

# *Fehlende Werte*

*Vorstellung der Seminararbeit zum Zusatzkurs  
Amtliche Statistik*

*Betreuung durch: Prof. Dr. Augustin  
Referent: Matthias Speidel*

# Gliederung

*Einleitung*  
*Die Fehlerarten*  
*Fehlende Werte*  
*Ergebnis*

# Einschub

Unterschied „Fehlende Werte“

und „falsche Werte“

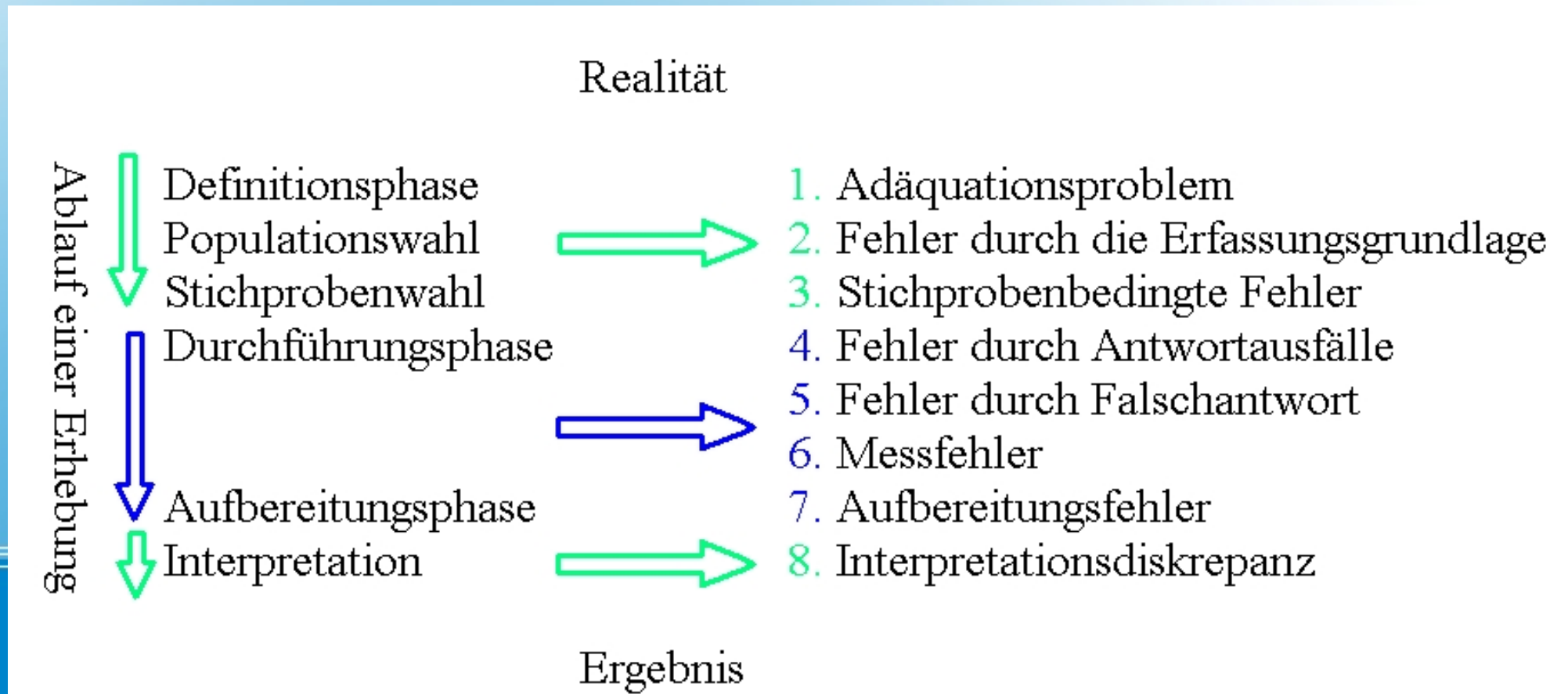
# Problematik

*statistische Ergebnisse haben weitreichende Folgen*

*Differenzen zur Realität sind Differenzen zur  
„optimalen Lösung“*

*Fall aus der Realität: Nutzen von Elterngeld*

# Die Fehlerarten



# 1. Adäquationsproblem

## *Problem:*

*Ein theoretisches Konstrukt muss auf die empirische Ebene gebracht werden*

## *Lösungsansätze:*

*sichere Kenntnisse (Substanzwissenschaftler)*

*valide und reliable Messinstrumente*

*Pre-Test und Vergleich mit anderen Instrumenten*

## 2. Fehler durch die Erfassungsgrundlage

### *Entstehung:*

*Die Erfassungsgrundlage weist eine Differenz zur eigentlichen Grundgesamtheit auf*

### *Vermeidung:*

*Genaue Bestimmung der Grundgesamtheit und ausloten der „Under-“ und „Overcoverage“*

## 3. Stichprobenbedingte Fehler

### *Entstehung:*

*Stichprobe nicht repräsentativ bezüglich des interessierenden Merkmals*

### *Vermeidung:*

*Systematische Verzerrungen vermeiden*

*Auswahlverfahren darf mit dem*

*Untersuchungsmerkmal nicht im Zusammenhang stehen*

*ausreichender Stichprobenumfang*

*wirklicher Zufallsmechanismus*

*ausreichender Stichprobenumfang*

## 4. Fehler durch Antwortausfälle

### *Entstehung:*

*Eine Einheit macht zu einer oder mehreren Merkmalsausprägungen keine Angaben oder: Informationsverlust*

### *Vermeidung:*

*Vertrauen durch Anonymität und Seriosität verständliche, relevante (Ausnahme: zu Beginn der Befragung) und nicht zu viele Fragen indirekte Fragen sichere / verlässliche Kommunikationswege wählen*

## 5. Fehler durch Falschantwort

### *Entstehung:*

*Es werden bewusst oder unbewusst falsche  
Angaben gemacht*

### *Vermeidung:*

*Anonymität*

*soziale Erwünschtheit eliminieren / abschwächen  
(schriftliche Befragung, indirekte Fragen,  
randomisierte Antwort)*

## 6. Messfehler

### *Entstehung:*

*Falsche Handhabung des Instrumentes  
Zufällige, unkontrollierte Einflüsse*

### *Vermeidung:*

*Vermitteln der exakten Handhabung  
Kontrollierte Bedingungen (geringe externe Einflüsse)*

# 7. Aufbereitungsfehler

## *Entstehung:*

*Bei der Aufbereitung der Messergebnisse für die Analyse*

## *Vermeidung:*

*Geschultes Personal einsetzen  
Konzentriertes Vorgehen*

## 8. Interpretationsdiskrepanz

### *Problem:*

*Von empirischen Resultaten muss auf die komplexe Realität zurück geschlossen werden*

### *Lösungsansätze:*

*Kommunikation der Ergebnisse an den Substanzwissenschaftler*

*Als Statistiker: Vermeiden von eigener Inhaltlicher Interpretation und Vermeiden von externer methodischer Fehlinterpretation*

# Fehlende Werte

Zufällig Fehlend

Abhängig Fehlend

Nicht Zufällig Fehlend

# Zufällig Fehlend

## Missing Completely At Random

*Der Antwortausfall ist rein zufällig Natur und steht weder mit der Ausprägung des Merkmals, noch mit einer anderen Variablen in Verbindung*

# Abhängig Fehlend Missing At Random

*Der Antwortausfall ist unabhängig  
von der Ausprägung des Merkmals,  
steht allerdings mit einer anderen Variablen in Verbindung*

# Nicht Zufällig Fehlend Missing Not At Random

*Der Antwortausfall ist abhängig  
von der Ausprägung des Merkmals, steht aber nicht (primär)  
mit einer anderen Variablen in Verbindung*

# Bestimmung der Arten

## Abhängig Fehlend

*Aufsplitten der vollständigen und unvollständigen Daten*

*Bestimmung der Verteilung*

*Differenzbildung zwischen den Verteilungen*

*Bei großen Differenzen scheint die Variable einen Einfluss auf das Fehlen zu haben*

# Bestimmung der Arten

## Zufällig Fehlend

*nicht Abhängig Fehlend und keine auffallend hohe Fehlende Werte Rate (im Vergleich zu anderen Variablen)*

# Bestimmung der Arten

## Nicht Zufällig Fehlend

*nicht Abhängig Fehlend und auffallend  
hohe Fehlende Werte Rate (im Vergleich zu  
anderen Variablen)*

# Complete & Available Case Analysis

## Complete Case Analysis

*Es werden nur die Einheiten untersucht, für die zu jeder Variablen die Ausprägung vorhanden ist*

## Available Case Analysis

*Es werden nur die Einheiten untersucht, für die zu jeder aktuell interessierenden Variable die Ausprägung vorhanden ist*

# Imputation

## Einfache Imputation

*Fehlende Werte sollen durch die „richtigen“  
Werte ersetzt werden  
Suche nach adäquater Variable (Matching)  
Einsetzen eines plausiblen Wertes*

## Multiple Imputation

*mehrmaliges Einsetzen eines Wertes  
(Wertevektor) und anschließende Analyse*

# Bewertung der Möglichkeiten

## Complete Case Analysis

*Einfach handhabbar*

*Vernachlässigt den Fehlendmechanismus*

*Bei vielen Variablen wird die Fallzahl gering*

## Available Case Analysis

*Nutzt mehr Information*

*Vernachlässigt den Fehlendmechanismus*

*Vergleich von Teilergebnissen wird erschwert*

# Bewertung der Möglichkeiten

## Einfache Imputation

*Informationserhalt*

*Unterschätzung der Unsicherheit*

## Multiple Imputation

*Belastbarer Informationserhalt*

*Technisch aufwendiger*

# Zusammenfassung

Allgemein

Im Studienverlauf

In einer Befragung

bei Fehlenden Werten

*Wissen um die Fehler*

*Zusammenarbeit mit*

*Substanzwissenschaftlern*

*Zusichern und Gewährleisten von*

*Anonymität*

*Zufällig Fehlend ist unproblematisch*

*Abhängig Fehlend in Bezug auf die*

*verursachende Variable kritisch*

*Nicht Zufällig Fehlend großes*

*Fehlerpotential in der Analyse*

Imputation

*Solides Verfahren allerdings kein echter*

*Informationsgewinn*

# Quellen

## Literaturangaben

*Fahrmeir, L. et al. (2007): Statistik*

*Göthlich, S. (2005): Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen*

*Groves, R. (1989): Survey Errors and Survey Costs*

*Küchenhoff, H. und Kauermann, G. (2008): Erkenntnisse aus Stichproben*

*Radermacher, W. und Körner, T. (2006): Fehlende oder fehlerhafte Daten in der amtlichen*

*Statistik. Neue Herausforderungen und Lösungsansätze*

*Schnell, R. (1991): Wer ist das Volk?*

# Quellen

## Internetquellen

[http://www.welt.de/diewelt/article1456180/Experte\\_sieht\\_Elterngeld\\_als\\_Ursache\\_fuer\\_steigende\\_Geburtenrate.html](http://www.welt.de/diewelt/article1456180/Experte_sieht_Elterngeld_als_Ursache_fuer_steigende_Geburtenrate.html)

<http://www.zeit.de/online/2008/34/geburten-anstieg-analyse>

<http://www.statistik-berlin.de/wir/amtliche-statistik/wir1.htm>

<http://www.nzzfolio.ch/www/21b625ad-36bc-48ea-b615-1c30cd0b472d/showarticle/fffc70d1-99f5-4326-912f-dfc7f23cbc48.aspx>

<http://www.jstor.org/pss/2283137>

<http://www.stat.psu.edu/~jls/mifaq.html>

Vielen Dank für Ihre Aufmerksamkeit!