

Otto-Friedrich-Universität Bamberg  
Lehrstuhl für Soziologie, insbes. Survey-Methodologie

Fortgeschrittene Methoden der Datenerhebung:  
Datenerhebung und Fehlerquellen  
Sommersemester 2020



Gibt es verdächtige Interviewer?

Eine Analyse zur Identifikation potenziell gefälschter  
Interviews im EES 2002

Dozent: Herr Prof. Dr. Mark Trappmann  
Prüfling: Kevin Glock  
E-Mail: kevin.glock@stud.uni-bamberg.de  
Semster: 2. FS, MiSS  
Abgabe: 30. März 2021

Zeichenanzahl (ohne Abbildungen und Tabellen): 39.728

Wörterzahl: 5.131



<b>1. EINLEITUNG.....</b>	<b>5</b>
<b>2. FORSCHUNGSSTAND .....</b>	<b>7</b>
2.1. FORMEN VON FÄLSCHUNGEN.....	7
2.2. VERZERRUNGEN DURCH FÄLSCHUNGEN.....	8
2.3. PSYCHOLOGISCHE SICHTWEISE AUF INTERVIEWFÄLSCHUNG .....	9
2.4. ERKENNUNG POTENZIELLER FABRIKATE .....	12
2.5. FORSCHUNGSFRAGE.....	14
<b>3. DATEN &amp; METHODIK .....</b>	<b>17</b>
3.1. RELIABILITÄT UND VARIABILITÄT.....	19
<b>4. ERGEBNISSE .....</b>	<b>23</b>
<b>5. ZUSAMMENFASSUNG .....</b>	<b>27</b>
<b>6. LITERATURVERZEICHNIS.....</b>	<b>30</b>
<b>7. ANHANG.....</b>	<b>34</b>
7.1 R CODE .....	34
7.1.1. <i>Required libraries</i> .....	34
7.1.2. <i>Load respondent's data</i> .....	34
7.1.3. <i>Merge data</i> .....	35
7.1.4. <i>Methodological info on dataset</i> .....	35
7.1.5. <i>Methodological info on subset</i> .....	37
7.1.6. <i>Variability approach</i> .....	37
7.1.7. <i>Calculate the means</i> .....	37
7.1.8. <i>Calculate variabilities</i> .....	38
7.1.9. <i>Get the test statistic T</i> .....	38
7.1.10. <i>Whole dataset</i> .....	39
7.1.11. <i>Resampling</i> .....	41
7.1.12. <i>Resample for single interviews</i> .....	42
7.1.13. <i>Qualitative evaluation</i> .....	44
7.2 EIGENSTÄNDIGKEITSERKLÄRUNG .....	45

## Tabellenverzeichnis

TABELLE 1 ANZAHL DER KONTAKTVERSUCHE DES ERSTKONTAKTES (OHNE 10 UND 99=SYS MIS), QUELLE ESS 2002 ROUND 1 HUNGARY .....	17
TABELLE 2 ANZAHL DER INTERVIEWER PRO INTERVIEWS BASIEREND AUF DEN ERSTELLTEN SUBSET VON KONTAKTVERSUCHEN, QUELLE: EIGENE BERECHNUNGEN .....	18
TABELLE 3 GLOBALES INTERVIEWER-RANKING ERSTE 28 PERSONEN NACH SCHÄTZER $T_i$ UND PLAUSIBILITÄT FÜR POTENZIELLE FÄLSCHUNG, QUELLE: EIGENE BERECHNUNGEN .....	24
TABELLE 4 RELATIVE ANZAHL DER POTENZIELL GEFÄLSCHTEN INTERVIEWS DER IN TABELLE 3 MARKIERTEN INTERVIEWER, QUELLE: EIGENE BERECHNUNGEN .....	25

## Abbildungsverzeichnis

ABBILDUNG 1 THEORETISCHE KOSTEN-NUTZEN-FUNKTIONEN VON BEFRAGENDEN UND AUFTRAGGEBERN .....	11
ABBILDUNG 2 POSTDIKTIVER ERKLÄRUNGSWEG .....	13
ABBILDUNG 3 THEORETISCHE ITEM-DIMENSIONALITÄT .....	19
ABBILDUNG 4 EMPIRISCHE (SCHWARZ) UND SIMULIERTE (GRAU) DICHTEFUNKTION DER TOTALEN VARIABILITÄT DER NACH INTERVIEWER GECLUSTERTEN INTERVIEWS, QUELLE: EIGENE BERECHNUNGEN .....	23

## Formelverzeichnis

(1) BIAS EINES SCHÄTZERS $\hat{\vartheta}$ .....	8
(2) SUMME DER ABWEICHUNGSQUADRATE .....	8
(3) SUMME DER ABWEICHUNGSQUADRATE .....	20
(4) BIAS DER VARIABILITÄT .....	20
(5) TESTSTATISTIK $T_{  }$ .....	21

## **Abstract**

Falsification is a central problem of empirical social research and can lead to significant distortions of parameter estimators. Prevention and detection of falsifications are therefore essential to ensure reliable data and valid estimates by interviews. This paper evaluates the procedure of identifying potential counterfeits using the variability method, applied on data of the European Social Survey Hungary 2002 (Round 1). 10 interviewers could thus be identified who require follow-up investigation on 35 interviews. The evaluated hypotheses find partial support. Potential falsification occurs more frequently with some interviewers than with others. The variability method can be implemented in a cost and time efficient manner. According to the literature it provides equally reliable results compared to advanced statistical methods to known falsifiers, but reliability is limited in case of unknown ones. To ensure reproducibility, the analysis is performed in R. The analysis code can be found in the Annex.

Keywords: Fake interviews, Interview fabrication, variability, R, resample, survey research

## **Zusammenfassung**

Interviewfälschungen sind ein zentrales Problem der empirischen Sozialforschung und können zu bedeutenden Verzerrungen von Parameterschätzern führen. Vermeidung und Erkennung von Fälschungen sind essenziell, um reliable Daten durch Befragungen und valide Schätzungen zu gewährleisten. Die Arbeit evaluiert das Erkennung von potenziellen Fälschungen mittels Variabilitätsmethode am Beispiel des European Social Surveys Ungarn 2002 (Round 1). 10 Interviewer konnten detektiert werden, die für 35 Interviews einer Nachuntersuchung bedürfen. Die untersuchten Hypothesen werden teilweise bekräftigt. Potenzielle Fälschungen häufen sich bei einzelnen Interviewern. Die Variabilitätsmethode ist kosten- und zeiteffizient implementierbar. Laut Literatur liefert sie gegenüber erweiterten statistischen Verfahren gleichwertige Ergebnisse zur Erkennung bekannter Fälscher, für unbekannte ist dies einzuschränken. Um die Reproduzierbarkeit zu gewährleisten erfolgt die Analyse in R. Der Analysecode findet sich im Anhang.

Keywords: Interviewfälschung, Interviewfabrikation, Variabilität, R, Resampling, Surveyforschung



## 1. Einleitung

Daten zu generieren, um valide Aussagen zu erhalten, ist in großangelegten Bevölkerungsbefragungen ein komplexes Problem. Viele Fragestellungen sind nur mittels der systematischen Methode der Befragung beantwortbar, trotz dass Befragungen allgemein als unzuverlässig gelten. Eine bedeutende Problematik stellen zudem steigende Unit-Non-Response-Rate dar (Schnell 2013), die dazu führt, dass größere Auswahlgesamtheiten definiert werden müssen, um hinreichende Ausschöpfungsquoten zu generieren. Diese bedingen eine zunehmend belastendere Feldarbeit für Interviewende durch Nichterreichbarkeit und mehr Arbeitsaufwand. Vorsätzliche Interviewfälschungen sind ein probates Mittel für Interviewende, um den Lohn, bei Minimierung des Arbeitsaufwandes, zu maximieren.

Systematische Verzerrungen durch Fälschungen stellen ein schwerwiegendes Problem für die Konstruktion erwartungstreuer und konsistenter Schätzer dar. Obwohl die Vermeidung und Detektion von Fälschungen zentrale Probleme in der empirischen Sozialforschung darstellen (Blasius und Thiessen 2012, 2013; Blasius und Friedrichs 2012), liegen nur mäßig viele Publikationen in Bezug auf Interviewfälschungen vor (Blasius 2014, S. 323).

Die vorliegende Arbeit verfolgt daher folgende Fragen: *Welche Faktoren sind ursächlich zur Erklärung einer Fälschung? Weisen einzelne Interviews signifikant abweichendes Verhalten auf? Sind potenzielle Interviewfälschungen adäquat anhand ihrer abweichenden Streuung zu detektieren?*

Dies erfolgt anhand des European Social Survey Ungarn 2002 (European Social Survey ERIC (ESS ERIC) 2003). Potenzielle Interviewfälschungen werden mittels Resampling der empirischen Streuung in jeweiligen Interview-Clustern datengetrieben identifiziert und deren Plausibilität als Fälschung geprüft.

Im Folgenden wird der Stand der Forschung evaluiert. Es erfolgt die Darstellung der Daten, Methodik und der Ergebnisse. Am Ende der Arbeit erfolgt eine ausführliche Diskussion. Da potenzielle Fälschungen betrachtet werden bleibt eine endgültige Entscheidung einer Hypothese bis zur Nachuntersuchung ausstehend.





## 2. Forschungsstand

Im Folgenden wird auf Forschungsergebnisse zu Interviewfabrikation hinsichtlich der Folgen für Schätzverfahren, psychologischer Beweggründe und der methodischen Aufdeckung potenzieller Fabrikate eingegangen.

### 2.1. Formen von Fälschungen

Unterschieden werden Fälschungen nach dem Grad der Fälschung. Einerseits werden Duplikate in Datensätzen als Fälschungen angesehen (Blasius und Thiessen 2012, 2015; Kuriakose und Robbins 2016), andererseits aber auch die falsche Umsetzung einer Interviewanweisung oder systematisch fabrizierte Interviews (Blasius 2014, S. 325), die nie eine zu befragende Person erreicht haben. Darüber hinaus gibt es speziellere Formen von Teilfälschungen, wie Nutzung von abweichenden Erhebungsmethoden oder das Eingreifen der befragenden Person, die stellvertretend antwortet oder aus eigenem Antrieb einen passenden Wert für ein Item angibt (Schräpler und Wagner 2005, S. 7).

Blasius (2014) führt aus, dass laut „AAPOR (2003: 2) [...] ein Interview bereits gefälscht [ist], wenn eine der folgenden Aussagen zutrifft: >>a) fabricating all or part of an interview – the recording of data that are not provided by a designated survey respondent and reporting them as answers of that respondent; b) deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a refusal case as ineligible for the sample; reporting a fictitious contact attempt); c) deliberately miscoding the answer to a question in order to avoid follow-up questions; d) deliberately interviewing a non-sampled person in order to reduce effort required to complete an interview; or e) otherwise, intentionally misrepresenting the data collection process to the survey management.<<“ (S. 325).

Schnell (1991) betrachtet Fälschungen von Interviewern als eine spezielle Form der Imputation, die auf den Annahmen der fälschenden Person beruht, wie einzelne Charakteristiken von Befragten streuen. Sie entsprechen einer Expertenschätzung (Schnell 1991, S. 26).

In der vorliegenden Arbeit werden Fabrikate als bewusste Fälschungen durch Imputation von Werten durch die befragende Person im Sinne von Schnell (1991) und Schräpler und Wagner (2005) verstanden. Dabei werden Teilfälschungen und auch vollständig fabrizierte Interviews als gleichwertig schwerwiegend verstanden.

## 2.2. Verzerrungen durch Fälschungen

Fälschungen von Interviews können zu bedeutenden Verzerrungen von Parameterschätzern führen und machen Daten im schlimmsten Fall unbrauchbar, weshalb ihre Erkennung und Aussortierung notwendig sind. Im Folgenden werden diese Verzerrungen erläutert.

Der Bias eines Schätzers  $\hat{\vartheta}$  ist die systematische Abweichung des Schätzers  $\hat{\vartheta}$  vom wahren Wert  $\vartheta$  in der Grundgesamtheit.

$$Bias(\hat{\vartheta}) = \hat{\vartheta} - \vartheta$$

(1)

Es ist bedeutend für die konsistente und erwartungstreue Schätzung mittels einer Stichprobe. Die systematische Imputation von Werten durch Befragende oder Befragte verursacht einen Bias und sollte minimal sein. Durch die Testkonstruktion, also Nutzung von Filterfragen, variablen Fragekonstruktionen und Sortierungen der Antwortkategorien kann ein Survey so konstruiert werden, dass der Test nicht ermüdend wirkt, Extremwerttendenz, Tendenz zur Mitte und anderen Testeffekte minimiert werden. Ziel der Testkonstruktion sollte daher sein, mögliche Varianzveränderungen adäquat herbeizuführen und damit eine hohe Variabilität des Antwortverhaltens von Befragten abdecken zu können.

Schnell (1991) betrachtet Fälschungen von Interviewern als eine spezielle Form der Imputation, die auf den Annahmen dieser beruht, wie einzelne Charakteristiken von Befragten streuen. Diese Imputation führt in der Folge zu einem Bias des Schätzers  $\hat{\vartheta}$  für einzelne Items, Gruppen von Items oder aber ganze Interviews, je nach Anteil der Fälschungen.

Schlussfolgernd daraus lässt sich damit herleiten, dass die Variabilität aller Stichprobenstatistiken  $Y_i$  mit  $i=1, \dots, n$ , als Summe der Abweichungsquadrate

$$SQ(Y_i) = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

(2)

hinreichend groß sein sollte, das heißt alle Kategorien durch Befragte abgedeckt werden und damit eine Tendenz zur Mitte nicht vorliegt. Extremwerttendenzen hingegen würden zu einer Maximierung von  $SQ(Y_i)$  führen, da sie in allen Fällen maximal vom

Mittelwert abweichen. Durch analytische Abschätzung schlussfolgert Schnell (1991), dass einfache Statistiken bei größeren Fallzahlen und kleinen Anteilen von Fälschungen robuste Werte abbilden (S. 35). Da Fälschungen deutliche Verzerrungen multivariater Analysen herbeiführen können (Menold et al. 2013; Schräpler und Wagner 2005; Schnell 1991), sind Fabrikate schwerwiegend. Ergebnisse solcher Interviews sind im ärgsten Fall unbrauchbar. Schnell (1991) erläutert, dass Imputationen durch Fälscher nicht problematisch für die Abschätzung der Verzerrung wären, sofern zufällig generierte Werte eingesetzt würden (S. 29). Das Problem ist die potenzielle Anwendung von Laientheorien über die Streuung der Ausprägungen, die zu einem systematischen Fehler führen (Schnell 1991, S. 29). Es ist wahrscheinlich, dass Interviewfälschungen ähnliche Antwortmuster aufweisen, innerhalb als auch zwischen Interviews einer befragenden Person. Item-Batterien werden ähnlich ausgefüllt, demografische Merkmale wiederholen sich unplausibel und Filterfragen werden übergangen. „[D]ie Konsistenz der Angaben in den gefälschten Interviews ist etwas größer als in echten Interviews“ (Schnell 1991, S. 28). Auch (Schäfer et al. 2004a, S. 11) stellen fest, dass Fälschungen geringere Variabilität innerhalb eines Interviewer-Clusters aufweisen. Dies wäre besonders ein Problem in Untersuchungen mit kleinen Stichprobengrößen. Da für größere Stichproben das schwache Gesetz der großen Zahl greifen würde, sofern alle  $X_i$  Ziehungen unabhängig sind und wenige Fälschungen vorhanden sind. Fraglich bleibt, welcher Wert eine hinreichende Größe für die Variabilität eines Interviews darstellt.

### 2.3. Psychologische Sichtweise auf Interviewfälschung

Interviewende produzieren aus unterschiedlichen Gründen Fälschungen im Rahmen ihrer Tätigkeit (Blasius 2014; Schäfer et al. 2004b). Ein Grund für Fälschungen sei die Demoralisation von Befragenden (Crespi 1945). Einerseits durch die Institute, andererseits durch die Befragenden (Blasius 2014). „So können eine zu lange Dauer des Interviews, zu viele Nachfragen [...], zu komplexe und zu schwierige Fragen den Interviewer demoralisieren. Ein weiterer möglicher Grund ist der Druck durch den Supervisor [...] eine bestimmte Anzahl von Interviews durchzuführen“ (Nelson und Kiecker 1996, S. 1110; Blasius 2014, S. 324).

Andererseits stimuliert riskantes Verhalten Individuen mit einem psychologischen Reiz der Belohnung, was als möglicher Grund angeführt wird, Interviewfälschung zu betreiben (Harrison und Krauss 2002). Insbesondere, wenn Teilfälschungen von Individuen betrieben werden, die ein hohes Wissen über Populationsparameter (bspw.

Studenten) und einen hohen Grad der Motivation aufweisen riskantes Verhalten ausprägen, ist das Risiko sehr hoch, dass diese Art der Fälschung betrieben wird. Dabei sollten weitere motivationale Faktoren, wie Persönlichkeitsmerkmale und Einstellungen (Buehl und Melchers 2017, S. 1) in Betracht gezogen werden, um fälschendes Verhalten zu erklären. Das Risiko entdeckt zu werden und die Konsequenzen sind tendenziell gering (Menold und Kemper 2014; Blasius 2014), was einen zusätzlichen Anreiz für fälschende Personen darstellt. Zwar kann dieses Phänomen auftreten, allerdings sieht der Autor dies als keine allgemeine ursächliche Erklärung für Fälschungen an. Diese Art des divergenten Verhaltens wird eher als Randphänomen betrachtet, da es kategorisch Interviewende als pathologische Fälscher kategorisiert und damit unplausibel erscheint. Sofern dies so wäre, wäre jede je durchgeführte Befragung unbrauchbar.

Einheitlich wird in der Literatur hingegen auf rationale Beweggründe verwiesen (Blasius 2014; Schäfer et al. 2004b). Der Erhebungsprozess bedeutet für Interviewende und für Erhebungsinstitute einen erheblichen Zeit- und Kostenaufwand. Einerseits nimmt die Bereitschaft an Befragungen teilzunehmen seit Jahren zunehmend ab (Schnell 2013). Eine zunehmend steigende Unit-Non-Response-Rate, sorgt in Folge für steigenden Druck der Interviewenden, im ursprünglich ersten Versuch ein Interview durchführen zu können. Andererseits führt dies durchschnittlich auch zu einer höheren Anzahl von Versuchen eine beliebige Kontaktperson in einem Wohngebiet zu erreichen und eine hinreichende Anzahl an Interviews in der Stunde durchführen zu können, um eine angemessene Bezahlung zu erreichen; vornehmlich wenn sie pro durchgeführten Interview und nicht pro Stunde entlohnt werden (Blasius und Friedrichs 2012) oder Re-Issuings nur einen geringeren Lohn bedeuten. Das Argument erscheint plausibel. Interviews führen, ist eher eine befristete Tätigkeit als ein Hauptberuf und tendenziell unterbezahlt, da großangelegte Surveys personaltechnisch teuer in der Umsetzung sind. "Einfache" Tätigkeiten durch befristete Beschäftigte durchzuführen ist daher ein ökonomisch probates Mittel; Personalfluktuationsrate ist hoch und eine Dauerbeschäftigung kalkulatorisch und personaltechnisch schlecht planbar.

Die rationalistische Vereinfachung bedeutet also, dass sich die individuellen Kosten-Nutzen-Funktionen von Interviewenden und Auftraggeber entgegenstehen (siehe Abbildung 1). Je mehr Re-Issuings erfolgen, desto größer wird dieser Trade-Off für die Interviewenden, was im Falle des ESS damit kompensiert wird, dass Re-Issuings, durch andere Interviewende durchgeführt werden. Betrachtet man Interviewende im

Szenario des Erhebungsprozesses als rationale Akteure, so lässt sich der Nutzen (Entlohnung per Interview) individuell erhöhen, wenn auf Ehrlichkeit und Arbeitsaufwand verzichtet wird und Interviews systematisch gefälscht werden, um so den individuell größten Nutzen bei geringsten Kosten (Zeitaufwand, körperlicher Einsatz durch Erreichen von Wohngebiet und Wohneinheiten, psychischer Aufwand durch potenziell harsche Abweisung durch Interview-Verweigerer) zu erreichen. Insbesondere unter Annahme der Informationsasymmetrie von Prinzipal (Erhebungsinstitute) und Agent (Interviewende).

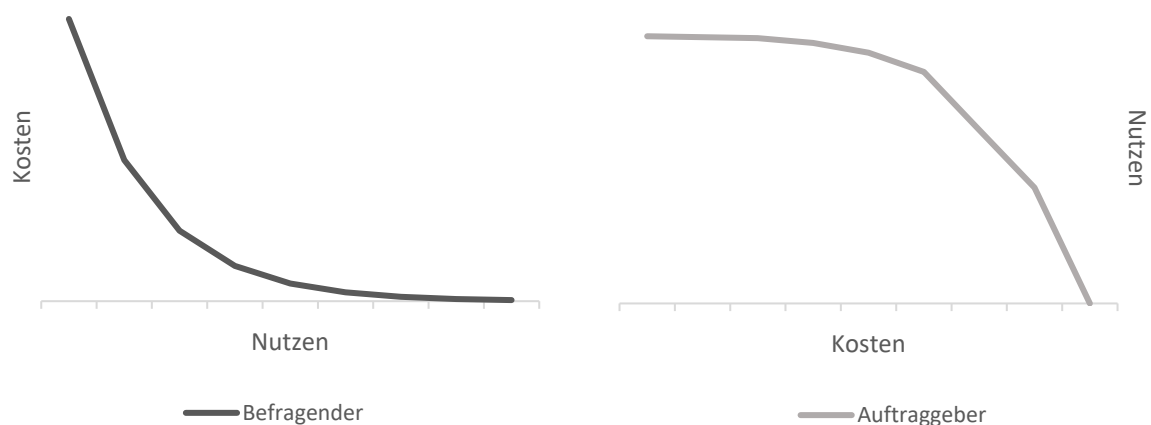


Abbildung 1 Theoretische Kosten-Nutzen-Funktionen von Befragenden und Auftraggebern

Betrachtet man Interviewfabrikation aus persönlichkeitspsychologischer Sicht, kann diese als Lügen oder Unehrlichkeit (Menold und Kemper 2014) verstanden werden. Je nach Reichweite des Fabrikationsbegriffs beabsichtigen Interviewer durch bewusste oder unbewusste Handlungen sich einen persönlichen Vorteil zu ermöglichen, dessen Erlangung ohne Lüge nur mit höherem Aufwand erreichbar wäre, beziehungsweise dessen Erfolgswahrscheinlichkeit subjektiv geringer ausfällt. Dies deckt sich mit den Annahmen der rationalistischen Sichtweise.

Fabrikation ist folglich eine komplexe Lügenkonstruktionen, weshalb es nötig wird diagnostische Denkweisen auf Daten anzuwenden. Schräpler und Wagner (2005) und Blasius (2014) führen an, dass „Interviewer at risk“ (Bushery et al. 1999) häufig den „shortest path through the interview“ (Hood und Bushery 1997, S. 821) nehmen, „[they] avoid choosing the category >>others, please specify<<“ (Menold und Kemper 2014, S. 45) und „producing survey participants who live in one person families and are >>white, non-smoker, [without] health problems and no health insurance<<“ (Hood und Bushery 1997, S. 821; Menold und Kemper 2014, S. 45); unauffällige

Durchschnittspersonen mit geringer Variation. Diese Definition deckt sich mit der theoretischen Annahme, dass Fälschungen eine geringere Variabilität aufweisen (Schäfer et al. 2004a, S. 11).

Menold und Kemper (2014) führen aus, dass sich Unterschiede in Verhaltens- und Einstellungsmerkmalen der Antwortenden zeigen. Verhaltensmerkmale seien sehr leicht durch Fälscher reproduzierbar. Hingegen seien Einstellungsmerkmale, das heißt persönliche Urteile und individuelles Verhaltenswissen schwieriger zu fälschen. Ferner seien individuell psychologische Aspekte sowie subjektive und kollektive Sichtweisen besonders schwer zu fälschen (Menold und Kemper 2014), was mögliche Vorteile zur Vermeidung durch Fragekonstruktion bietet. Aus den Erläuterungen ergeben sich zwei Dimensionen der Intelligenz, die Fälscher vereinen müssen. Einerseits müssen sie ein breites Abstraktionswissen über individuelle und Populationsparameter und -anteile besitzen, insbesondere in Subgruppen und andererseits Vorhersagefähigkeit mitbringen, da sie abschätzen müssen, welche Antwortmuster zu welchen Ergebnissen führen. Diese Überlegung wird dadurch gestützt, dass gefälschte Interviews nach Professionalität des Interviewers variieren. Langjährige Interviewer modifizieren eher und produzieren weniger Totalfälschungen, unerfahrene Interviewer hingegen produzieren tendenziell schlechtere Teil- und mehr Totalfälschungen (Schnell 1991, S. 28). Dies deckt sich mit der rationalen Annahme der Informationsasymmetrie, dass Fälscher eine rational Handlungsoption hinsichtlich ihrer Situation und Verdienstsituation nach ihrem potenziellen Wissen über das Verhalten des Prinzipals abwägen. Personen mit geringerer Arbeitserfahrungen fälschen tendenziell mehr, da diese von höherer Informationsasymmetrie ausgehen. Sie nehmen an, das Institut gehe nicht von Fälschern aus und oder nicht die Mittel oder Kenntnis darüber hat, eine Fälschung zu finden. Hingegen Personen mit höherer Arbeitserfahrung Informationsasymmetrien besser abschätzen und unauffälliger fälschen, so dass sie unentdeckt bleiben.

### 2.4. Erkennung potenzieller Fabrikate

Im Folgenden wird dargestellt, was die statistische Erkennung potenzieller Fabrikate bedeutet und mit welchen Methoden sie erfolgen kann. Dabei wird sich auf bereits vorhandene Fabrikate in Datensätzen beschränkt, die im Nachhinein (ex-post) kausal erklärt und aufgedeckt werden sollen (Postdiktives Verfahren). Damit erfolgen die Analysen im Sinne von Pearl und Mackenzie (2018) kontrafaktisch und basieren auf der Argumentation, wie müsste ein Outcome ausgeprägt sein, wenn es keine Fälschung

gegeben hätte und schlussfolgern daraus was bedeutet eine Fälschung, für ein Outcome.

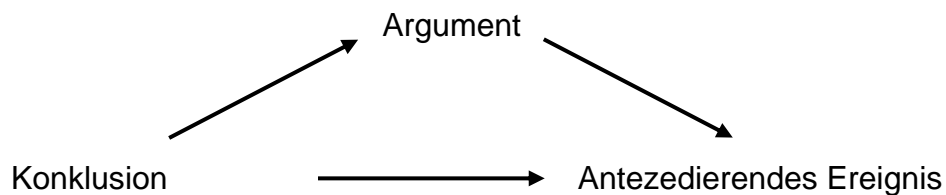


Abbildung 2 Postdiktiver Erklärungsweg

Für die Vermeidung von Interviewfälschungen (Präventive Verfahren) sei allgemein auf Gwartney (2013) und Schnell (2019) verwiesen.

Die Forschung zur Detektion von Fabrikaten nutzt das gesamte methodische Repertoire der Statistik sowie datengetriebene und algorithmische Ansätze (Bredl et al. 2012; Murphy et al. 2004; Kemper und Menold 2014; Haas und Winker 2016; Birnbaum 2012; Kuchen 2018; Schäfer et al. 2004b; Bredl et al. 2011; Bergmann et al. 2019). Einfache deskriptive Verfahren, die sich auf Korrelationswerte und Varianzen fokussieren, können Tendenzen abbilden, ob einzelne oder mehrere Interviews auffällig vom Gesamt abweichen. Diese Verfahren beruhen auf der Annahme der internen Konsistenz von Item-Skalen. Einzelne Items einer Skala sollten eine gewisse Korrelation aufweisen, die bei Fälschungen abweicht. Es ist anzunehmen, dass höhere Korrelationen für Item-Batterien vorliegen, da Fälscher ein repetitives und damit konsistenteres Antwortverhalten zeigen, wenn sie auf eine schnelle Beantwortung mittels kürzesten Weg durch das Interview setzen.

Datengetriebene und algorithmische Methoden haben das größte Potenzial für die zukünftige Evaluation von Surveydaten, um Fälschungen zu detektieren. Bieten im vorliegenden Fall jedoch den Nachteil, dass sie Trainingsdaten benötigen, das heißt vorhandene Fälschungen bekannt sein müssen, um auf weitere Fälschungen zu schließen.

Interpretiert man das Aufdecken von Fabrikaten mit dem klassischen Kommunikationsmodell (Röhner und Schütz 2015, S. 21–25) ist die Erkennung von Fabrikate der Zugriff auf Informationen einer Nachricht als Signal ohne spezifische Zusammenarbeit mit dem Absender. Um potenzielle Fabrikate zu erkennen ist es nötig das Vorhandensein von Rauschen in Kommunikationssignalen zu detektieren, um Verzerrungen zu extrahieren. Auffällig ist, dass gefälschte Interviews einen höheren Anteil an Middle Responses und einen geringeren Anteil an Extremwertaussagen aufweisen (Bredl et

al. 2012; Schäfer et al. 2004b). Zudem zeigen gefälschte Interviews deutlich häufiger geringere Anteile an „anderes“-Kategorien, da fälschende Personen den kürzesten Weg durch ein Interview nehmen (Menold et al. 2013, S. 30). Daraus ergibt sich, dass Varianzen in gefälschten Daten deutlich geringer ausfallen. Das gilt auch für andere gefälschte experimentelle Daten (Al-Marzouki et al. 2005). Die fälschende Person geht dabei davon aus, dass insbesondere quasi-experimentelle Daten, wie Bevölkerungsumfragen in ihren Response-Mustern nicht so leicht reproduzierbar wären und damit eher als abweichende Einzelfälle betrachtet würden, denn als vorsätzlich produzierte Daten. Fälschungen signalisieren also tendenziell ein überangepasstes Antwortverhalten, da sie versuchen unauffällig zu bleiben. Das Raschen wird hier also durch den Sender selbst verursacht. Die Erkennung von Rauschen in verdächtigem Material, basiert aber nur auf der Wahrscheinlichkeit, mit der ein möglicherweise verborgenes Signal enthalten ist, da statistische Verfahren nicht vollständig klären können, ob es sich um eine Fälschung handelt. Dies ist ein Nachteil der Aufdeckung. Weshalb immer Nachuntersuchungen der Interviews wie Rekontaktierungen (Schnell 2019, Kap 8.7.1) und die Erhebung von Paradata (Felder et al. 2014) erfolgen sollte, um diesen Nachteil zu minimieren. Dieses Problem besteht auch mit der genutzten Variabilitätsmethode, die potenzielle Fälscher aufgrund zu geringer geclustelter Totalvariabilität filtert, die einer Nachbefragung zugeführt werden sollten.

Um vorhandene Daten zu reproduzieren und auf eine Population hochzurechnen, macht sich das Resampling von Daten nützlich, das  $n$  Stichproben mit zurücklegen aus der bereits vorhandenen Stichprobe zieht. Durch die Simulation der Stichprobe ergeben sich  $N$  Stichproben, die eine Quasi-Grundgesamtheit abbilden, die konsistente Schätzungen für die wahre Grundgesamtheit erlauben.

### 2.5. Forschungsfrage

Die Erkennung von Fabrikaten stellt den Zugriff auf Informationen ohne spezifische Zusammenarbeit mit dem Absender dar. Fälscher signalisieren ein kostenminimierendes Verhalten. Ihre Kommunikation weicht vom üblichen Antwortverhalten ab und verursacht Rauschen, das auf eine Fälschung deutet. Um zu erkennen welches Interview eine Fälschung ist, müssen enthaltene Signale gefiltert werden, die mittels einer Kenngröße von echten Interviews als abweichend entschieden werden können. Rauschen in Daten, ist ein repetitives und konsistentes Antwortverhalten. Fälscher vermeiden Filterfragen, da sie dem Nutzenkalkül widersprechen. Extremwerte und



Ausfallkategorien werden vermieden, Middle-Response bevorzugt. Für Item-Batterien ergibt sich erhöhte Konsistenz. *Folglich weisen Fälschungen durchschnittlich eine geringere Gesamtstreuung der Antwortmuster auf.*

Aus dem Nutzenkalkül unter Informationsasymmetrie ergibt sich, dass Interviewer, die von einer stärkeren Asymmetrie zu ihrem Vorteil ausgehen, diese ausnutzen, um eine höhere Entlohnung zu erhalten. Interviewenden können unterschiedliche Wahrscheinlichkeiten zugeordnet werden, zu fälschen. *Fälschungen häufen sich bei einzelnen Interviewern, besonders mit geringer Arbeitserfahrung.*

Da keine Daten mit bekannten Fälschungen vorliegen, fokussiert die Arbeit die Detektion potenziell gefälschter Interviews anhand deskriptiver Werte, die mittels Simulation geschätzt werden.



### 3. Daten & Methodik

Im Folgenden werden die verwendeten Daten und Methodik vorgestellt. Die Analyse erfolgte mittels der Programmiersprache R und Datenpakete des *tidyverse*. Der genutzte Algorithmus wird in Form von Code im Anhang aufgeführt. Um eine volle Reproduzierbarkeit zu gewährleisten, ist der R Code unter <https://kevinglock.github.io/Int-Fabric/index.html> mit CC BY-SA Lizenz zugänglich.

In der vorliegenden Arbeit wird der European Social Survey 2002 Ungarn (ESS Round 1 Hungary) genutzt. Der European Social Survey (ESS Round 1 Hungary) enthält Daten von Befragten der ungarischen Wohnbevölkerung. Erhoben wurden die Daten im Wohnumfeld der Befragten mittels *computer-assisted personal interviews* (CAPI). Insgesamt gibt es 2484 Zielpersonen. 1685 verwertbare Interviews wurden realisiert. Das entspricht einer Unit-Non-Responserate von 32,17 Prozent. Und einer Ausschöpfung von 67,83%.

Für den Erstkontakt waren 198 Personen eingeteilt, 2461 Interviews sollten als Erstkontakt gestartet werden, 23 sind als NA vermerkt. 689 Befragte wurden nicht durch den Erstkontakt aufgesucht. 436 Befragte wurden einmal durch den Erstkontakt aufgesucht. 224 Befragte wurden zweimal aufgesucht. 124 Personen wurden dreimal durch den Erstkontakt angesteuert.

Kontaktversuche durch Erstkontakt	Anzahl der Befragten	Kontaktversuche durch Erstkontakt	Anzahl der Befragten
0	689	5	16
1	436	6	8
2	224	7	2
3	124	8	1
4	45	9	2

Tabelle 1 Anzahl der Kontaktversuche des Erstkontaktes (ohne 10 und 99=sys mis), Quelle ESS 2002 Round 1 Hungary

Nach dem vierten Kontaktversuch durch den Erstkontakt geht die Anzahl der Kontaktversuche gegen Eins, wobei zwei Personen einen neunten und einer einen zehnten Kontaktversuch gestartet haben. 19 Personen waren für ein Re-Issuing eingeteilt, das nicht von der Person für den Erstkontakt durchgeführt wurde. Wiederum 2 Personen waren als Drittinterviewer eingeteilt. 838 der 1685 realisierten Interviews wurden

innerhalb der ersten vier Kontaktversuche durch einen der drei Interviewer realisiert, die meisten durch den Erstkontakt.

Die Angaben, wie viele Kontaktversuche insgesamt unternommen wurden, Unit-Non-Responses eingerechnet, fehlen in den meisten Fällen: für die Person des Erstkontaktes fehlen 937, für die Zweitinterviewer 2449 und für den Dritrinterviewer 2479 Angaben. Hier ist nicht klar, wie viele Kontaktversuche unternommen wurden, auch wenn kein Interview realisiert werden konnte.

Aus den Ergebnissen folgend wurde für die Analyse ein Subset gebildet, das Interviews bis zum neunten Kontaktversuch beinhaltet. Dabei wurden sowohl die Erstkontaktpersonen sowie die Ersatzinterviewer miteinbezogen. So ergeben sich insgesamt 208 Interviewer mit unterschiedlicher Anzahl an Interviews. 1685 Interviews wurden miteinbezogen.

**Anzahl der Interviewer pro Anzahl der Interviews**

<b>1</b>	20	<b>13</b>	1
<b>2</b>	7	<b>14</b>	5
<b>3</b>	8	<b>15</b>	4
<b>4</b>	15	<b>17</b>	2
<b>5</b>	9	<b>18</b>	1
<b>6</b>	13	<b>19</b>	1
<b>7</b>	16	<b>20</b>	1
<b>8</b>	30	<b>21</b>	1
<b>9</b>	31	<b>22</b>	3
<b>10</b>	14	<b>26</b>	1
<b>11</b>	9	<b>29</b>	1
<b>12</b>	10	<b>30</b>	1
<b>Summe</b>	<b>1685 Interviews bei 208 Interviewern</b>		

*Tabelle 2 Anzahl der Interviewer pro Interviews basierend auf den erstellten Subset von Kontaktversuchen, Quelle: eigene Berechnungen*

Der Datensatz umfasst diverse Themenbereiche, die über insgesamt 560 Variablen abgedeckt werden, wobei nicht alle Variablen im Datensatz abgefragt wurden. Da es sich um einen integrierten Datensatz handelt, sind auch Variablen enthalten, die nicht im Fragebogen für Befragte in Ungarn enthalten sind. Solche Variablen, die nicht für die Erhebung relevant sind, wurden vor der Analyse aus dem Datensatz entfernt bzw. für

die Analyse als Nullwert umkodiert, da eine Berechnung mit *logischen NAs* selbst im Ergebnis ein *NA* zurückgibt.

Als Grundlage der Analyse dienen die Interviewer-Nummern, wobei ausschließlich solche der ursprünglich ersten Interviews genutzt werden. Da im ESS ein Re-Issuing nicht durch dieselben Personen durchgeführt wird, werden diese in der vorliegenden Analyse nicht betrachtet, sollten bei weiteren Analysen aber separat überprüft, oder in die Analyse mit aufgenommen werden, da ein Restrisiko der Fälschung für Re-Issuings verbleibt.

### 3.1. Reliabilität und Variabilität

Laut Schnell (2008) stellt Messen die regelgeleitete Zuordnung von Zahlen und Objekten dar. Der Vorgang der Messung ist somit eine homomorphe Abbildung eines empirischen Relativs in ein numerisches (Schnell 2008). Im Rahmen der Überprüfung der Güte dieser Zuordnung ist das Ziel eine korrekte Messung zu gewährleisten. Um insbesondere die Zuverlässigkeit einer Messung eines Konstruktes zu gewährleisten, wird in der Literatur häufig Cronbachs Alpha herangezogen, um die Item-Reliabilität zu überprüfen. Die Variabilitätsmethode macht sich dieses Konzept auf einfache Weise nutzbar. Durch abweichendes Antwortverhalten ergeben sich Unterschiede in der Streuung der Ausprägungen einzelner Instrumente. So weisen Fälschungen eine geringere Variabilität in den Ausprägungen einzelner Items auf, da nicht eine wahre Antwort gegeben wird, sondern eine willkürliche Imputation der fälschenden Person erfolgt. Diese Imputation entspricht damit nicht der Itemdimension, sondern dem Eigeninteresse der fälschenden Person, wodurch es zu einer Abweichung zur erwarteten Streuung des Items für eine einzelne Person bzw. ein Cluster von Interviews kommt.

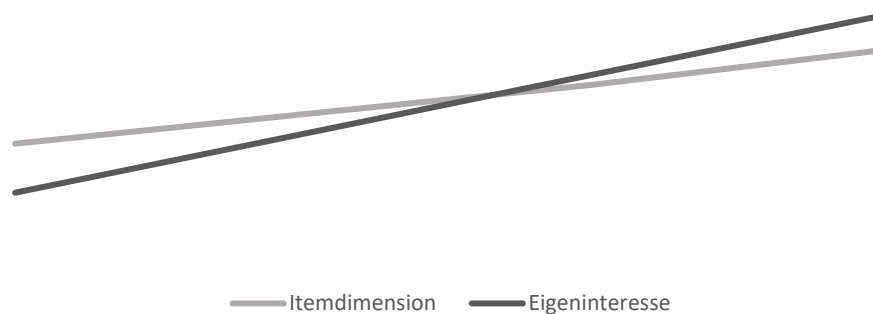


Abbildung 3 Theoretische Item-Dimensionalität

Kontrafaktisch wollen sich aber nicht widersprechen und unauffällig antworten, weshalb es zu einer Überanpassung kommt. Aus dem Nutzenkalkül ergibt sich ebenso ein konsistentes Fälschungsverhalten über mehrere Fragebögen. Folglich ist für echte Befragte die Summe der Abweichungsquadrate

$$SQ_i = \sum_{i=1}^{n_i} (x_i - \bar{x}_i)^2 \quad (3)$$

des Antwortverhaltens charakteristisch. Wobei  $SQ_i$  die „wahre“ Streuung der i-ten ursprünglich befragten Person ist,  $n_i$  die Anzahl der Items der i-ten ursprünglich befragten Person,  $x_i$  das i-te Item ist und  $\bar{x}_i$  der i-te Mittelwert eines Items, das die ursprünglich befragte Person beantwortet hätte.

Wenn nun Interviewer systematisch Fälschungen produzieren, bzw. Werte mittels Imputation schätzt, führt das zu einer systematisch abweichenden Variabilität, die entweder höher oder niedriger ist als diese von durchschnittlichen Befragten. Was zu einem Bias der Variabilität

$$BIAS_{li}(SQ) = \left( \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2 \right) - (SQ_i) \quad (4)$$

innerhalb des Clusters von j Interviews des i-ten Interviews I führt. Somit wird für die Umsetzung der Methode zuerst die jeweilige Variabilität innerhalb des li-ten Interviewer-Fragebogenclusters berechnet, wobei  $m_i$  die Anzahl aller Interviews j ist, die die befragende Person geführt hat. In einem zweiten Schritt wird die Statistik mit der erwarteten Varianz für einen Fragebogen-Cluster der gegebenen Größe auf der gesamten Umfrage verglichen (Schäfer et al. 2004).

„Somit ergeben sich  $l_i$ , mit  $i = 1, \dots, n$ , Interviewern  $l_i$  und  $n$ , die Anzahl der Interviewer, die die Umfrage durchgeführt haben. Die Anzahl der Fragebögen  $Q_j$  wird mit  $j = 1, \dots, m$  angegeben. Wobei  $m = m_1 + \dots + m_i$ , mit  $m_i$  die Anzahl der vom Interviewer  $l_i$  gelieferten Fragebögen bezeichnet. Ohne Rücksicht auf die Bedeutung der Antworten – ob 5 für '5 Jahre' oder für 'Ich bin anderer Meinung' kodiert – wird die Summe der Abweichungsquadrate für jede Frage  $Q_j(k)$  mit  $k = 1, \dots, l$  für alle Fragebögen  $Q_j$  eines Interviewers  $l_i$  berechnet und über alle Fragen summiert:

$$T_{li} = \sum_{k=1}^l \sum_{j=1}^{m_i} (Q_j(k) - \overline{Q(k)})^2$$

(5)

Hier bezeichnet  $\overline{Q(k)}$  den Mittelwert für Frage  $Q(k)$  und der Index  $j$  erfasst alle Fragebögen  $Q_j$  mit  $j = m_{i1}, \dots, m_{i1}$  des Interviewers  $i$ . Die Verteilung der Teststatistik  $T_{li}$  wird mit einem Resampling-Ansatz für die gesamte Umfrage geschätzt. Aus dieser Verteilung kann eine Wahrscheinlichkeit für die beobachteten Wert abgeleitet werden. Im Folgenden werden diese Wahrscheinlichkeiten mit Plausibilität bezeichnet. Durch Sortierung der Interviewer in Bezug auf die erreichte Plausibilität erhält man ein Interviewer Ranking. Die Interviewer mit der geringsten Plausibilität stehen an der Spitze der Rangfolge. Sie werden als potenzielle Fälscher betrachtet. Der Wert von  $T_{li}$ , der dem Interviewer  $i$  zugeordnet ist, wird mit der entsprechenden Verteilung von  $T$ , die mit einem Resampling-Ansatz geschätzt wird, verglichen. Die Fläche unter der Dichtekurve auf der linken Seite der Realisierung  $T_{li}$  definiert die Plausibilität. Wenn die Plausibilität für ein echtes Interview zu gering ist, wird der Interviewer als potenzieller Fälscher angesehen. Das Verfahren entspricht einem einseitigen statistischen Test“ (Schäfer et al. 2004a, S. 11–12).





## 4. Ergebnisse

Im Folgenden Abschnitt werden die Ergebnisse der Variabilitätsanalyse der Interviewer-Cluster vorgestellt.

Abbildung 4 zeigt das Ergebnis des Resamplings. Insgesamt wurden 208 Interviewer mit 1685 Interviews in die Analyse miteingeschlossen. Durchschnittlich ergibt sich eine totale Variabilität von 177.189.578. Interviewer mit auffällig niedriger Variabilität ergeben sich aus dem linksseitigem Testverfahren (unteres Quantil=0.05). 46 Interviewenden, wurden demnach als potenzielle Fälscher eingestuft. 257 Interviews fallen unter diese Interviewer-Cluster. Dies entspricht einem Anteil von 15,25% insgesamt, der potenzielle Interviewfälschungen enthält.

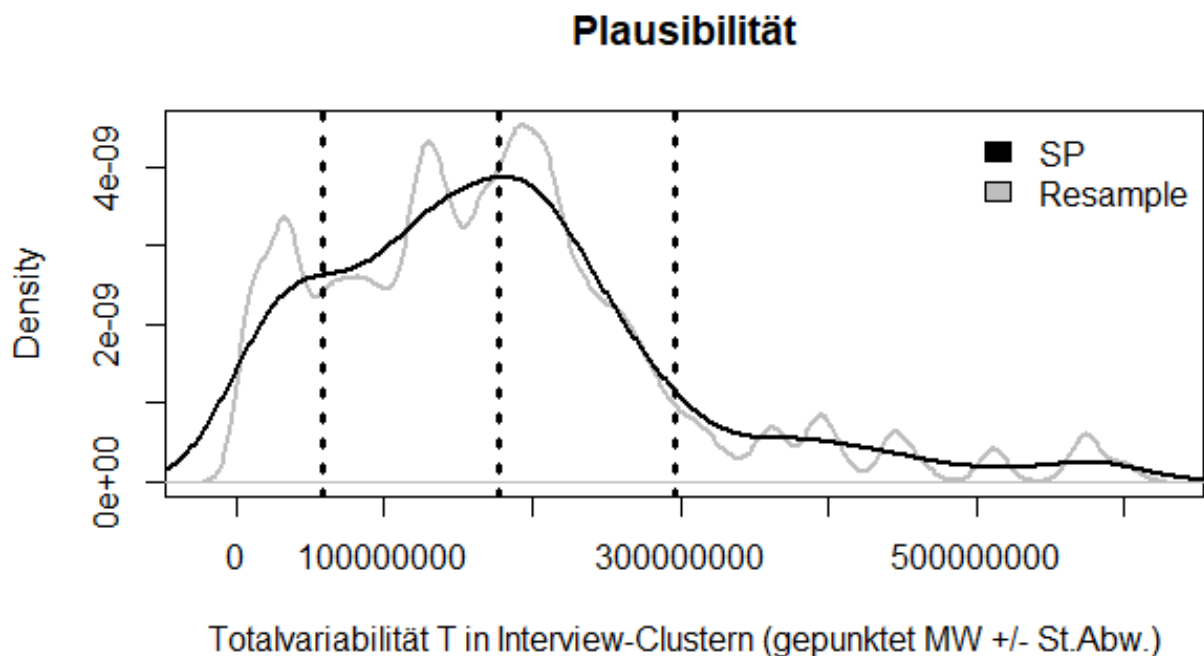


Abbildung 4 Empirische (schwarz) und simulierte (grau) Dichtefunktion der totalen Variabilität der nach Interviewer geclusterten Interviews, Quelle: eigene Berechnungen

Tabelle 3 zeigt das Ranking der ersten 28 Interviewer bis zu einer kumulierten Plausibilität bis 0,02. Insbesondere Interviewer mit nur einem Interview weisen sehr niedrige Plausibilitätswerte auf, so sind alle 11 Erst-Interviewer, die ein Interview realisierten unter den potenziellen Fälschern. Dies erscheint nachvollziehbar, da die summierte Totalstreuung eines Interviewers mit nur einem Interview geringer ist als die eines Interviewers mit mehreren Interviews. Dies muss bei der Evaluation berücksichtigt werden.

INTNUM1	TOTALVARIABILITÄT	PLAUSIBILITÄT	KUM PLAUSIBILITÄT	ABS ANZAHL INT
2086	6044302	0.000164	0.000164	1
<b>4114</b>	<b>6727375</b>	<b>0.000183</b>	<b>0.000347</b>	<b>6</b>
<b>10039</b>	<b>8882091</b>	<b>0.000242</b>	<b>0.000589</b>	<b>16</b>
<b>15051</b>	<b>8987421</b>	<b>0.000245</b>	<b>0.000834</b>	<b>6</b>
24097	9833070	0.000268	0.001101	1
3130	10259613	0.000279	0.001380	4
3131	10575177	0.000288	0.001668	1
39057	10851075	0.000295	0.001963	1
3023	13360272	0.000364	0.002327	1
8234	15281764	0.000416	0.002743	1
<b>4134</b>	<b>17409997</b>	<b>0.000474</b>	<b>0.003216</b>	<b>4</b>
<b>2074</b>	<b>23016376</b>	<b>0.000626</b>	<b>0.003843</b>	<b>5</b>
<b>19002</b>	<b>25141598</b>	<b>0.000684</b>	<b>0.004527</b>	<b>13</b>
10152	26059754	0.000709	0.005236	1
<b>18026</b>	<b>28163642</b>	<b>0.000766</b>	<b>0.006002</b>	<b>10</b>
<b>3010</b>	<b>29665678</b>	<b>0.000807</b>	<b>0.006809</b>	<b>17</b>
9118	29720965	0.000809	0.007618	1
34048	29848579	0.000812	0.008430	1
<b>21021</b>	<b>30718697</b>	<b>0.000836</b>	<b>0.009266</b>	<b>12</b>
8787	32983275	0.000897	0.010163	1
<b>10050</b>	<b>33693909</b>	<b>0.000917</b>	<b>0.011080</b>	<b>16</b>
3135	34447554	0.000937	0.012017	1
<b>22052</b>	<b>35017774</b>	<b>0.000953</b>	<b>0.012970</b>	<b>27</b>
19021	37094844	0.001009	0.013979	4
1054	37950168	0.001033	0.015012	3
4213	40684578	0.001107	0.017186	3
9009	42519354	0.001157	0.018343	4
25053	44614890	0.001214	0.019557	4
:	:	:	:	:

Tabelle 3 Globales Interviewer-Ranking erste 28 Personen nach Schätzer  $T_i$  und Plausibilität für potenzielle Fälschung, Quelle: eigene Berechnungen

Auffällig ist, dass Interviewern (Intnum1) mit einer hohen Anzahl Interviews, ähnlich geringe totale Variabilität zugeordnet wird, wie Interviewern mit einer sehr viel geringeren Anzahl an Interviews (Abs. Anzahl Int). So bilden die ersten vier Interviewer im Ranking ein jeweils ähnliches Wertpaar, trotz der unterschiedlichen Interviewzahlen.

Auf Grundlage des Testes sollten Interviews dieser Interviewenden einer Nachuntersuchung unterstellt werden. Da nur gruppierte Werte abgebildet werden und die Nachuntersuchung von 257 Interviews als nicht effizient angesehen wird, wurden für die Entscheidung in einem zweiten Schritt, die Interviews der potenziellen Fälscher mit Variabilitätsschwellenwerten unter 10.000.000 extrahiert. Somit ergeben sich 35 Interviews von 11 Personen (in Tabelle 3 fett markiert) in Tabelle 4 gruppiert nach Interviewer und Befragten-ID, die eine auffällig geringe Variabilität innerhalb eines Interviewer-Clusters aufweisen. Insgesamt sind dies rund 2 Prozent aller Interviews.

INTNUM1	BEFRAGTEN-ID	TOTALVARIABILITÄT	REL ANZAHL INT
<b>4114</b>	1896	5.640.451	<b>2/6</b>
<b>4114</b>	1900	6.727.381	
<b>10039</b>	1486	6.512.203	<b>5/16</b>
<b>10039</b>	1497	8.882.094	
<b>10039</b>	1527	8.037.321	
<b>10039</b>	1529	8.051.469	
<b>10039</b>	1533	8.033.013	
<b>15051</b>	82	6.759.770	<b>3/6</b>
<b>15051</b>	83	7.344.882	
<b>15051</b>	100	8.987.427	
<b>2074</b>	1121	7.310.791	<b>2/5</b>
<b>2074</b>	1132	3.629.677	
<b>19002</b>	131	9.136.788	<b>4/13</b>
<b>19002</b>	139	6.754.736	
<b>19002</b>	1998	9.043.844	
<b>19002</b>	1999	6.147.009	
<b>18026</b>	44	9.759.763	<b>3/10</b>
<b>18026</b>	46	8.101.205	
<b>18026</b>	47	9.456.914	
<b>3010</b>	1843	6.333.800	<b>2/17</b>
<b>3010</b>	1870	8.428.112	
<b>21021</b>	286	9.251.519	<b>2/12</b>
<b>21021</b>	307	4.991.735	
<b>10050</b>	71	6.288.748	<b>4/16</b>
<b>10050</b>	1492	7.465.428	
<b>10050</b>	1526	7.297.936	
<b>10050</b>	1530	7.323.056	
<b>22052</b>	943	9.078.136	<b>8/27</b>
<b>22052</b>	948	9.136.610	
<b>22052</b>	1099	6.293.232	
<b>22052</b>	1103	9.955.617	
<b>22052</b>	1106	4.507.172	
<b>22052</b>	2541	8.190.717	
<b>22052</b>	2542	2.454.477	
<b>22052</b>	2543	7.416.645	

Tabelle 4 Relative Anzahl der potenziell gefälschten Interviews der in Tabelle 3 markierten Interviewer, Quelle: eigene Berechnungen

Innerhalb der Interviewer-Cluster weichen die Variabilitätswerte der extrahierten Interviews teilweise deutlich ab. Im Falle des Interviewers 4134 wurde kein Interview extrahiert, da keines der durchgeführten einen Wert unter 10.000.000 aufweist.

Die detektierten Interviewer weisen durchschnittlich einen Anteil von 29,84 Prozent der Interviews mit einem Variabilitätswert unter 10.000.000 auf. Zudem weichen die Variabilitätswerte innerhalb ihres Clusters deutlich von den übrigen Interviews ab, die meist einen 3 bis 4-fach höheren Wert aufweisen. Ebenso wird deutlich, dass die extrahierten Interviews des Clusters starke Ähnlichkeit der Variabilität aufweisen, was darauf hindeutet, dass sie ähnlich beantwortet wurden.

Hinsichtlich der ersten Arbeitshypothese lässt sich festhalten, dass 11 Interviewer für 35 Interviews auffällig niedrige Variabilitätswerte aufweisen, die sich stark ähneln. Dies kann an der Gesamtlänge, dem übergehen von Filterfragen, oder der gleichmäßigen Beantwortung von Item-Batterien liegen. Die Interviews zeigen ähnliche Kennwerte. Damit bilden diese Interviews eine hohe Konsistenz des Antwortverhaltens innerhalb des Clusters und innerhalb der Interviews ab. Dies ist jedoch lediglich ein Indiz für eine Fälschung. Auch „wahre“ Befragte weisen diese Art des Verhalten auf. Zudem ist nicht eindeutig, ab welchen Wert, das Interview als plausible Fälschung angesehen werden sollte. Eine empirische Entscheidungsgröße fehlt. Das Argument der geringeren Streuung lässt nicht den eindeutigen kausalen Schluss auf das antezedierendes Ereignis der Fälschung zu, weshalb die erste Arbeitshypothese vorläufig nicht bestätigt werden kann. Hierfür bedarf es einer Nachuntersuchung zur eindeutigen Entscheidung.

Unter den plausiblen Fälschern sind 7 Personen mit einer zweistelligen Anzahl an Interviews. Interviewer 15051 und 22052 haben dabei beide einen Anteil an gerundet 30 Prozent an Fälschungen und damit die höchsten Anteile. Diese sollten explizit nachuntersucht werden. Aus dem Nutzenkalkül unter Informationsasymmetrie ergibt sich die zweite Arbeitshypothese. Fälschungen häufen sich bei einzelnen Interviewern. Die Ergebnisse in Tabelle 3 und 4 zeigen, dass insbesondere Interviewer mit einem hohen Anteil an Interviews als potenzielle Fälscher detektiert wurden. Die zweite Arbeitshypothese findet teilweise Bestätigung in den Daten. Potenzielle Fälschungen häufen sich bei einzelnen Interviewenden. Die Arbeitserfahrung der Interviewenden kann nicht überprüft werden, da diese Daten nur intern vorliegen. Eine endgültige Entscheidung bedarf auch hier einer Nachuntersuchung.

Die evaluierten Interviews zeigen signifikante Abweichungen vom vorhandenen Datensatz und sollten basierend auf der Analyse einer Nachuntersuchung zugeführt werden, um Fälschungen valide ausschließen zu können. Bestätigte Fälschungen könnten dann für weitere Nachuntersuchungen genutzt werden.

## 5. Zusammenfassung

Die Arbeit hat die Variabilitätsmethode zur Detektion potenzieller Fabrikate anhand des Datensatzes des European Social Surveys 2002 in Ungarn vorgestellt. Die Plausibilität für eine Fälschung wurde mittels Resampling-Ansatzes ermittelt und stellt die linksseitige Fläche unter der Verteilung dar. Die Frage, ob *einzelne Interviews signifikant abweichendes Verhalten aufweisen*, lässt sich eindeutig bestätigen. Es ergaben sich bei einer Grenzwert von 0,02 für die kumulierte Plausibilität 28 potenziell fälschende Personen, wobei unter denen mit mehr als einem Interview 11 Interviewende als vorrangig extrahiert wurden, die einer Prüfung bei 35 Interviews bedürfen. Dies entspricht 2 Prozent aller Interviews. Sie weichen signifikant in ihrer Totalvariabilität innerhalb eines Interviewer-Fragebogenclusters nach unten ab.

Die Forschungsfrage, ob *potenzielle Interviewfälschungen adäquat anhand ihrer abweichenden Streuung zu detektieren sind*, kann nicht eindeutig beantwortet werden. Die Gesamtstreuung ist kein trennscharfes Kriterium für die Entscheidung. Das antezedierende Ereignis der Konklusion „geringe Streuung“ ist nicht zwingend eine Fälschung, sondern auch konsistentes Antwortverhalten einer befragten Person, die andere Gründe haben kann. Geringe Streuung ist keine hinreichende Bedingung für eine Fälschung. Sie ist auch keine notwendige, sofern die Fälschung sehr gut ist. Bestätigt werden kann, dass sich potenzielle Fälschungen bei einzelnen Interviewenden häufen. Die identifizierten Interviews sollten einer Nachuntersuchung unterstellt werden, um eine Fabrikation ausschließen oder bestätigen zu können.

Ein großer Vorteil der vorgestellten Methodik ist die einfache Implementierung und Umsetzung, und die theoretisch objektive Begründbarkeit. Aufwendigere Verfahren liefern im Vergleich tendenziell nicht immer reliablere Ergebnisse bei der Aufdeckung von Fälschern (vgl. Schäfer et al. 2004b) und bedeuten einen wesentlich höheren Ressourcenaufwand und/oder Wissensstand.

Die Methodik weist bei nicht bekannten Fälschenden Reliabilitätsschwächen auf. Dass die identifizierten Interviews tatsächlich Fälschungen sind, ist nicht eindeutig. Diese

Schwäche haben ex-post Methodiken grundsätzlich inne. Die Identifikation von Fälschungen ist nur in Folge der Nachuntersuchung möglich.

Die Methodik beruht auf der Annahme, dass Fälschungen eine höhere Konsistenz des Antwortverhaltens abbilden. Dies ist nicht trennscharf auf eine Fälschung zurückzuführen. Es muss eine hinreichend große Anzahl an Interviews pro Interviewer vorliegen, um zu entscheiden, dass die Variabilität innerhalb eines Clusters geringer ist. Professionelle Fälscher weisen ein höheres und breiteres Erfahrungswissen über Parameter und den Fragebogen auf, sodass sie von dieser Methodik insgesamt unentdeckt blieben. Ein eindeutiger Grenzwert für die Entscheidung, ob die Variabilität zu gering ist, ist nicht vorhanden, sondern wird datengetrieben entschieden, womit die Kenngröße je nach Stichprobe unterschiedlich ausfällt. Hier bedarf es weiterer Untersuchungen.

Grundsätzlich bestehen zwei Ansätze Fälschungen anzugehen: (1) sie postdiktiv ursächlich zu erklären, damit ex-post aufzudecken oder (2) die Prävention. Um sie ex-post aufzudecken, kann aus den Ausführungen von Menold und Kemper (2014) abgeleitet werden, dass die Möglichkeit bestünde, Fragebögen mit zusätzlichen Kontrollvariablen zu versehen, die grundsätzlich schwer zu fälschen sind und deren Verteilung in der Grundgesamtheit bekannt ist. Dafür eignen sich individuell psychologische Aspekte sowie subjektive und kollektive Sichtweisen (Menold und Kemper 2014). Hieraus ergeben sich mögliche Kontrollparameter für Fälschungen.

Es besteht ein breites Spektrum an Sichtweisen auf Fälschungen, wobei psychologische den größten Teil ausmachen. Zur Beantwortung der Frage, *welche Faktoren ursächlich zur Erklärung einer Fälschung sind*, sieht der Autor eine rationalistische Erklärung für Fälschungen als ursächlich an. Fälschende Personen verfolgen ein kostenreduzierendes Kalkül unter der Annahme der Informationsasymmetrie. Die Detektion von Fälschungen ist potenziell methodisch aufwendig, daher ist eine wichtige Maßnahme die Prävention von Fabrikaten durch geeignete Maßnahmen, wie Supervision, angemessene Lohn- und Fortbildungsmaßnahmen (Schnell 2019, 8.5; 8.6; 8.7.2), um Informationsasymmetrien zu verringern und dem rationalen Kalkül entgegenzuwirken. Die psychologischen Ausführungen bieten ferner Anlass zur Prävention durch psychologische Eignungstests des anzustellenden Personals. Risikoverhalten, spezifisches Fachwissen zu Verteilungen, Tendenzen zu unehrlichen Verhalten wären zu prüfen. Personen mit hohen Testwerten, könnten abgelehnt werden; oder

eingestellt und bei der Nachuntersuchung besonders berücksichtigt. Hierbei sollte ethisch geklärt werden, ob Personal einzustellen, deren Loyalität von Beginn an in Abrede gestellt würde, ein probates Mittel wäre. Ein simpleres Mittel wäre eine angemessenere Bezahlung und weniger Arbeitsdruck. Also Vermeidung von Demoralisation und prekärer Beschäftigung.

Der Autor erachtet, die Wichtigkeit der Prävention als vorrangig, gegenüber der Aufdeckung von Fabrikaten. Da eine Aufdeckung von Betrug schwer methodisch zu erreichen ist und keine letztendliche Garantie besteht, dass einzelne Fälschungen oder Teilfälschungen unentdeckt bleiben. Die Arbeit konnte einen Beitrag zur Erweiterung der Forschungsergebnisse zu bestehenden Methodiken zu Prävention und Detektion von Interviewfälschungen liefern. Weitere Forschung, um robuste Schätzungen aus Bevölkerungsumfragen zu gewinnen, bleibt nötig.

## 6. Literaturverzeichnis

Al-Marzouki, Sanaa; Evans, Stephen; Marshall, Tom; Roberts, Ian (2005): Are these data real? Statistical methods for the detection of data fabrication in clinical trials. In: *BMJ* 331 (7511), S. 267–270. DOI: 10.1136/bmj.331.7511.267.

Bergmann, Michael; Schuller, Karin; Malter, Frederic (2019): Preventing interview falsifications during fieldwork in the Survey of Health, Ageing and Retirement in Europe (SHARE). In: *Longitudinal and Life Course Studies* 10 (4), S. 513–530. Online verfügbar unter <https://doi.org/10.1332/175795919X15694136530293>.

Birnbaum, Benjamin (2012): Algorithmic Approaches to Detecting Interviewer Fabrication in Surveys.

Blasius, Jörg (2014): Fälschungen von Interviews. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 323–329.

Blasius, Jörg; Friedrichs, Jürgen (2012): Faked Interviews. In: Samuel Salzborn, Eldad Davidov und Jost Reinecke (Hg.): *Methods, Theories, and Empirical Applications in the Social Sciences*. Festschrift für Peter Schmidt. Wiesbaden: Springer VS, S. 49–56.

Blasius, Jörg; Thiessen, Victor (2012): *Assessing the Quality of Survey Data*. London: Sage.

Blasius, Jörg; Thiessen, Victor (2013): Detecting Poorly Conducted Interviews. In: Peter Winkler, Natalja Menold und Rolf Porst (Hg.): *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*. Frankfurt: Peter Lang, S. 67–88.

Blasius, Jörg; Thiessen, Victor (2015): Should we Trust Survey Data? Response Simplification and Data Fabrication. In: *Social Science Research*.

Bredl, Sebastian; Storfinger, Nina; Menold, Natalja (2011): A literature review of methods to detect fabricated survey data. Discussion Papers. Justus Liebig University Gießen, Center for international Development and Environmental Research (ZEU) (56). Online verfügbar unter <https://EconPapers.repec.org/RePEc:zbw:zeudps:56>.

Bredl, Sebastian; Winker, Peter; Koetschau, Kerstin (2012): A statistical approach to detect interviewer falsification of survey data. In: *Survey Methodology* 38, S. 1–10.



Buehl, Anne-Kathrin; Melchers, Klaus G. (2017): Individual Difference Variables and the Occurrence and Effectiveness of Faking Behavior in Interviews. In: *Frontiers in psychology* 8, S. 1–15. DOI: 10.3389/fpsyg.2017.00686.

Bushery, John M.; Reichert, Jennifer W.; Albright, Keith A.; Rossiter, John C. (1999): Using Date and Time Stamps to Detect Interviewer Falsification. In: *Proceedings of the Survey Research Methods Section, ASA* (9), S. 316–320. Online verfügbar unter [http://www.asasrms.org/Proceedings/papers/1999\\_053.pdf](http://www.asasrms.org/Proceedings/papers/1999_053.pdf), zuletzt geprüft am 27.01.2021.

Crespi, Leo P. (1945): The Cheater Problem in Polling. In: *Opinion Quarterly* (9), S. 431–445.

European Social Survey ERIC (ESS ERIC) (2003): European Social Survey (ESS), Round 1 - 2002.

Felder, Barabara; Birg, Alexandra; Kreuter, Frauke (2014): Paradata. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 357–366.

Gwartney, Patricia (2013): Mischief Versus Mistakes: Motivating Interviewers to not Deviate. In: Peter Winker, Natalja Menold und Rolf Porst (Hg.): *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*. Frankfurt am Main: Peter Lang, S. 195–215.

Haas, Samuel de; Winker, Peter (2016): Detecting Fraudulent Interviewers by Improved Clustering Methods – The Case of Falsifications of Answers to Parts of a Questionnaire. In: *Journal of Official Statistics* 32 (3), S. 643–660. DOI: 10.1515/jos-2016-0033.

Harrison, David E.; Krauss, Stefanie L. (2002): Interviewer Cheating: Implications for Research on Entrepreneurship in Africa. In: *Journal of Developmental Entrepreneurship* (7), S. 319–330.

Hood, C. C.; Bushery, John M. (1997): Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers". In: *Proceedings from Section on Survey Research Methods, ASA*, S. 820–824, zuletzt geprüft am 27.01.2021.

Kemper, Christoph J.; Menold, Natalja (2014): Nuisance or Remedy? The Utility of Stylistic Responding as an Indicator of Data Fabrication in Surveys. In: *Methodology* 10 (3), S. 92–99. DOI: 10.1027/1614-2241/a000078.

Kuchen, Robert (2018): Detecting Fabricated Interview Data in a Two-Level Approach Using the Mahalanobis Distance. Johannes Gutenberg-Universität Mainz, Universitätsmedizin.

Kuriakose, Noble; Robbins, Michael (2016): Don't get duped: Fraud through duplication in public opinion surveys. In: *SJI* 32 (3), S. 283–291. DOI: 10.3233/SJI-160978.

Menold, Natalja; Winker, Peter; Storfinger, Nina; Kemper, Christoph J. (2013): A method for ex-post identification of falsifications in survey data. In: Peter Winkler, Natalja Menold und Rolf Porst (Hg.): Interviewers' deviations in surveys. Frankfurt am Main: Peter Lang, S. 25–48. Online verfügbar unter <https://search.gesis.org/publication/gesis-bib-141178>.

Menold, Natalja.; Kemper, Christoph. J. (2014): How Do Real and Falsified Data Differ? Psychology of Survey Response as a Source of Falsification Indicators in Face-to-Face Surveys. In: *International Journal of Public Opinion Research* 26 (1), S. 41–65. DOI: 10.1093/ijpor/edt017.

Murphy, Joe; Baxter, Rodney; Eyerman, Joe; Cunningham, David; Kennet, Joel (2004): A System for Detecting Interviewer Falsification. [American Association for Public Opinion Research 59th Annual Conference]. Hg. v. American Association for Public Opinion Research Section on Survey Research Methods. Phoenix, Arizona. Online verfügbar unter <https://www.amstat.org/Sections/Srms/Proceedings/y2004/Files/Jsm2004-000517.pdf>, zuletzt geprüft am 27.01.2021.

Nelson, James E.; Kiecker, Pamela L. (1996): Marketing Research Interviewers and Their Perceived Necessity of Moral Compromise. In: *Journal of Business Ethics* (15), S. 1107–1117.

Pearl, Judea; Mackenzie, Dana (2018): The Book of Why. The New Science of Cause and Effect. New York, NY: Basic Books.

Röhner, J.; Schütz, A. (2015): Psychologie der Kommunikation: Springer Fachmedien Wiesbaden (Basiswissen Psychologie).

Schäfer, Christin; Schräpler, Jörg-Peter; Müller, Klaus-Robert (2004a): Identification, Characteristics and Impact of Faked and Fraudulent Interviews in Surveys. Deutsches Institut für Wirtschaftsforschung (DIW). Berlin. Online verfügbar unter [https://www.diw.de/documents/dokumentenarchiv/17/diw\\_01.c.41963.de/paper2004\\_schaeferetal.pdf](https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.41963.de/paper2004_schaeferetal.pdf), zuletzt geprüft am 27.01.2021.

Schäfer, Christin; Schräpler, Jörg-Peter; Müller, Klaus-Robert; Wagner, Gert G. (2004b): Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methods. Deutsches Institut für Wirtschaftsforschung (DIW). Berlin (DIW Discussion Papers, 441). Online verfügbar unter <http://hdl.handle.net/10419/18293>, zuletzt geprüft am 02.05.2020.

Schnell, Rainer (1991): Der Einfluss gefälschter Interviews auf Survey-Ergebnisse. In: *Zeitschrift für Soziologie* (20), S. 25–35.

Schnell, Rainer (2013): Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen: Springer-Verlag.

Schnell, Rainer (2019): Survey-Interviews. Methoden standardisierter Befragungen, 2. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden, zuletzt geprüft am 27.01.2021.

Schräpler, Jörg-Peter; Wagner, Gert G. (2005): Characteristics and impact of faked interviews in surveys. An analysis of genuine fakes in the raw data of SOEP. In: *Allgemeines Statistisches Archiv* 89, S. 7–20, zuletzt geprüft am 02.05.2020.

## 7. Anhang

Im Folgenden ist der genutzte Algorithmus zur Erstellung und Auswertung der genutzten Daten sowie die Eigenständigkeitserklärung zu finden. Der genutzte Code findet sich unter CC BY-SA Lizenz ebenso unter <https://kevinglock.github.io/IntFabric/index.html>.

### 7.1 R Code

#### 7.1.1. Required libraries

These libraries are used for the analysis.

```
library(foreign) # read.spss
library(dplyr)   # using pipes
library(haven)   # read_por
```

#### 7.1.2. Load respondent's data

Get the data on respondents.

```
resp <- read.spss(
  "D:/IntFabric/data/ESS1HU.sav", # please specify the directory
  to.data.frame=T,
  use.value.labels = F
) # data on respondents

dim(resp) # 1685 entries and 566 variables
```

Get the data on interviewers.

```
int <- read_por("D:/IntFabric/data/ESS1cfHU.por" # please specify the directory
) # data on interviewers

dim(int) # 2484 entries and 176 variables
```

### 7.1.3. Merge data

Get rid of not used variables and merge the sets to a new merged subset.

```
data1 <- subset.data.frame(resp,
                           select = -c(name, essround,
                                       edition, proddate,
                                       cntry, pweight,
                                       pspwght, dweight,
                                       ctzship, cntbrth,
                                       lnghoma, lnghomb)
                           ) # remove irrelevant and character variables

dim(data1) # 1685 entries and 554 variables

data2 <- subset.data.frame(int,
                           select = c(IDNO, INTNUM1, INTNUM2, INTNUM3,
                                       TOTCINT1, TOTCINT2, TOTCINT3)
                           )
dim(data2) # 2484 entries and 7 variables

mergeddata <- merge(data1, data2,
                    by.x = "idno", by.y = "IDNO",
                    sort = T, no.dups = F)

mergeddata <- mergeddata %>% select(idno, INTNUM1, TOTCINT1,
                                INTNUM2, TOTCINT2, INTNUM3, TOTCINT3, everything()
                                )
dim(mergeddata) # 1685 entries and 560 variables
head(mergeddata, n = 10)
```

### 7.1.4. Methodological info on dataset

Total non-response rate.

```
2484 - 1685 #=799
(799 * 100)/2484 #=32.17 %
1 - 0.3217 # =67.83 %
(1685/2484) * 100 # =67.83%
```

How many interviewers?

```
unique(int$INTNUM1)
# 198 unique interviewer who started first contact (one NA)
unique(int$INTNUM2) # 19 (one NA)
unique(int$INTNUM3) # 2 (one NA)
sum(table(int$INTNUM1))
# 2461 first contacts 2484-2461=23 NAs
sum(table(int$INTNUM2)) # 35 re-issues
sum(table(int$INTNUM3)) # 4 re-issues
table(int$TOTCINT1)
table(int$TOTCINT2)
table(int$TOTCINT3)
# 436+224+124+45+7+2=838 436+224+124+45+937=1766
rm(list = c("resp", "int"))
```

How many total contacts being made by the interviews?

```
data5.0 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 0)
data5.1 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 1)
data5.2 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 2)
data5.3 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 3)
data5.4 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 4)
data5.5 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 5)
data5.6 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 6)
data5.7 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 7)
data5.8 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 8)
data5.9 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 9)
data5.10 <- mergeddata %>% subset(mergeddata$TOTCINT1 == 10)

dim(data5.0)
# 689 for all desired interviews resp. 513 for all performed interview
dim(data5.1) # 436 resp. 336
dim(data5.2) # 224 resp. 147
dim(data5.3) # 124 resp. 69
dim(data5.4) # 45 resp. 20
dim(data5.5) # 16 resp. 7
dim(data5.6) # 8 resp. 2
dim(data5.7) # 2 resp. 1
dim(data5.8) # 1 resp. 0
dim(data5.9) # 2 resp. 0
dim(data5.10) # 0 resp. 0

data5.0.1 <- mergeddata %>% subset(mergeddata$TOTCINT2 == 0)
data5.1.1 <- mergeddata %>% subset(mergeddata$TOTCINT2 == 1)
data5.2.1 <- mergeddata %>% subset(mergeddata$TOTCINT2 == 2)

dim(data5.0.1) # 7 resp. 7
dim(data5.1.1) # 4 resp. 4
dim(data5.2.1) # 0 resp. 0

data5.0.2 <- mergeddata %>% subset(mergeddata$TOTCINT3 == 0)
data5.1.2 <- mergeddata %>% subset(mergeddata$TOTCINT3 == 1)
data5.2.2 <- mergeddata %>% subset(mergeddata$TOTCINT3 == 2)

dim(data5.0.2) # 1 resp. 1
dim(data5.1.2) # 1 resp. 0
dim(data5.2.1) # 0 resp. 0

rm(list = c("data5.0", "data5.1", "data5.2", "data5.3", "data5.4",
  "data5.5", "data5.6", "data5.7", "data5.8", "data5.9", "data5.10",
  "data5.0.1", "data5.1.1", "data5.2.1", "data5.3.1", "data5.4.1",
  "data5.5.1", "data5.6.1", "data5.7.1", "data5.8.1", "data5.9.1",
  "data5.10.1", "data5.0.2", "data5.1.2", "data5.2.2"))
```

### 7.1.5. Methodological info on subset

How many interviewers?

```
length(unique(mergeddata[, "INTNUM1"])) # 193 interviewers
unique(mergeddata[, "INTNUM2"]) # 15 interviewers (plus one NA)
unique(mergeddata[, "INTNUM3"]) # 0 interviewer (plus one NA)
# 193+15=208 interviewers
```

### 7.1.6. Variability approach

For the variability approach in the first step the variance for each interviewer and every question is calculated and summed up. The difference between each observation for a question by interviewer and the mean of the question will provide a t-statistic which can be tested against a resampled distribution.

```
dim(mergeddata)

anew <- subset.data.frame(mergeddata, select = -c(3, 5, 7)) %>%
  group_by(idno, INTNUM1)
# sort it by interviewer one and match each interview

head(anew)

anew[is.na(anew)] <- 0 # ALL NAs should be 0 because of calculations

dim(anew)
head(anew) # Now the data is tidy; 1685 obs on 557 var
```

### 7.1.7. Calculate the means

Pay attention for 1685 entries, 557 total columns but idno and INTNUM1-3 should not be included.

```
bnew <- anew[, c(5:557)] %>% colMeans() %>% rbind.data.frame()

names(bnew) <- colnames(anew[5:557])

head(bnew) # now we have the mean values for every variable

bnew <- bnew[rep(seq_len(ncol(bnew)), each = 1685), ]
# each must be adapted

head(bnew, n = 5)
tail(bnew, n = 5)

bnew <- bnew[-c(1686:nrow(bnew)), ]
# make sure there are no redundant rows

nrow(bnew) # 1685

# now we have a df with all means for every variable for each observation
# we have to apply the calculation for each observation
head(anew)
```

```
cnew <- anew[, -(1:4)]
# we have to remove idno and INTNUM1-3 to make row wise calculations in the next step

dim(cnew) # stated values for each interview; 1685 obs and 553 vars
dim(bnew) # mean for each question; 1685 obs and 553 vars
```

### 7.1.8. Calculate variabilities

Variability for each question over each interview

```
calc <- (cnew - bnew)^2
head(calc)

calc2 <- cbind(anew[, c("idno", "INTNUM1", "INTNUM2", "INTNUM3")],
               calc)

dim(calc2) # like before, 1685 obs on 557 vars
head(calc2)
```

How many interviews per interviewer?

```
sort(table(calc2[, 3])) # 25 interviews by IntNum 2
sort(table(calc2[, 4])) # 0 interviews by IntNum 3
sort(table(calc2[, 2])) # 1685-25 = 1660 interviews by IntNum 1
length(unique(calc2$INTNUM1))
length(unique(calc2$INTNUM2))

calc3 <- calc2 %>% group_by(INTNUM1, INTNUM2) %>% summarise(.groups = "keep",
                  across(.cols = 5:554, .fns = sum))
# sum up for each interviewer over all questions

rm(list = c("anew", "bnew", "cnew"))

dim(calc3)
# 208 unique interviewer because of grouping by INTNUM1-2 and 552 vars
head(calc3)
```

### 7.1.9. Get the test statistic T

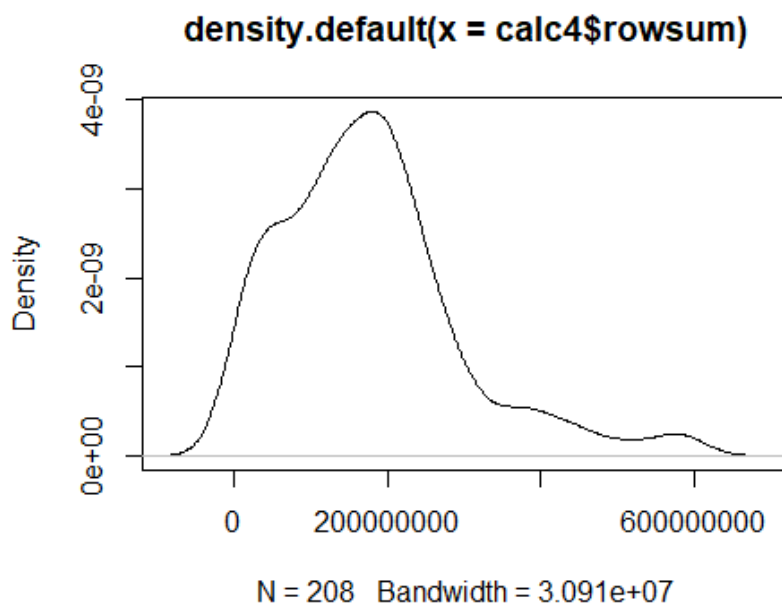
Get the test statistic T (total variability) for the clustered interviewer and for the whole survey.

```
calc4 <- cbind.data.frame(calc3[, c("INTNUM1", "INTNUM2")], rowsum = rowSums(
  calc3[, 3:552]))
# create a new object containing the INTNUMs and the total variability aggregated by INTNUMs

sort.calc4 <- calc4[order(calc4$rowsum), ]
# resort it by the rowsums to get ascending values of the variabilities

head(sort.calc4, n = 5)
plot(density(calc4$rowsum))
```





```
options(scipen = 5) # the density of variability over all interviewers

sd(calc4$rowsum) # 119066901
mean(calc4$rowsum) # 176699764

# 176699764-119066901=57632863 176699764+119066901=295766665

sort(table(calc4$INTNUM1)) # we can see that some are double or triple th
is is because they are re-issued by a second and third person
head(calc4, n = 20) # Look on the entries 6 and 7 for example

length(which(sort(calc4$rowsum) < 1e+08)) # fifty-seven interviews show su
spiciously low total variabilities at first sight

head(sort(calc4, n = 5) # Lowest value 6044290 for interviewer 2086; keep
in mind that this are aggregated values!
tail(sort(calc4, n = 5) # highest value 599664843 for interviewer 6049
```

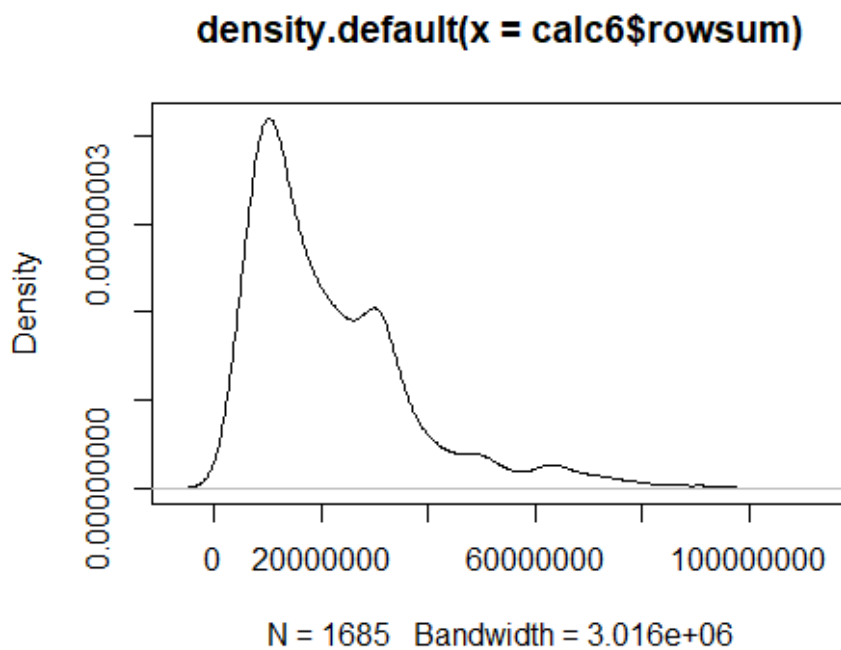
#### 7.1.10. Whole dataset

Get independent variabilities for the interviews.

```
calc5 <- cbind(calc2[c("idno", "INTNUM1")], calc)
head(calc5, n = 5)

calc6 <- cbind.data.frame(calc5[c("idno", "INTNUM1")], rowsum = rowSums(ca
lc5[3:554]))
head(calc6, n = 5)

plot(density(calc6$rowsum))
# density plot for unique interviews not grouped by interviewers
```



```
sd(calc6$rowsum) # 15601634
mean(calc6$rowsum) #21812202

sort.calc6 <- calc6[order(calc6$rowsum), ]
# now sorted by rowsums not by respondents ID

head(sort.calc6, n = 5) # lowest value 2352503 for interview 2308; keep i
n mind that this are separated interview values not aggregated
tail(sort.calc6, n = 5) # highest value 104602583 for interview 2089

write.csv2(calc6, "data.csv")
getwd()
```

### 7.1.11. Resampling

Resample from the subset and check of plausibility.

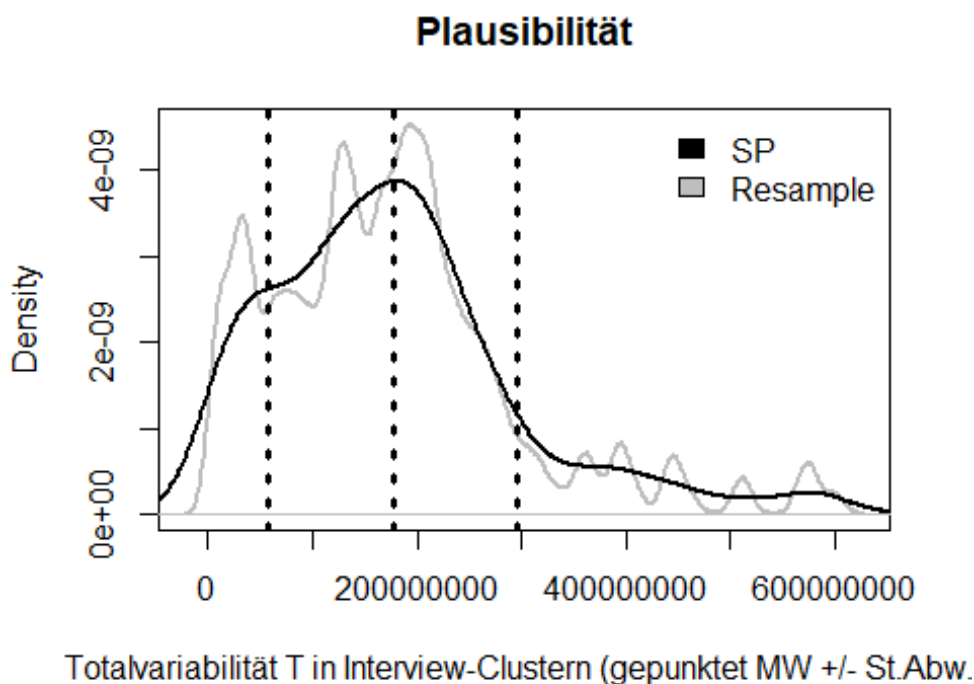
```
N <- 100000
SP <- sample(calc4$rowsum, N, replace = T)
# draw a sample of size N=100,000 from the aggregated rowsums

sd(SP) # 119385175
mean(SP) # 177189578
# 177189578-119385175=57804403 177189578+119385175=296574753

mean(calc4$rowsum) # 176699764
sd(calc4$rowsum) # 119066901
# 176699764-119066901=57632863 176699764+119066901=295766665

plot(density(SP), col = "grey", main = "Plausibilität", xlab = "Totalvaria-
bilität T in Interview-Clustern (gepunktet MW +/- St.Abw.)",
     lwd = 2)
lines(density(calc4$rowsum), col = "black", lwd = 2)
abline(v = c(57804403, 177189578, 296574753), col = "grey", lty = "dotted"
,
      lwd = 3)
abline(v = c(57632863, 176699764, 295766665), col = "black",
      lty = "dotted", lwd = 3)
legend("topright", legend = c("SP", "Resample"), fill = c("black",
"grey"), bty = "n")

options(scipen = 5)
# plot the density of the original rowsums and the resampled ones
```



Get the densities.

```
dat2 <- data.frame(V1 = calc4$rowsum)
round(sort(with(dat2, V1/sum(V1))), digits = 6)

round(cumsum(sort(with(dat2, V1/sum(V1)))), digits = 6)

length(which(round(cumsum(sort(with(dat2, V1/sum(V1)))), digits = 6) <
  0.05)) # 46 interviewer are on the left side below the lower 5 percent
```

Get the interviewer id and rowsums ordered the same way as the densities. Get the number of interviews.

```
sort.calc4$INTNUM1
sort.calc4$rowsum
```

### 7.1.12. Resample for single interviews

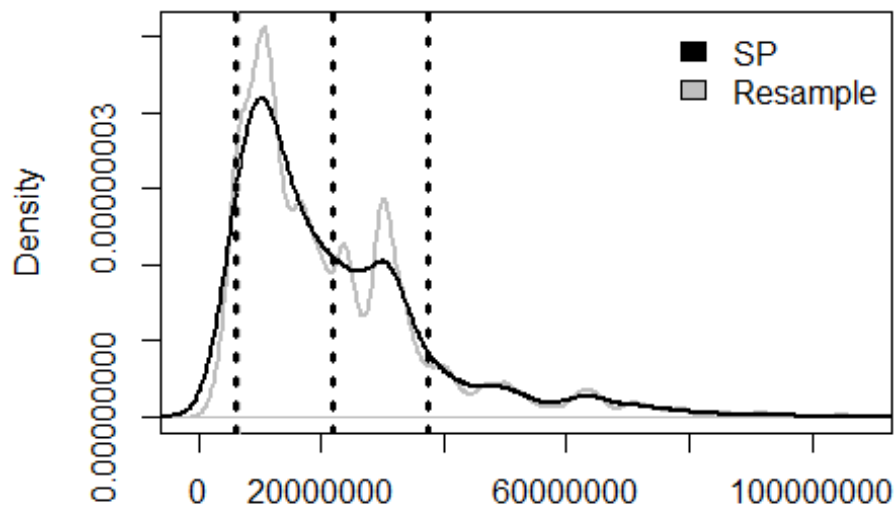
```
N2 <- 100000
SP2 <- sample(calc6$rowsum, N2, replace = T)
# now sample the same resample for the separated interviews

mean(SP2) # 21768691
sd(SP2) # 15497107
# 21768691-15497107=6271584 21768691+15497107=37265798

mean(calc6$rowsum) # 21812202
sd(calc6$rowsum) # 15601634
# 21812202-15601634=6210568 21812202+15601634=37413836

plot(density(SP2), col = "grey", main = "Plausibilität", xlab = "Totalvari-
  abilität T in Einzelinterviews (gepunktet MW +/- St.Abw.)",
  lwd = 2)
lines(density(calc6$rowsum), col = "black", lwd = 2)
abline(v = c(6271584, 21768691, 37265798), col = "grey", lty = "dotted",
  lwd = 3)
abline(v = c(6210568, 21812202, 37413836), col = "black", lty = "dotted",
  lwd = 3)
legend("topright", legend = c("SP", "Resample"), fill = c("black",
  "grey"), bty = "n")
```

## Plausibilität



Totalvariabilität T in Einzelinterviews (gepunktet MW +/- St.Abw.)

```
options(scipen = 5)

dat3 <- data.frame(V1 = calc6$rowsum)

round(sort(with(dat3, V1/sum(V1))), digits = 6)

round(cumsum(sort(with(dat3, V1/sum(V1)))), digits = 6)

length(which(round(cumsum(sort(with(dat3, V1/sum(V1)))), digits = 6) <
  0.05)) # 281 interviews are on the left side below the lower 5 percent
```

## 7.1.13. Qualitative evaluation

Get tidy data for clustering. Write a csv and extract info to get an overview of the ranking.

```
getwd()

write.csv2(sort.calc6, "2sorted_ranking_calc6.csv", sep = ";")

sort(table(mergeddata$INTNUM1))
head(mergeddata, n = 5)

Intnum_count_each <- mergeddata[, -c(3:length(mergeddata))] %>%
  rowwise %>% mutate(time = list(rep(INTNUM1)))
# repeat the INTNUM1 based on its counts

head(Intnum_count_each, n = 5)

length(unique(Intnum_count_each$INTNUM1)) # 193 unique interviewers

Intnum_count_each2 <- Intnum_count_each %>% group_by(INTNUM1) %>%
  mutate(count = n()) # put in an occurrence variable

head(Intnum_count_each2, n = 5)

Intnum_count_each3 <- Intnum_count_each2[, -c(3)] %>% group_by(INTNUM1,
  count) %>% mutate(count = n()) # grouping and erase irrelevant var

rm(list = c("Intnum_count_each", "Intnum_count_each2"))

head(Intnum_count_each3, n = 5)

rowsum_by_idno_intnum3 <- cbind.data.frame(calc6, Intnum_count_each3[,
  3])

write.csv2(rowsum_by_idno_intnum3, "data1.csv")
getwd()

unique_int_rowsums <- rowsum_by_idno_intnum3[!duplicated(rowsum_by_idno_in
tnum3$INTNUM1),
]
write.csv2(unique_int_rowsums, "data2.csv")

unique_int_rowsums <- read.csv2("data2.csv", header = T)
unique_int_rowsums <- unique_int_rowsums[, -c(6:13)]
unique_int_rowsums2 <- unique_int_rowsums[, -c(1, 3)]
# evaluate which interviews rowsums are significantly lower in a cluster
```

## 7.2 Eigenständigkeitserklärung

Ich, Kevin Glock, versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 163 Abs.1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Bamberg, 30. März 2021

Unterschrift