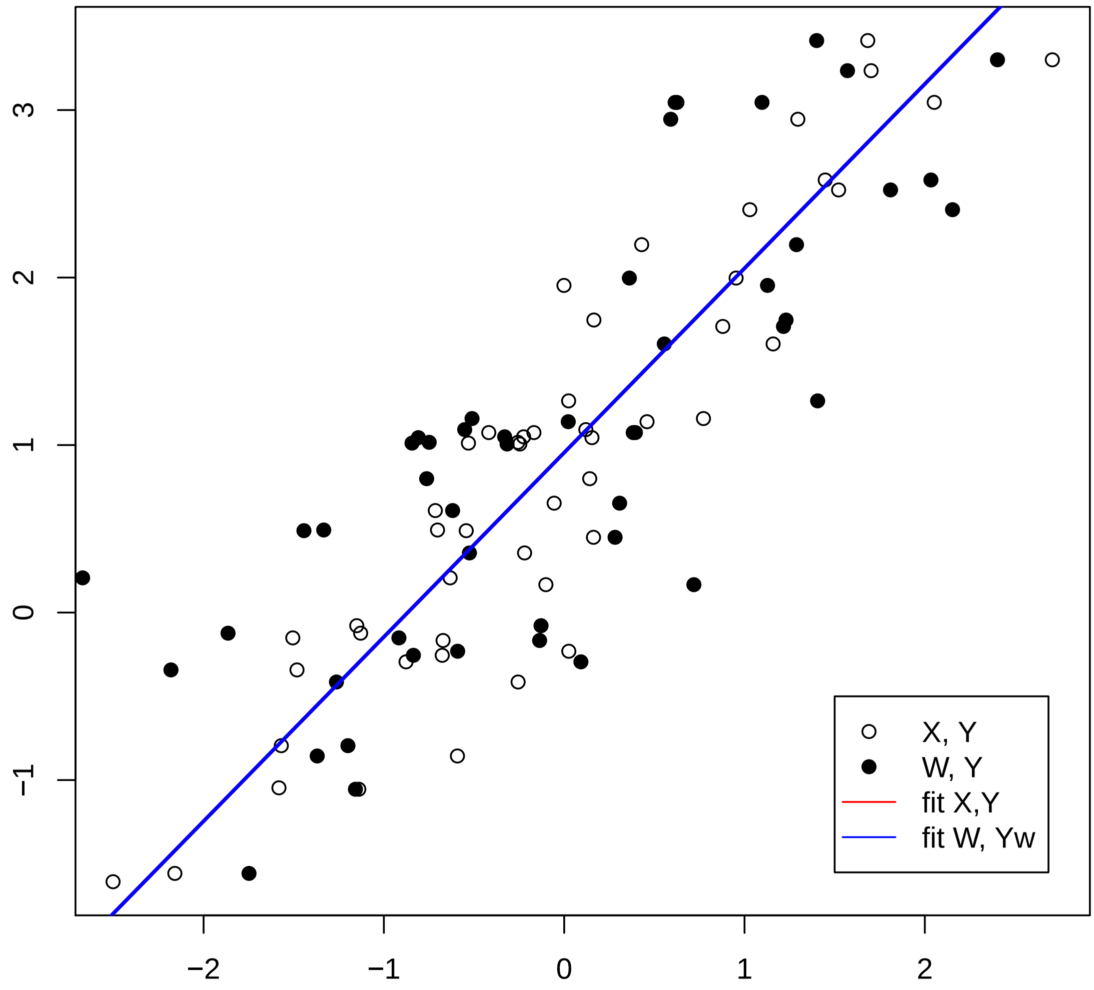

Abbildung 5: Auswirkung des Berkson-Fehlers in der linearen Regression: KQ-Regressionsgleichungen sind fast identisch und es ergibt sich in Folge keine Verzerrung der Regressionskoeffizienten.

$$Var(Y|W) = Var(Y) + Var(E(Y|W)) = \sigma_\epsilon^2 + \beta_X^2\sigma_U^2 > Var(Y|X).$$

(Carrol et al., 2006, S. 45)

Die Varianz der Regression (Y, W) ist größer als die Varianz der Regression mit den wahren Werten X (σ_ϵ^2) als Folge, dass W ein Surrogat ist.

3.3.3 Kombination von klassischen und Berkson-Fehlern

Wir betrachten ein Regressionsmodell, bei dem die Einflussvariable sowohl klassische als auch Berkson-Komponenten beinhaltet. Aus dem klassischen Fehlermodell ist bekannt, dass $W = X + U$ und der beste linearer Prädiktor für $(X|W)$ ist

$$X = \lambda W + (1 - \lambda)E(U) + U^*$$

wobei $U^* = (1 - \lambda)(X - E(X)) - \lambda U$ und der Reliabilitätsfaktor $\lambda = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2} < 1$ sind. Es ist offensichtlich, dass U^* und W unkorreliert sind. Das bedeutet, dass der klassische Fehler in einen Berkson-Fehler überführt wird.

Da das Berkson-Modell und die entsprechende Parameterschätzung unverzerrt sind, empfiehlt sich diese Betrachtung . Die Fehlerstruktur soll durch das stochastische Berkson-Fehlermodell, verzerrt durch den systematischen, proportionalen Fehler λ, dargestellt werden. (Carrol et al., 2006, S. 29)

3.3.4 Multivariate lineare Regression

In der multivariaten linearen Regression sind die Auswirkungen der Messfehler komplexer, auch für das klassische additive Fehlermodell.

Wir betrachten eine skalare X und zusätzlich eine Kovariable Z, die fehlerfrei gemessen wurde. Das lineare Modell sieht jetzt folgendermaßen aus:

$$Y = \beta_0 + \beta_x X + \beta_z^t Z + \epsilon,$$

wobei Z und β_z Spaltenvektoren sind und β_z^t der Zeilenvektor ist. Die Kleinste-Quadrate-Schätzer für die Regressionskoeffizienten sind im Fall der multivariaten Regression wie im Fall einer einfachen linearen Regression verzerrt. Statt β_x wird $\lambda_1 \beta_x$ mit

$\lambda_1 = \frac{\sigma_{x|z}^2}{\sigma_{w|z}^2} = \frac{\sigma_{x|z}^2}{\sigma_{x|z}^2 + \sigma_u^2}$, und statt β_z wird $\beta_z + \beta_x(1 - \lambda_1)\Gamma_z$ geschätzt. Γ_z ist der Regressionskoeffizient von X auf Z, d.h. $E(X|Z) = \Gamma_0 + \Gamma_z^t Z$. (Carrol et al., 2006, S. 52)

Damit wird deutlich, dass Messfehler auch auf die Koeffizientenschätzer der anderen fehlerfreien Variablen Auswirkungen haben können und deshalb nicht vernachlässigt werden dürfen.

4 Ausgewählte Methoden zur Korrektur eines Messfehlers

4.1 Momentenmethode

Wie in den vorangegangenen Kapiteln gezeigt wurde, ist der KQ-Schätzer bei fehlerbehafteten Daten verzerrt. Richtung und Umfang der Verzerrung sind vom Regressionsmodell, der Verteilung des Messfehlers und der Korrelation zwischen den geschätzten Variablen abhängig.

In der einfachen linearen Regression mit klassischem additivem Messfehlermodell wird der KQ-Schätzer als $\lambda\beta_x$ geschätzt, wobei $\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$ der Reliabilitätsfaktor ist. Wenn dieser bekannt ist, dann erhält man einen unverzerrten geschätzten β_x durch Multiplikation mit $\frac{1}{\lambda}$.

Der Reliabilitätsfaktor ist jedoch selten bekannt und muss daher geschätzt werden. Bei systematischen Messfehlern ist das nur dann möglich, wenn die Fehlerparameter (Erwartungswert und Varianz) bekannt sind. Damit kann man die Daten reinigen oder die geschätzten Parameter korrigieren. Wenn $\widehat{\sigma_u^2}$ die Schätzung der Messfehlervarianz und $\widehat{\sigma_w^2}$ die Schätzung der Varianz von W ist, dann lautet die konsistente Schätzung des Reliabilitätsfaktors

$$\widehat{\lambda} = \frac{(\widehat{\sigma_w^2} - \widehat{\sigma_u^2})}{\widehat{\sigma_w^2}}.$$

Die resultierende Schätzung ist demnach

$$\widehat{\beta_x} = \frac{\widehat{\beta_{x*}}}{\widehat{\lambda}}$$

In kleinen Stichproben ist die Stichprobenverteilung von $\widehat{\beta_{x*}}/\widehat{\lambda}$ normalerweise stark verzerrt. In diesen Fällen ist eine modifizierte Version der Momentenmethode ratsam. (Carrol et al., 2006, S. 55)

4.2 Orthogonale Regression

Es sei $Y = \beta_0 + \beta_x X + \epsilon$ und $W = X + U$, wobei ϵ und U unkorreliert sind. Während die Momentenmethode verlangt, dass die Messfehlervarianz σ_u^2 bekannt ist oder geschätzt werden kann, verlangt die orthogonale Regression das selbe für den Parameter $\eta = \sigma_e^2/\sigma_u^2$.

Der orthogonale Regressionsschätzer minimiert den orthogonalen Abstand, d.h. den kürzesten Abstand von (Y, W) zu der Gerade $y = \beta_0 + \beta_y x$, wenn $\eta = 1$:

$$SSR = \sum (y_i - \beta_0 - \beta_1 x_i^*)^2 + (x_i - x_i^*)^2) \;\rightarrow\; \min_{\beta_0\beta_1 x_i^*} SSR.$$

Die Parameter zur Lösung des Minimierungsproblems ergeben sich wie folgt

$$\beta_1 = \frac{s_y^2 - s_x^2 + \sqrt{(s_y^2 - s_x^2)^2 + 4s_{xy}^2}}{2s_{xy}}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$x_i^* = x_i + \frac{\beta_1}{\beta_1^2 + 1}(y_i - \beta_0 - \beta_1 x_i)$$

Hierbei sind (x_i^*, y_i^*) die Fußpunkte auf der Geraden, s_x^2 und s_y^2 korrigierte Stichprobenvarianzen der x_i und y_i, s_{xy}^2 die korrigierte Kovarianz von (x_i, y_i), $\bar{x}$ und $\bar{y}$ die arithmetischen Mittel von x und y.

Der orthogonale Regressionsschätzer ist der funktionale Maximum-Likelihood-Schätzer unter Annahme, dass $(X_1, \ldots, X_n)$ unbekannte feste Konstanten und die Fehler (ϵ, U) unabhängig und normalverteilt sind.

Die orthogonale Regression wird öfters als Momentenmethodenschätzung verwendet, in der nur die Kennzahl η der Fehlervarianzen bekannt oder geschätzt sein muss. Das Problem ist, dass η nicht richtig spezifiziert oder geschätzt werden kann. Die Verwendung der orthogonalen Regression mit falsch spezifiziertem Wert η ergibt oft eine unakzeptabel große Überkorrektur, die eine Abschwächung des Messfehlers verursacht.

Folglich ist das Verfahren trotz seiner Einfachheit mit Problemen verbunden. (Carrol et al., 2006, S. 57)

Weitere Korrekturmethoden bei der Parameterschätzung sind die Regressionskalibrierung und SIMEX bzw. MCSIMEX, die im Rahmen des Seminars thematisiert wurden. Die Regressionskalibrierung wurde bereits auf Seite 13 erwähnt. Die SIMEX Methode wird vor allem für die Korrektur der additiven Messfehler verwendet. Diese Methode ist besonders für komplexe Modelle mit einfacher Fehlerstruktur geeignet. Der theoretische Hintergrund sowie der SIMEX-Algorithmus wurden im Seminarbericht (Hölzl, 2014) erläutert.

5 Schlussfolgerung

In der Seminararbeit wurde versucht, die bekanntesten und in der Literatur beschriebenen Messfehlermodelle zu erläutern. Dabei sind die nicht-differentiellen und Berkson-Messfehler von großer Bedeutung. In der Praxis sind die gesuchten Größen oft latent oder nicht messbar. Um diese zu erforschen, werden sogenannte Surrogate in Betracht gezogen. Im klassischen Fehlermodell wird die fehlerbehaftete Größe als unverzerrte Schätzung der wahren fehlerfreien Variablen dargestellt, d.h. die Messfehler werden hierbei nicht beachtet. Im Gegensatz dazu betrachtet das Berkson-Fehlermodell den wahren Wert als Ergebnis der fehlerbehafteten, kontrollierbaren Beobachtung.

Weiterhin wurden die Auswirkungen von Messfehlern in der linearen Regression beschrieben. Das zentrale Problem bei einem linearen Modell mit fehlerbehafteten Daten ist die Parameterschätzung. Wenn die fehlerfreie Variable nicht direkt beobachtbar ist, muss man auf die fehlerbehaftete beobachtbare Variable zurückgreifen und mit ihrer Hilfe möglichst genaue Schätzungen konstruieren.

Für alle Arten von Messfehlern kommt es bei der linearen Regression zu inkonsistenten und verzerrten Parameterschätzern. Das verursacht im meisten Fällen die systematische Unterschätzung der Steigung der Regressionsgeraden. Die Varianz der Regression mit fehlerbehafteten Variablen bzw. Surrogaten ist deutlich größer als die Varianz der wahren Regression.

Bei den bekannten Messfehlerparametern wie Erwartungswert und Varianz ist es einfach, die Messfehlerkorrektur vorzunehmen. Jedoch sind diese in der Praxis selten bekannt. In entsprechenden Fällen sollte auf weitere Korrekturverfahren, wie die orthogonale Regression, die Regressionskalibrierung, SIMEX usw. zugegriffen werden.

Das Ziel der Seminararbeit war es, die Grundlagen zu erarbeiten und aufzuzeigen, dass die Berücksichtigung von Abweichungen in beobachteten Daten von großer Bedeutung ist. Die Auswirkungen von fehlerbehafteten Variablen in der linearen Regression hängt von der Fehlerstruktur, derer Verteilung und dem Zusammenhang mit den anderen Kovariablen ab.

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis

Albers, S., Klapper, D., Konradt, U., Walter, A. and Wolf, J. (2007). *Methodik der empirischen Forschung*, 2. edn, GWV Fachverlage GmbH, Wiesbaden.

Augustin, T. and Wiencierz, A. (2013). Wirtschafts- und Sozialstatistik Foliensatz WiSe 13/14. 14.10.2014`http://www.statistik.lmu.de/institut/ag/agmg/lehre/2013_WiSe/Wiso/WiSo_folien_kap_1.pdf`.

Carrol, R., Ruppert, D., Stefanski, L. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2. edn, Chapman Hall, Boca Raton.

Gräber, P.-W. (2009). Systemanalyse. Foliensatz Automatisierungstechnik in der Wasserwirtschaft. 14.11.2014 `http://tu-dresden.de/die_tu_dresden/fakultaeten/fakultaet_forst_geo_und_hydrowissenschaften/fachrichtung_wasserwesen/iaa/systemanalyse/studium/folder.2009-01-29.lehre/folder.2009-04-03.at/AT%206.pdf`.

Hartung, J., Elpert, B. and Klösener, K. (2009). *Statistik: Lehr- und Handbuch der angewandten Statistik*, 15. edn, Oldenbourg Wissenschaftsverlag GmbH, München.

Hölzl, A. (2014). Simulation extrapolation. 01.03.2015`http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014_WiSe/BA_MA_Seminar/Vorbereitungsmaterial/Hoelzl.pdf`.

Höpcke, W. (1980). *Fehlerlehre und Ausgleichsrechnung*, de Gruyter, Berlin.

Le, M.-A. (2014). Regressionskalibrierung. 26.02.2015`http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014_WiSe/BA_MA_Seminar/Vorbereitungsmaterial/Le.pdf`.

Markovic, N. (2014). Fehler in der abhängigen variable. 25.02.2015`http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014_WiSe/BA_MA_Seminar/Vorbereitungsmaterial/Markovic.pdf`.

Rathmann, P. (2014). *Medienbezogene Effekte von Product Placement: Theoretische Konzeption und empirische Analyse*, Springer Gabler, Wiesbaden.

Schneeweiß, H. and Mittag, H.-J. (1986). *Lineare Modelle mit fehlerbehafteten Daten*, Physica-Verlag Heidelberg Wien, Würzburg.

Eidesstattliche Erklärung

Ich, Hanna Marshalava, versichere hiermit, dass ich meine Seminararbeit mit dem Thema

Überblick über Messfehler und ihre Auswirkungen in der linearen Regression

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, wobei ich alle wörtlichen und sinngemäßen Zitate als solche gekennzeichnet habe.

München, den 15. März 2015

H. Marshalava

Hanna Marshalava