

2023

FZDW TECHNICAL REPORTS [NO. 2]



FORSCHUNGSZENTRUM
DEMOGRAFISCHER
WANDEL (FZDW)

ENTWICKLUNG UND VERBREITUNG VON FORSCHUNGSDATEN-
MANAGEMENT AN FACHHOCHSCHULEN UND HOCHSCHULEN
FÜR ANGEWANDTE WISSENSCHAFTEN (EVER_FDM)
GENERIERTE VARIABLEN IM EVER_FDM DATENSATZ
ROBERT WERTH



Autor:

Robert Werth 

Zitationsempfehlung:

Werth, R. (2023). Generierte Variablen im EVER_FDM Datensatz. *FZDW Technical Reports*, 2. Frankfurt am Main, Forschungszentrum Demografischer Wandel (FZDW), Frankfurt University of Applied Sciences

Die Durchführung der EVER_FDM-Studie war eine große Gemeinschaftsleistung. Der Beitrag der folgenden Personen wird dabei dankend anerkannt:

Andreas Klocke
Arnela Balic
Sven Stadtmüller
Meike Backes
Lukas Magnin

Kontakt:

Frankfurt University of Applied Sciences
Forschungszentrum Demografischer Wandel (FZDW)
Nibelungenplatz 1
60318 Frankfurt am Main
E-Mail: info@fzdw.de
Web: www.fzdw.de



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).

EVER_FDM wurde:



Förderkennzeichen: 16DFH201

Inhaltsverzeichnis

- 1. Über diesen Bericht 4
- 2. Generierte Variablen 4

1. Über diesen Bericht

Im Rahmen der Datenedition wurden 57 neue Variablen generiert. Dabei handelt es sich um Codierungen und Recodierungen offener Angaben sowie um Hilfs- und Servicevariablen. Im vorliegenden Dokument sind diese Variablen und ihre Entstehungslogiken detailliert beschrieben und es werden die zur Erstellung verwendeten Variablen gelistet. Ggf. sind auch die verwendete Syntax sowie Anmerkungen zur Nutzung dokumentiert. Die Reihenfolge entspricht der Abfolge der Variablen im Datensatz. Die Syntaxen wurden in Stata (Version 14) erstellt.

2. Generierte Variablen

Variablenname: sy_misratio_g1	Ursprungsvariable(n): -
<p>Beschreibung: Gibt für jede befragte Person an, wie groß der prozentuale Anteil an der Gesamtzahl der Variablen ist, bei welchem kein bzw. kein gültiger Wert vorliegt.</p>	
<p>Anmerkungen: Diente ursprünglich dazu, die Fälle zu bereinigen, die ausschließlich fehlende Werte beinhalteten. Es ist demnach nur eine Annäherung an den Prozentsatz der nicht beantworteten inhaltlichen Fragen, da jeder fehlende Wert mit in die Berechnung einfließt (bspw. auch filterbedingt fehlende Werte). Offene Angaben fließen nicht in die Berechnung mit ein.</p>	
<p>Syntax: <pre> preserve keep sy_id ac_q03_01-comment_s drop *_s // String-Variablen dropen (fließen nicht in die Berechnung mit ein) *** Gesamtzahl der Variablen berechnen *** quietly: describe gen totalvar_g1 = r(k) - 1 // -1 ist die ID-Variable *** Anzahl der fehlenden Angaben berechnen *** egen misscount_g1 = rowmiss(_all) *** Missingquote berechnen *** gen sy_misratio_g1 = misscount_g1/totalvar keep sy_id misscount_g1 sy_misratio_g1 totalvar tempfile miss1 save `miss1' *** Informationen anhängen *** restore merge 1:1 sy_id using `miss1' assert _merge==1 _merge==3 drop _merge totalvar_g1 misscount_g1 order sy_misratio_g1, a(sy_complete) </pre> </p>	

Variablenname: sd_q02_g1	Ursprungsvariable(n): sd_q01, sd_q02
Beschreibung: Die Variable zeigt an, ob der/die Teilnehmende an der aktuellen Hochschule momentan Forschung betreibt oder dies in der Vergangenheit getan hat. Mit dieser Variablen kann also zwischen Forschenden (Zielgruppe) und Nicht-Forschenden unterschieden werden.	
Anmerkungen: -	
Syntax: gen sd_q02_g1 = . replace sd_q02_g1 = 1 if sd_q01==1 sd_q02==1 replace sd_q02_g1 = 0 if sd_q02==2 sd_q02==3 replace sd_q02_g1 = -966 if sd_q01==998 & sd_q02==998	

Variablenname: ac_q03_01_g1 - ac_q03_11_g1	Ursprungsvariable(n): ac_q03_01 - ac_q03_11, ac_q03_11_s
Beschreibung: Die Variablen enthalten die recodierten Versionen der Ursprungsvariablen zur Herkunft der Forschungsdaten. Dabei wurden die offenen Angaben bei der Nennung von „sonstige Quellen, und zwar“ (ac_q03_11_s) geprüft und diese ggf. den bereits vorhandenen Kategorien nachträglich zugeordnet.	
<u>Codierkonventionen:</u> - Georeferenzdaten → Amtliche Statistik und Referenzdaten (ac_q03_08_g1) - (Quell-)Code → Textdokumente (ac_q03_10_g1)	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser, statt der Ursprungsvariablen. Aus Datenschutzgründen wurde die Ursprungsvariable ac_q03_11_s aus dem Datensatz entfernt.	
Syntax: -	

Variablenname: ac_q06_g1	Ursprungsvariable(n): ac_q06, ac_q06_s
Beschreibung: Die Variable enthält die in numerische Werte überführten Angaben (= Codierung) aus der offenen Abfrage zur Archivierungsdauer nach Projektende.	
<u>Codierkonventionen:</u> - Bei Angabe einer Löschung nach Projektende → 0 - Bei Angabe einer unbegrenzten Archivierungsdauer → 99 - Bei Angabe eines Zeitraums → Niedrigste genannte Zahl	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser, statt der Ursprungsvariable ac_q06. Aus Datenschutzgründen wurde die Ursprungsvariable ac_q06_s aus dem Datensatz entfernt.	
Syntax: -	

Variablenname: ac_q06_g2	Ursprungsvariable(n): ac_q06_g1
Beschreibung: Kategorisierung der numerischen Angabe zur Archivierungsdauer.	
Anmerkung: Aufgrund der besonderen Bedeutung des 10-Jahres-Kriteriums (empfohlene Archivierungsdauer laut Leitlinie 17 zur Sicherung guter wissenschaftlicher Praxis der DFG) wurde hierfür eine eigene Kategorie erstellt.	
Syntax: recode ac_q06_g1 (0 = 1 "gar nicht") /// (0.5/4 = 2 "weniger als 5 Jahre") /// (5/9 = 3 "5-9 Jahre") /// (10 = 4 "10 Jahre") /// (11/19 = 5 "11-19 Jahre") /// (20/29 = 6 "20-29 Jahre") /// (30/98 = 7 "30-98 Jahre") /// (99 = 8 "99 Jahre oder länger"), gen(ac_q06_g2)	

Variablenname: ac_q12_g1	Ursprungsvariable(n): ac_q12_s
Beschreibung: Die Variable enthält die codierten offenen Angaben zu weiteren Hinderungsgründen bei der Veröffentlichung von Forschungsdaten. Teilweise wurden auch Gründe genannt, die bereits in der geschlossenen Abfrage gelistet waren.	
Anmerkung: Aus Datenschutzgründen wurde die Ursprungsvariable aus dem Datensatz entfernt.	
Syntax: -	

Variablenname: ac_q13_07_g1	Ursprungsvariable(n): ac_q13_07, ac_q14_s
Beschreibung: Die Variable enthält die recodierten Angaben zu anderen Richtlinien oder Standards im aktuellen bzw. letzten Forschungsprojekt. Es wurden lediglich Werte umcodiert, bei denen in der offenen Nennung anderer Richtlinien oder Standards (ac_q14_s) ungültige Angaben gemacht wurden.	
Codierkonventionen:	
<ul style="list-style-type: none"> - „Keine Ahnung“ / „Weiß nicht“ → -999 - Satz- oder Sonderzeichen → -966 - „Keine“ → 0 	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser, statt der Ursprungsvariablen. Aus Datenschutzgründen wurde die Ursprungsvariable ac_q14_s aus dem Datensatz entfernt.	
Syntax: -	

Variablenname: ac_q16_01_g1 - ac_q16_08_g1	Ursprungsvariable(n): ac_q16_01 - ac_q16_08, ac_q16_08_s
Beschreibung: Die Variable enthält die recodierten Angaben zu sonstigen externen Kooperationspartnern. Dabei wurden die offenen Angaben bei der Nennung von „Sonstige, und zwar“ (ac_q16_08_s) geprüft und ggf. den bereits vorhandenen Kategorien nachträglich zugeordnet.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser, statt der Ursprungsvariablen. Aus Datenschutzgründen wurde die Ursprungsvariable ac_q16_08_s aus dem Datensatz entfernt.	
Syntax: -	

Variablenamen: ne_q17_g1 - ne_q17_g12	Ursprungsvariable(n): ne_q17_s
<p>Beschreibung: Die Variablen enthalten die codierten Angaben zu den Bedarfen hinsichtlich des Umgangs mit Forschungsdaten.</p> <p>Codierkonventionen:</p> <ul style="list-style-type: none"> - „Ressourcen“ / „Mittel“ → Mehr Geld (ne_q17_g2) - „Kapazität“ → Mehr Zeit (ne_q17_g3) - „Server“ → Spezifische Hardware (ne_q17_g6) - „Informationen“ → Beratungsangebote (ne_q17_g8) - „Wissen“ / „Know-How“ → Schulungsangebote (ne_q17_g9) 	
<p>Anmerkungen: Häufig wurden mehrere Bedarfe genannt. Daher wurde eine Mehrfachauswahl codiert. Aus Datenschutzgründen wurde die Ursprungsvariable aus dem Datensatz entfernt.</p>	
<p>Syntax: -</p>	

Variablenamen: need_01_g1 - need_12_g1	Ursprungsvariable(n): ne_q18_01 - ne_q18_12, ne_q19_01 - ne_q19_12
<p>Beschreibung: Bedarfsindikator mit vier Ausprägungen:</p> <p>Nimmt den Wert „1“ an, wenn ein Thema als „unwichtig“ oder „etwas wichtig“ eingestuft wurde und gleichzeitig „keine Kenntnisse“ oder „geringe Kenntnisse“ angegeben wurden.</p> <p>Nimmt den Wert „2“ an, wenn ein Thema als „unwichtig“ oder „etwas wichtig“ eingestuft wurde und gleichzeitig „ziemlich gute Kenntnisse“ oder „sehr gute Kenntnisse“ angegeben wurden.</p> <p>Nimmt den Wert „3“ an, wenn ein Thema als „ziemlich wichtig“ oder „sehr wichtig“ eingestuft wurde und gleichzeitig „keine Kenntnisse“ oder „geringe Kenntnisse“ angegeben wurden.</p> <p>Nimmt den Wert „4“ an, wenn ein Thema als „ziemlich wichtig“ oder „sehr wichtig“ eingestuft wurde und gleichzeitig „ziemlich gute Kenntnisse“ oder „sehr gute Kenntnisse“ angegeben wurden.</p>	
<p>Anmerkungen: -</p>	
<p>Syntax:</p> <pre>foreach n in 01 02 03 04 05 06 07 08 09 10 11 12 { gen need_`n'_g1 = . replace need_`n'_g1 = 1 if (ne_q18_`n'==1 ne_q18_`n'==2) & (ne_q19_`n'==1 ne_q19_`n'==2) replace need_`n'_g1 = 2 if (ne_q18_`n'==1 ne_q18_`n'==2) & (ne_q19_`n'==3 ne_q19_`n'==4) replace need_`n'_g1 = 3 if (ne_q18_`n'==3 ne_q18_`n'==4) & (ne_q19_`n'==1 ne_q19_`n'==2) replace need_`n'_g1 = 4 if (ne_q18_`n'==3 ne_q18_`n'==4) & (ne_q19_`n'==3 ne_q19_`n'==4) replace need_`n'_g1 = -966 if missing(ne_q18_`n') missing(ne_q19_`n') }</pre>	

Variablenname: ne_q20_11_g1 - ne_q20_15_g1	Ursprungsvariable(n): ne_q20_11, ne_q21_s
Beschreibung: Die Variablen enthalten die codierten Angaben zur potenziellen Nutzung anderer Software. Dabei wurde jeweils der codierten Variable der Wert der Variable ne_q20_11 zugewiesen. Wurden mehrere Softwarelösungen genannt, so erhielten alle den Wert der Variable ne_q20_11. Codierkonvention: - „(Daten-)Analysesoftware“ → Software zur Auswertung quant. Daten (ne_q20_12_g1)	
Anmerkungen: Häufig wurden mehrere Softwarelösungen genannt. Daher wurde eine Mehrfachauswahl codiert. Aus Datenschutzgründen wurde die Ursprungsvariable ne_q21_s aus dem Datensatz entfernt.	
Syntax: -	

Variablenname: sd_q29_g1	Ursprungsvariable(n): sd_q29
Beschreibung: Die Variable enthält Informationen zur Statusgruppe der Befragten. Die Statusgruppen der Laboringenieure (n = 22), Lehrbeauftragten (n = 17) und der studentischen bzw. wissenschaftlichen Hilfskräfte (n = 15) waren dabei kaum vertreten, was eine Re-Identifikation der Befragten begünstigen würde. Daher wurden diese drei Statusgruppen der Kategorie „anderer Status“ hinzugefügt.	
Anmerkung: Aus Datenschutzgründen wurde die Ursprungsvariable aus dem Datensatz entfernt.	
Syntax: clonevar sd_q29_g1 = sd_q29 recode sd_q29_g1 (5/7=8)	

Variablenname: sd_q31_g1	Ursprungsvariable(n): sd_q31
Beschreibung: Aus Datenschutzgründen wurde die Angabe eines diversen Geschlechts der Kategorie „keine Angabe (Antwortkategorie)“ hinzugefügt. Insgesamt betrifft dies acht Fälle.	
Anmerkung: Aus Datenschutzgründen wurde die Ursprungsvariable aus dem Datensatz entfernt.	
Syntax: clonevar sd_q31_g1 = sd_q31 mvdecode sd_q31_g1, mv(3=-997)	

Variablenname: comment_g1	Ursprungsvariable(n): comment_s
Beschreibung: Die Variable enthält die codierten offenen Angaben zu weiteren Anmerkungen zum Fragebogen oder zum Thema Forschungsdatenmanagement.	
Anmerkung: Aus Datenschutzgründen wurde die Ursprungsvariable aus dem Datensatz entfernt.	
Syntax: -	