

2021

00

Statistische
Grundlagen und
Übersichten

Neuchâtel 2023

Forschung und Entwicklung 2021

Stichprobenplan, Gewichtung und Schätzverfahren

Methodenberichte der Sektion Statistische Methoden

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zugrunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicherzustellen, und andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar

Forschung und Entwicklung 2021

Stichprobenplan, Gewichtung und Schätzverfahren

Redaktion Jann Potterat, BFS;
Daniel Assoulin, BFS

Inhalt Sektion Statistische Methoden, BFS

Herausgeber Bundesamt für Statistik (BFS)

Neuchâtel 2023

Herausgeber: Bundesamt für Statistik (BFS)

Auskunft: Jann Potterat, Tel. +41 58 463 65 90
jann.potterat@bfs.admin.ch
Daniel Assoulin, Tel. +41 58 463 67 51
daniel.assoulin@bfs.admin.ch

Redaktion: Jann Potterat, BFS
Daniel Assoulin, BFS

Inhalt: Sektion Statistische Methoden, BFS

Reihe: Statistik der Schweiz

Themenbereich: 00 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Layout: Publishing und Diffusion PUB, BFS
Dieses Dokument wurde automatisch aus einer Datenbank
heraus erzeugt. Es entspricht daher nicht dem üblichen
typografischen Standard der BFS-Publikationen

Online: www.statistik.ch

Print: www.statistik.ch
Bundesamt für Statistik, CH-2010 Neuchâtel,
order@bfs.admin.ch, Tel. +41 58 463 60 60
Druck in der Schweiz

Copyright: BFS, Neuchâtel 2023
Wiedergabe unter Angabe der Quelle
für nichtkommerzielle Nutzung gestattet

BFS-Nummer: 338-0086

ISBN: 978-3-303-00727-3

Inhaltsverzeichnis

Vorwort und Zusammenfassung	5
1 Einleitung	7
2 Grundgesamtheit	7
2.1 Ausweitung Grundgesamtheit aufgrund Unternehmensgruppen	8
2.2 Bestimmung einflussreicher Beobachtungen	10
3 Stichprobenplan	11
3.1 Ansätze zur Reduktion der Stichprobengrösse	11
3.2 Analyse Antworten Vorerhebungen	12
3.3 Modellierung Wahrscheinlichkeiten	13
3.4 Definitiver Stichprobenplan	14
3.5 Ziehung der Stichprobe im SVS	16
4 Analyse Screening: Schätzung Anzahl Unternehmen mit Forschungstätigkeit	17
4.1 Grobe Schätzung basierend auf Erfahrungswerten der Vorerhebungen 2015-2019	17
4.2 Schätzung pro NOGA x Grössenklassen via logistische Regression	18
5 Analyse Haupterhebung	19
6 Behandlung Merkmalsausfälle	20
7 Gewichtung	22
7.1 Vergleich dreier Methoden zur Antwortkorrektur	22
7.1.1 Detaillierter Beschrieb der Methoden zur Antwortkorrektur	22
7.1.2 Statistiken zur Evaluation der Antwortkorrektur	23
7.1.3 Auswirkungen der Modelle der Antwortkorrektur auf die Schätzungen	24
7.2 Behandlung einflussreiche Beobachtungen und Gruppen	25
7.3 Resultierende Gewichtung	26
8 Schätzung der Resultate mit Varianzschätzung	27
8.1 Varianzschätzung ohne Berücksichtigung Einsetzung	27
8.2 Varianzschätzung: Einsetzungsvarianz «last observation carry-forward»	27
9 Schlussfolgerungen und Ausblick	30
Literatur	31
Anhang	32
A Fragebogen der F+E 2021	32
Methodenberichte der Sektion Statistische Methoden des BFS	37

Vorwort

Forschung und Entwicklung (F+E) spielt eine wichtige Rolle im Innovationsprozess und stellt eine zentrale Antriebskraft im wirtschaftlichen Wachstum dar. Die Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft wird vom Bundesamt für Statistik (BFS) mit Unterstützung des Dachverbands der Schweizer Wirtschaft «economiesuisse» durchgeführt. Neben den Resultaten der F+E der Privatwirtschaft publiziert das BFS auch die F+E-Ausgaben des Hochschulsektors und des Bundes.

Der Sektion Statistische Methoden METH des BFS wurde die Überarbeitung der statistischen Methoden der F+E-Erhebung 2021 anvertraut. Um die Belastung der Befragten und die Kosten der Erhebung zu verringern, mussten die Methoden, die seit 2012 grösstenteils unverändert geblieben sind, revidiert beziehungsweise optimiert werden.

Die Autoren bedanken sich bei den Sektionen Monetäre Unternehmensstatistik (MON) und Wirtschaftsstruktur und -analysen (WSA), welche für die Durchführung und Auswertung der Erhebung zuständig sind, insbesondere bei M. Ferrari, L. Milhaud Duvernay, S. Plaza Chardon und P. Sollberger, für die sehr gute und angenehme Zusammenarbeit. Ein weiterer Dank geht an J.-P. Renfer und D. Kilchmann von METH für die methodische Unterstützung und die aufmerksame Durchsicht des Berichts sowie an L. Hoessly (METH) für seine Analysen, die ebenfalls in diesen Bericht einfließen.

Zusammenfassung

Die Erhebung des Bundesamtes für Statistik zur Forschung und Entwicklung in der schweizerischen Privatwirtschaft (F+E) wird seit 2015 jedes zweite Jahr durchgeführt. Sie besteht aus zwei Phasen: einem Screening zur Identifikation der Unternehmen mit F+E-Ausgaben und der Haupterhebung zur Ermittlung der eigentlichen Ausgaben.

Dieser Methodenbericht beschreibt die wichtigsten methodischen Aspekte der beiden Phasen. Ein Hauptaugenmerk wird dabei auf die Neukonzeption des Screenings gelegt, wobei für die Erhebung 2021 die bisherige Vollerhebung durch eine ressourcensparende Alternative basierend auf einer Stichprobe abgelöst wurde.

Aus der Grundgesamtheit (ca. 22 000 vorwiegend grössere Unternehmen in klar definierten Branchen) wurden dabei in der ersten Phase rund 14 000 Unternehmen gezogen. Von den gezogenen Unternehmen haben über 90% geantwortet, wobei aber nur rund jedes fünfte Unternehmen angab, F+E-Ausgaben zu tätigen. In der zweiten Phase haben etwas mehr als 60% dieser Unternehmen (ca. 1650) Angaben zu den effektiven Ausgaben mitgeteilt.

Die Gewichtung wurde an die Tatsache angepasst, dass es sich bei der ersten Phase der Erhebung nicht mehr um eine Vollerhebung handelte. Es wurden verschiedene Gewichtungsvarianten für die Korrektur von Antwortausfällen getestet, darunter eine beruhend auf einer logistischen Regression. Aus Gründen der Vergleichbarkeit wurde die diesbezügliche Methodik schlussendlich aber nicht verändert.

Im Rahmen der F+E gibt es erfahrungsgemäss einige wenige Unternehmen mit aussergewöhnlich hohen Ausgaben und entsprechend starkem Einfluss auf die Resultate. Um die Vergleichbarkeit der Resultate über die Zeit zu gewährleisten, wurde eine systematische Identifikation dieser Unternehmen basierend auf vorangehenden Erhebungswerten eingeführt. Dies erlaubt es, diese Unternehmen sowohl im Stichprobenplan, als auch in der Datenerhebung und der Hochrechnung gesondert zu behandeln.

Am Ende des Berichts wird die Varianzschätzung beschrieben, welche dazu dient, die Genauigkeit der im Rahmen der F+E erstellten Schätzungen beurteilen zu können und so auch hilft, die Resultate oder Entwicklungen besser interpretieren zu können.

1 Einleitung

Die Erhebung zu Forschung und Entwicklung (F+E) wurde von 1992 bis 2012 alle vier und ab 2015 jedes zweite Jahr durchgeführt. Die Sektionen MON und WSA im Bundesamt für Statistik sind zuständig für die Durchführung bzw. Auswertung der Erhebung.

Die F+E-Erhebung besteht seit 2012 aus zwei Phasen:

1. Ein Screening von Unternehmen, um festzustellen, ob ein Unternehmen Forschungs- und Entwicklungsaktivitäten hat. Die entsprechende Frage lautet wie folgt (siehe auch Abbildung 1):
«Hatte Ihr Unternehmen im Jahr 20xx Forschungs- und Entwicklungsaufwendungen in einer oder mehreren der unten aufgeführten Kategorien getätigt oder geplant, solche zu tätigen?»
 - Für eigene Zwecke oder im Auftrag Dritter durchgeführte F+E (Intramuros F+E).
 - F+E an Dritte vergeben (Extramuros F+E).
 - Für eigene Zwecke oder im Auftrag Dritter durchgeführte F+E in Ihren Arbeitsstätten und/oder Zweigunternehmen im Ausland.»
2. Ein detaillierter Fragebogen zu den F+E-Aktivitäten wird an Unternehmen verschickt, die das Screening mit Ja beantwortet haben. Dieser enthält dann z.B. Fragen zu Kosten, Personal, Art, Ziel und Technologiebereich (siehe Abbildungen 2 bis 5).

Das Screening bestand bis 2019 aus einer Vollerhebung der Branche «72 Forschung und Entwicklung» ab 1 Beschäftigten, plus einer Selektion aller Unternehmen in gewissen Branchen ab 10 Beschäftigten. Dies führte in den letzten Erhebungen zu jeweils rund 21 500 Unternehmen. Für die Erhebung 2021 war das Ziel, nicht mehr als etwa 14 000 Unternehmen im Screening zu befragen, um einerseits die Belastung der Befragten und die Kosten der Erhebung zu verringern und andererseits auf diese Weise die verfügbaren Ressourcen zu optimieren.

Entsprechend wurde die Vollerhebung (Screening) der ersten Phase durch eine Stichprobenerhebung ersetzt. In diesem Bericht werden der entsprechende Stichprobenplan sowie die Überlegungen, welche dazu führten, beschrieben. Ebenso wird das Gewichtungs- und Hochrechnungsverfahren vorgestellt, welches im Laufe der Zeit optimiert worden ist.

2 Grundgesamtheit

Die Grundgesamtheit der Unternehmen, die den Rahmen des Screening 2021 bildet, wird im Stichprobenrahmen gemäss Betriebs- und Unternehmensregister (BUR) vom 7. Dezember 2020 nach den folgenden Kriterien abgegrenzt:

1. Privatsektor (Betriebstyp: E13) plus einige öffentliche Unternehmen (Betriebstyp: E27) grosser Unternehmensgruppen.
2. NOGA08¹: Sekundärer und tertiärer Sektor gemäss der Liste der NOGA08-Codes in Tabelle 1. Zu beachten ist, dass nur ein Teil der NOGA08-Zweige berücksichtigt wird. Zusätzlich hinzugefügt werden wie bisher Unternehmen ausserhalb der relevanten Branchen, die jedoch eine Arbeitsstätte in der NOGA 72 (Forschung) haben. Dies sind rund 10 Unternehmen, welche im Prinzip der Branche 9 zugeordnet werden könnten. Um sie jedoch zu unterscheiden, werden diese als 9.9 bezeichnet.

¹Wirtschaftsaktivität «Nomenclature générale des activités économiques» (gemäss Allgemeine Systematik der Wirtschaftszweige: www.noga.bfs.admin.ch).

3. $Betot \geq 10$, wobei $Betot$ die Gesamtzahl der Beschäftigten im Unternehmen ist. Ausnahme für NOGA08-Branche Nr. 9 (NOGA08=72), wo alle Unternehmen mit $Betot > 0$ ausgewählt werden.

Tabelle 1 Liste der NOGA-Branchen (gemäss NOGA08) zur Bestimmung des Rahmens für die Screening-Population.

Index für aggregierte NOGA08 (bra_noga)	Name der Aggregierung	enthaltene NOGA08
1	Lebensmittel, Tabak	10, 11, 12 (12 ab 2017 für Publikation in «Andere»)
2	Chemie	19, 20, 22
3	Apotheke	21
4	Metall	24, 25
5	Maschinen	27, 28, 29, 30 (ohne 30.3)
6	High-Tech-Instrumente	26.5, 26.7, 30.3
7	IKT*-Fertigung	26 (ohne 26.5, 26.7)
8	IKT*-Dienstleistungen	46.5, 58.2, 61, 62, 63.1, 95.1
9 (9.9)	Forschung und Entwicklung	72
10.1	Andere Zweige 1	69, 70, 71, 73
10.2	Andere Zweige 2	5-9, 13-18, 23, 31, 32, 35-42, 53, 58 (ohne 58.2), 59-60, 75
Für die Publikation werden Branchen ≥ 10.1 zusammengefasst (zu Andere «99»).		
*für «Informations- und Kommunikationstechnologien»		

Die Screening-Grundgesamtheit – definiert nach den Kriterien 1, 2 und 3 – setzt sich im 2021 aus 22 077 Unternehmen zusammen (2019: 21 572, 2017: 21 502). Dies ist der (Netto-) Rahmen, wobei wie gesagt keine Vollerhebung mehr stattfindet, sondern eine Stichprobe gezogen werden muss.

Dieser Rahmen wurde für das Screening analog zu den Vorjahren wie im nächsten Abschnitt beschrieben, leicht erweitert, was zum Brutorahmen führt.

2.1 Ausweitung Grundgesamtheit aufgrund Unternehmensgruppen

In der Schweiz ansässige Unternehmen können zu sogenannten Unternehmensgruppen oder Konzernen gehören, wobei der Sitz der Gruppe in der Schweiz oder im Ausland sein kann. Das für die Gruppendifinition wichtige Kontrollverhältnis zwischen zwei Unternehmen bemisst sich dabei an der Kapitalbeteiligung – d.h. die kontrollierende Einheit besitzt mehr als 50% der Stimmrechte – oder der Möglichkeit, die allgemeine Unternehmenspolitik zu bestimmen.

Die Liste, welche Gruppen es gibt und wer sich darin befindet, wird unter Berücksichtigung mehrerer Quellen jährlich im BFS aktualisiert. Eine dieser Quellen war ursprünglich jene von Dun&Bradstreet (D&B), daher werden diese Gruppen im Folgenden als D&B-Gruppen bezeichnet. Für eine F+E-Erhebung wird die Liste mit Stand per Ende Vorjahr verwendet, also z.B. Ende 2020 für die Erhebung 2021.

Diese Gruppen werden bei der Erhebung häufig auf eine der folgenden zwei Arten antworten:

- Sammelantworten: 1 Antwort umfasst die F+E-Ausgaben der ganzen Gruppe. Diese Ausgaben werden mit einem Schlüssel auf alle Mitglieder verteilt.
- Antwort auf Niveau Unternehmen (UNT) von den Unternehmen, die effektiv die Forschung für die Gruppe betreiben. Hier gibt es keine Verteilung und für den Rest der Gruppe wird dann Ausgaben=0 gesetzt.

Aus früheren F+E-Erhebungen ist bekannt, dass Unternehmen einer Gruppe im Screening oder der Erhebung auf diese Weise auch für Unternehmen derselben Gruppe antworten, die sich ausserhalb der Netto-Population befinden (andere NOGA, oder weniger als 10 Beschäftigte).

Die Netto-Grundgesamtheit der F+E wird vergrössert, um diesem Aspekt Rechnung zu tragen, was zur Brutto-Grundgesamtheit führt. So wurde gemäss der Liste aller D&B-Gruppen eine Teilmenge, bestehend aus 41 Gruppen bestimmt, für welche aufgrund früherer Erhebung hohe F+E-Ausgaben vermutet werden. Diese Gruppen umfassen im 2021 insgesamt 684 Unternehmen, davon sind 356 in der oben beschriebenen Netto-Grundgesamtheit enthalten, die restlichen 328 Unternehmen wären es eigentlich nicht, wurden aber für das Screening in die Brutto-Grundgesamtheit integriert. Mit dieser schon seit längerem angewandten Methode will man verhindern, dass man hohe Forschungsausgaben von Unternehmen ausserhalb der Netto-Grundgesamtheit «verpasst».

Anmerkung: Im BFS gibt es noch eine andere Definition von Unternehmensgruppen, nämlich jene des Profiling. Das Profiling basiert auf partnerschaftlicher Zusammenarbeit zwischen Unternehmen und dem BFS. Diese Partnerschaften ermöglichen die vereinfachte Datenerhebung zwecks Aktualisierung des BUR und eine optimale Unterstützung und Koordination für Statistiken des BFS.

Die Unternehmen von Profiling-Gruppen, welche u.a. quartalsweise ihre Beschäftigtendaten ans BFS liefern, können sich von den eben beschriebenen D&B-Unternehmensgruppen unterscheiden. Zudem gibt es auch grosse Gruppen gemäss D&B, die nicht im Profiling sind. Damit die in der F+E verwendeten Gruppen alle möglichen Gruppenmitglieder in einer Art Maximalliste umfassen, hat METH die beiden Listen miteinander «zusammengeführt». Diese Liste wird dann beim Screening dem vermuteten Verantwortlichen der entsprechenden Gruppe geschickt, welcher die Mitglieder noch berichtigen kann.

Eine zusätzliche Erweiterung der Netto-Grundgesamtheit gibt es im Rahmen der F+E 2021, um die Option einer Panelstudie bzw. Spezialauswertung offen zu halten. Bei einem Panelansatz wäre bei der Gewichtung / Hochrechnung eine genauere Analyse der Entwicklung zwischen zwei Erhebungen möglich, wobei hingegen Populationsveränderungen nicht berücksichtigt würden. So wurden im 2021 zusätzlich 95 Unternehmen zum Brutorahmen hinzugefügt, welche in der Erhebung 2019 F+E-Ausgaben hatten und immer noch aktiv sind, aber nicht mehr zur neuen Population gehören würden. Diese werden direkt befragt, jedoch ist es geplant, deren Antwort nur in der allfälligen Spezialauswertung zu berücksichtigen.

Schliesslich ergibt sich im 2021 ein Brutorahmen von 22 500 Unternehmen. Die Tabelle 2 zeigt eine Übersicht der Einheiten nach NOGA-Branche (bra_noga08 bzw. gr_noga08) und Grössenklassen (grkl2).

Tabelle 2 Liste der Grundgesamtheit nach NOGA und Grössenklasse.

NOGA-Branche		Größenklasse (grkl2)											
bra_noga08	gr_noga08	1 bis 9		10 bis 49		50 bis 99		100+		«Big-Units»		Total	
		N	betot	N	betot	N	betot	N	betot	N	betot	N	betot
1	1012	2	11	1 221	24 773	146	10 015	153	44 728	7	12 767	1 529	92 294
2	1922	1	9	378	8 629	121	8 508	102	25 344	9	8 117	611	50 607
3	2100	.	.	58	1 374	26	1 827	55	24 330	2	20 602	141	48 133
4	2425	3	16	1 582	31 982	216	14 988	160	32 728	7	3 772	1 968	83 486
5	2730	9	58	853	19 278	232	16 391	247	71 891	7	7 189	1 348	114 807
6	2657	2	9	323	7 182	83	5 905	126	56 764	9	5 515	543	75 375
7	2600	2	17	190	4 395	52	3 540	78	29 282	3	4 136	325	41 370
8	6162	27	157	1 864	37 153	218	14 812	192	66 366	7	1 962	2 308	120 450
9	7200	1 685	3 713	206	4 093	32	2 229	33	6 129	9	7 135	1 965	23 299
9.9	7201	1	2	2	35	1	77	5	6 350	.	.	9	6 464
10.1	6973	49	306	5 075	93 673	406	27 247	245	66 679	2	716	5 777	188 621
10.2	8000	32	159	4 551	93 319	665	45 417	429	163 549	.	.	5 677	302 444
12	9000	76	348	91	2 346	49	3 495	80	39 859	3	769	299	46 817
Total		1 889	4 805	16 394	328 232	2 247	154 451	1 905	633 999	65	72 680	22 500	1 194 167

Die Branche 12 enthält die Unternehmen, welche sich nicht in den Branchen 1-10 befinden und nur aufgrund einer Gruppenzugehörigkeit in die Grundgesamtheit aufgenommen werden. Aus demselben Grund befinden sich auch einige kleine Unternehmen – mit weniger als 10 Be-

schäftigten – und ausserhalb der Branche 9 (Forschung und Entwicklung) in der Grundgesamtheit.

Die Grössenklasse «Big-Units» enthält die Unternehmen, die in Vergangenheit sehr grosse Ausgaben hatten (Erklärung folgt im Abschnitt 2.2) und hier unabhängig ihrer Grössenklasse separat dargestellt werden. Die Tabelle 3 zeigt die Übersicht des Brutto- und des Nettorahmens.

Tabelle 3 Übersicht Brutto- und Nettorahmen

Rahmen FE21	Anzahl UNT
Nettopopulation	22 077 *
davon in Gruppen	356
Hinzufügen UNT ausserhalb Population:	
Gruppenmitglieder ohne Vorerhebungswerten	325
Gruppenmitglieder mit Vorerhebungswerten	3 *
Zusätzliche Unternehmen Panelstudie	95 *
Bruttopopulation	22 500
*benutzt für Modellierung Wahrscheinlichkeiten (vgl. Abschnitt 3.2)	22 175

2.2 Bestimmung einflussreicher Beobachtungen

Eine Analyse der F+E-Ausgaben 2015 und 2017 hatte für die F+E 2019 zum Ziel, jene Unternehmen mit hohen Ausgaben in den letzten Erhebungen bei der nächsten Erhebung unabhängig von der Anzahl Beschäftigten in eine separate Schicht zu legen, bei der Erhebung speziell zu verfolgen und im Falle von Antwortausfällen einzusetzen. Diese Einheiten wurden bestimmt, indem die Ausgaben folgender drei Hauptvariablen analysiert wurden:

- B0245 (F+E Intramuros Ausgaben)
- G0295 (F+E Extramuros Ausgaben)
- J0205 (F+E Ausgaben im Ausland)

Für jedes Unternehmen werden zu zwei Zeitpunkten 2015 und 2017 (sowie für den Mittelwert von 2015 und 2017) der geschätzte Anteil seiner F+E-Ausgaben für alle drei Variablen in ihrer NOGA berechnet. Falls ein Unternehmen nur in einem der beiden Jahre Ausgaben geliefert bzw. geantwortet hat, wurde dieser Wert sowohl fürs entsprechende Jahr wie auch für den Mittelwert gesetzt. Bei der Berechnung des Mittelwerts mit einem fehlenden Wert, hätte man alternativ fürs fehlende Jahr auch Missing setzen können, doch schlussendlich wollte man die Einheiten nicht von der Prozedur ausschliessen.

Als grosse (Einzel-)Unternehmen wurden jene bestimmt, deren Ausgabenanteil in der Branche für eine der neun Ausprägungen (3 Variablen x 3 Jahre) über 5% bzw. 10% liegt. Dabei wurde beim Mittelwert das 5%- und bei den restlichen das 10%-Kriterium angewandt, wie es folgende Tabelle zeigt:

	2015	2017	Mittel
B0245	10%	10%	5%
G0295	10%	10%	5%
J0205	10%	10%	5%

Für die F+E 2019 umfasste die Liste der Einheiten mit hohen Ausgaben, welche folgend als «Big-Units» bezeichnet werden, basierend auf obiger Analyse, 61 Unternehmen. Diese Einheiten sollten auch im Rahmen der F+E 2019 geflaggt werden. Aufgrund von möglichen Mutationen (Inaktivierungen, Übernahmen, Fusionen) waren nur noch 56 Unternehmen aktiv. Die Liste wurde also kleiner, auch wenn bei den Fusionen der «Big-Unit-Flag» des Vorgängers an den Nachfolger übergeben wurde.

Die Idee ist es diese Liste bei jeder folgenden Erhebung zu aktualisieren. So wurde für F+E 2021 die eben beschriebene Analyse bzw. Prozedur wiederholt, neu jedoch als Input mit den Ausgaben der Erhebungen von 2017 und 2019 (und dem Mittelwert der beiden Jahre). Dies ergab neu eine Liste von 65 «Big-Units», die in Tabelle 2 einer separaten Schicht zugeordnet wurden. Falls ein ehemals als gross bezeichnetes Unternehmen in zwei aufeinanderfolgenden Erhebungen nicht mehr antwortet, wird es für die nächste Erhebung aus der Liste der «Big-Units» gestrichen, also nicht mehr speziell behandelt. Dies könnte natürlich potentiell problematisch sein und bei einem erneuten Antwortausfall, der nicht speziell behandelt wird, zu einer Unterschätzung der Resultate führen, was aber in Kauf genommen wird (s. Abschnitt 7.2).

3 Stichprobenplan

3.1 Ansätze zur Reduktion der Stichprobengrösse

Im Frühjahr 2021 wurde mitgeteilt, dass für das Screening der F+E-Erhebung die Kosten verringert werden müssen. Um zu beurteilen wie eine mögliche Reduzierung aussehen könnte, wurden mehrere Varianten diskutiert, welche sich aus der Tabelle 4 herleiten lassen. Die Unternehmen im Rahmen sollten in drei Kategorien eingeteilt werden: 1) neue Unternehmen, 2) solche die eher nein antworten, also keine Forschung betreiben und 3) solche die eher ja antworten, also Forschung betreiben. Das Kriterium, ob ein Unternehmen eher ja oder nein antwortet, kann aus den Vorerhebungen eruiert werden, wobei die Methode noch erläutert wird.

Nach der Zuordnung der Unternehmen zu den Kategorien muss beurteilt werden, wie diese Kategorien behandelt werden können. Aus der Tabelle ist ersichtlich, dass es im Prinzip drei Möglichkeiten (a-c) gibt. Beispielsweise a): man übernimmt eine vorhandene Screening-Antwort des Unternehmens aus der Vorerhebung, ohne es neu zu befragen und falls es mit ja geantwortet hat, würde es direkt den detaillierten Fragebogen erhalten. Die anderen Möglichkeiten sind: Aus allen Einheiten entweder eine Teil-Stichprobe zu ziehen (b) oder wie bisher direkt alle zu befragen (c). Da die Anwendung von Letzterem auf die Unternehmen der Kategorie 2 nicht zu einer Reduktion des Stichprobenumfangs beitragen würde, wird Option c für die Kategorie 2 nicht in Betracht gezogen. Ebenso kann man bei neuen Unternehmen keinen Vorerhebungswert übernehmen, also gibt es keine Option 1a.

Tabelle 4 Ideen Screening

Typ	Screening 21		
	a) kein, also Antwort aus Vorerheb. übernehmen	b) Teil-Stichprobe ziehen	c) voll erheben
1: Neue Unt		1b	1c
2: Ausgaben=nein	2a	2b	
3: Ausgaben=ja	3a	3b	3c

Jede dieser Möglichkeiten hat Vor- und Nachteile. Die Varianten, welche auf der Annahme «gleiches Verhalten wie früher» beruhten, hätten insbesondere den Vorteil, dass sie einfach und schnell umzusetzen gewesen wären, was in der damaligen Situation attraktiv erschien, da eigentlich keine Zeit für eine Art Methodenrevision geplant war und man kurzfristig von den neuen Tatsachen etwas überrascht wurde.

Die Analysen konnten die Annahmen für ein solches Vorgehen aber nicht zufriedenstellend stützen, sondern wiesen darauf hin, dass man sich mit den «Ausschlussmethoden» einem im Grunde unnötigen Biasrisiko aussetzte.

Aus diesem Grund wurde empfohlen einen Stichprobenplan für die erste Phase zu erarbeiten, welcher den Erfordernissen (Stichprobengrösse, Effizienz) möglichst Rechnung tragen konnte.

3.2 Analyse Antworten Vorerhebungen

Es war ziemlich naheliegend, den Stichprobenplan aufgrund der Informationen aus den letzten Erhebungen zu konstruieren. Hierzu wurden die letzten Erhebungen 2012, 2015, 2017 und 2019 bezüglich des Merkmals analysiert, ob ein Unternehmen Forschung betreibt oder nicht. Für jedes der vier Erhebungsjahre kommen als mögliche Codes x vier Ausprägungen in Frage.

Tabelle 5 Kodierung Variablen aus Vorerhebungen

x	Beschrieb
1	ja (mit Forschung)
2	nein (keine Forschung)
5	keine Antwort (ob Forschung ja/nein)
. (0)	nicht im Erhebungsrahmen*
* Der Code «.» wird im Plan (z.B. in Tabellen 9 und 10) durch «0» ersetzt	

Es gäbe also maximal insgesamt $x^t = 4^4 = 256$ Kombinationen (mit x : Anzahl Codes, t =Anzahl Jahre), welche die «F+E-Geschichte» eines Unternehmens repräsentieren. Die Codes aller vier Jahre werden in eine kategorielle Variable «rep_f_2012_2019» zusammengefasst (konkateniert), wobei die erste Position den Code des ersten Jahres (2012) und die letzte Position der Code des neuesten Jahres (2019) wiedergibt. Beispiel zum besseren Verständnis: «rep_f_2012_2019=1211» bedeutet für ein Unternehmen, dass es in den Jahren 2012, 2017 und 2019 angegeben hat, Forschung zu betreiben, im Jahr 2015 betrieb es jedoch keine.

Die Tabelle 6 zeigt die Analyse der Anzahl Fälle bei Verbindung aller 4 Erhebungen (FE12-19) via Unternehmensidentifikator (entid). Dabei findet man jeweils die Anzahl Unternehmen in allen Vorerhebungen (31 197) und die Einschränkung auf die noch aktiven im Rahmen der FE21 (22 175) gemäss Tabelle 3. In der linken Hälfte sind die Kategorien aggregiert, in der rechten Hälfte findet man die Aufschlüsselung einiger Kombinationen.

Tabelle 6 Analyse Antworten in den letzten vier Screenings 2012-2019

Merkmals	FE12-19	Rahmen FE21*
4x nein	7 250	6 660
3x nein /1x nicht im Rahmen	4 121	2 928
3x nein /1x nicht antwortend	995	845
3x nein /1x ja	992	934
falls Antwort: immer nein (1x-2x)	11 685	3 888
unterschiedl. Antwort (1x / 2x)	2 327	1 783
falls Antwort: immer ja (1x-2x)	1 331	667
3x ja /1x nein	503	479
3x ja/1x nicht antwortend	155	152
3x ja /1x nicht im Rahmen	252	214
4x ja	647	636
nicht beantwortet	939	251
Neue Einheiten		2 738
Total	31 197	22 175

* im BUR (von Dez 2020, Basis für F+E 2021) noch in Population aktiv

Beschrieb von rep_f_2012_2019

1 Ziffer: FE 12	1=F&E-Ausgaben=ja
2 Ziffer: FE 15	2=F&E-Ausgaben=nein
3 Ziffer: FE 17	. = Nicht im Rahmen
4 Ziffer: FE 19	5=fehlend/k.Antwort

Aufschlüsselung einiger Kombinationen:

rep_f_2012_2019	Anzahl	Anzahl
1115	49	48
1151	36	35
1511	60	59
5111	10	10
1111	647	636
1112	139	126
1121	78	76
1211	88	84
2111	198	193
2225	285	198
2252	283	261
2522	360	324
5222	67	62
1222	250	229
2122	353	331
2212	186	177
2221	203	197
2222	7 250	6 660
.111	216	205
1.11	3	3
11.1	3	3
111.	30	3
.222	3 076	2 657
2.22	136	95
22.2	85	63
222.	824	113
Total		4 707

Wie ersichtlich hat es vom oben beschriebenen Fall: «rep_f_2012_2019=1211» genau 88 Unternehmen, die in den Jahren 2012, 2017 und 2019 angegeben haben, dass sie Forschung betreiben, im Jahr 2015 betreiben sie jedoch keine.

Damit unverfälschte Resultate schätzbar sind, muss gemäss Theorie (z.B. [Särndal, C.E., et al., 1992](#)) jedes der Unternehmen eine Ziehungswahrscheinlichkeit grösser 0 haben. Da zudem die Antwort von Erhebung zu Erhebung ändern kann, kam man zum Schluss, dass es methodisch nicht vertretbar ist, die Werte aus der Vorerhebung zu übernehmen. Man beschloss daher, anhand aller Werte aus den Vorjahren ein Modell bzw. die Wahrscheinlichkeiten zu schätzen, ob ein Unternehmen in einer nächsten Erhebung Forschung betreibt oder nicht und diese Schätzungen in den Stichprobenplan einfließen zu lassen.

3.3 Modellierung Wahrscheinlichkeiten

Die Idee ist eine logistische Regression anzuwenden, um die Wahrscheinlichkeit zu schätzen, dass ein Unternehmen im Jahr 2019 Forschung betreibt, gegeben die Antworten von 2012, 2015 und 2017. Als Basis dienen die Inputdaten mit allen Unternehmen, die im Jahr 2019 im Screening ja oder nein geantwortet haben (binäre Zielvariable x_{19}) mit den Hilfsvariablen bzw. Einflussgrössen den Variablen der Vorjahre (x_{12} , x_{15} und x_{17} , jeweils kategoriell mit 4 Ausprägungen). Dies lässt sich wie folgt formalisieren:

- t : Zeitpunkt aktuelle Erhebung 2021. Die Vorgängererhebungen werden der Einfachheit halber mit $t - 1$ (2019), $t - 2$ (2017) etc. indiziert.
- U_t : Unternehmenspopulation in Erhebung t .
- k : Unternehmen in U_t
- $x_{t-1,k}, x_{t-2,k}, x_{t-3,k}, x_{t-4,k}$: Hilfsvariablen aus den vier vorangehenden Erhebungen.
- $\hat{p}_{t-1,k} = \hat{f}_{t-1}(x_{t-2,k}, x_{t-3,k}, x_{t-4,k})$: geschätzte Wahrscheinlichkeit, dass k in $t - 1$ (2019) F+E betreibt. Dabei basiert $\hat{f}$ auf einer gewichteten logistischen Regression mit Zielvariable $y_{t-1} = 1/0$ (F+E / kein F+E in Erhebung $t - 1$, respektive 2019).

Folgender SAS-Code wurde implementiert, um das Modell mit einer logistischen Regression zu schätzen:

```
proc logistic data=logreg3 outmodel=model_sel outest=mod_param nocov;
class x_12 x_15 x_17 ;
weight weights_1;
model x_19(event='1')=x_12 x_15 x_17 /rsquare lackfit;
quit;
```

Als Gewicht wird jenes gewählt, dass die Antwortausfälle im Screening 2019 kompensiert ($weights_1$). Obiges Resultat ($model_sel$) ergibt mit folgendem Code für jedes Unternehmen eine geschätzte Wahrscheinlichkeit, ob es Forschung betreibt (P_1), oder nicht (P_2):

```
proc logistic inmodel=model_sel(type=logismod);
score clm data = logreg3 out=outp2(rename=(P_1=p_ja P_2=p_nein )) ;
run;
```

Es werden die Wahrscheinlichkeiten je Kreuzung von x_{12} , x_{15} und x_{17} gebildet:

```
proc means data=outp2 noprint;
class x_12 x_15 x_17;
var p_ja p_nein ;
output out=wkeiten_log mean(p_ja p_nein)=mean_pja mean_pnein;
run;
```

Die Resultate sind auszugsweise in der Tabelle 7 dargestellt.

Tabelle 7 Resultate logistische Regression

x_12	x_15	x_17	Anz_UNT (FE 19)	mean_pja	mean_pnein
2	2	2	7 453	0.03080	0.96920
.	2	2	3 160	0.03427	0.96573
5	2	2	70	0.05475	0.94525
2	2	.	88	0.06158	0.93842
2	.	2	143	0.06631	0.93369
2	5	2	382	0.06720	0.93280
.	2	.	251	0.06828	0.93172
.	.	2	1 101	0.07349	0.92651
.	5	2	205	0.07446	0.92554
2	2	5	306	0.07594	0.92406
1	2	2	302	0.08271	0.91729
.	2	5	191	0.08406	0.91594
5	2	.	1	0.10684	0.89316
5	.	2	1	0.11463	0.88537
5	5	2	10	0.11608	0.88392
etc.	.	.	2 521	0.14074	0.85926
etc.	.	.			
1	1	1	786	0.88199	0.11801
5	1	1	12	0.82762	0.17238
1	5	1	74	0.78668	0.21332
1	.	1	3	0.78428	0.21572
.	1	1	291	0.74627	0.25373
2	1	1	296	0.72480	0.27520
5	5	1	3	0.70318	0.29682
1	2	1	155	0.61928	0.38072
Total			20 191	0.15124	0.84876

Zur Erklärung bzw. Interpretation sollen folgende Beispiele dienen, die farbig markiert sind:

gelb: Die Wahrscheinlichkeit, dass ein UNT in FE19=2 (nein) hat, gegeben es hatte bereits 3x nein gesagt, ist: $\text{mean_pnein} = P(\text{FE19}=2 \mid \text{FE12}=\text{FE15}=\text{FE17}=2) = 96.9\%$. Umgekehrt ist $\text{mean_pja} = P(\text{FE19}=1 \mid \text{FE12}=\text{FE15}=\text{FE17}=2) = 3.1\%$

grün: Die Wahrscheinlichkeit, dass ein UNT in FE19=2 (nein) hat, gegeben es ist ein neues Unternehmen (bzw. 3x nicht im Rahmen) $\text{mean_pnein} = P(\text{FE19}=2 \mid \text{FE12}=\text{FE15}=\text{FE17}=\text{.}) = 85.9\%$. Umgekehrt ist $\text{mean_pja} = P(\text{FE19}=1 \mid \text{FE12}=\text{FE15}=\text{FE17}=\text{.}) = 14.1\%$

orange: Die Wahrscheinlichkeit, dass ein UNT in FE19=1 (ja) hat, gegeben es hatte bereits 3x ja gesagt, ist: $\text{mean_pja} = P(\text{FE19}=1 \mid \text{FE12}=\text{FE15}=\text{FE17}=1) = 88.2\%$. Umgekehrt ist $\text{mean_pnein} = P(\text{FE19}=2 \mid \text{FE12}=\text{FE15}=\text{FE17}=1) = 11.8\%$

Interpretation der Resultate:

Es gibt eine grosse Diskrepanz zwischen den geschätzten Wahrscheinlichkeiten. Je nach Kategorie können diese sehr unterschiedliche Werte annehmen. Es gibt aber keine, die exakt 0 und 1 sind.

Da eine hohe Korrelation zwischen diesen Wahrscheinlichkeiten und der Zielvariablen des Screenings (Ausgaben Ja oder Nein) erwartet werden darf, wird das geschätzte Modell dazu verwendet, eine entsprechende Hilfsvariable für den Stichprobenplan zu konstruieren. Mehr dazu im nächsten Abschnitt 3.4.

3.4 Definitiver Stichprobenplan

Für die Erstellung des Plans werden die obigen Resultate der logistischen Regression auf die neue Erhebung «übertragen». Bei diesem Vorgehen wird davon ausgegangen, dass sich das Mo-

dell und die geschätzten Regressionsparameter auf die nächste Erhebung übertragen lassen, um die entsprechenden Wahrscheinlichkeiten für F+E-Ausgaben zu schätzen. Es geht hier nur um die Kreation einer Hilfsvariablen zum Erstellen eines Plans. Falls also die Annahme eines unveränderten Modells verletzt wird, hat es einfach zur Folge, dass die Hilfsvariable weniger stark mit der Zielvariablen korrelieren und somit der Stichprobenplan weniger effizient würde. Etwas detaillierter:

- Schätzung der Wahrscheinlichkeit $\hat{p}_{t,k}$, dass k in der Erhebung t (2021) F+E betreibt mittels $\hat{f}_{t-1}(x_{t-1,k}, x_{t-2,k}, x_{t-3,k})$. Die Tabelle 8 zeigt einen Auszug wie diese Wahrscheinlichkeiten «übertragen» werden.

Tabelle 8 Geschätzte Wahrscheinlichkeiten $\hat{p}_{t,k}$ (=mean_pja) für den Plan 2021.

x_15	x_17	x_19	Anz_UNT (FE 21)	mean_pja	mean_pnein
2	2	2	9 608	0.03080	0.96920
.	2	2	933	0.03427	0.96573
etc.	.	.	2 837	0.14074	0.85926
etc.	.	.			
1	1	1	1 044	0.88199	0.11801
5	1	1	113	0.82762	0.17238
Total			22 175	0.15124	0.84876

- Konstruktion pps-Plan (pps: «probability proportional to size») mittels Hilfsvariable $\hat{p}_{t,k}$ so dass für die Erhebung 2021 eine Stichprobengrösse von ca. 14 000 Unternehmen resultierte. Dabei haben viele Unternehmen – vor allem solche mit vorangehender F+E-Aktivität – eine Ziehungswahrscheinlichkeit von 1. Die minimale Ziehungswahrscheinlichkeit liegt bei ca. 30%. Dies ergibt die Stichprobe s_t fürs Screening 2021.

Die Konstruktion des Stichprobenplans mit dem pps-Ansatz erfolgte in mehreren Schritten folgendermassen:

1. Aufteilung der in Abschnitt 2.1 bestimmten erweiterten Grundgesamtheit der 22 500 Unternehmen in drei Teile:
 - 2047 UNT: A priori voll erhobene Unternehmen im Nettorahmen: Grössenklassen 4 und 5, d.h. ab 100 Beschäftigten und «Big-Units», plus Gruppenmitglieder in Population und Paneleinheiten.
 - 328 UNT: Gruppenmitglieder ausserhalb Nettorahmen
 - 20 125 UNT: übrige Unternehmen der Population: Daraus wird die pps-Stichprobe gezogen.
2. Festlegung einer erhofften Stichprobengrösse gemäss folgender pps-Prozedur, ohne die voll erhobenen Einheiten, also theoretisch rund 14 000 - 2047 $\approx$ 12 000. Aufgrund der Tatsache, dass später (siehe Punkt 4) noch provisorische Werte abgerundet werden, führte $tot_samp = 13\,500$ zum gewünschten Resultat.
3. Berechnung der erwarteten Anzahl Unternehmen mit Forschung je Schicht:
 $Np = anzunt \times mean_pja$, wobei $anzunt$ die Anzahl Unternehmen in der Schicht ist (Kreuzung von x_15 x_17 x_19) und $mean_pja$ die geschätzte Wahrscheinlichkeit, dass ein Unternehmen in der entsprechenden Schicht F+E-Ausgaben hat.
4. Provisorische pps-Einschlusswahrscheinlichkeit je Schicht:
 $\pi_{pps0} = \min(1, \frac{tot_samp \times mean_pja}{\sum Np})$. Achtung: das Resultat des Quotienten könnte in einzelnen Schichten grösser als 1 sein, darum das Minimum-Kriterium.
5. Provisorisch erwartete Anzahl Einheiten in Stichprobe je Schicht:
 $samp_size_pps0 = \pi_{pps0} \times anzunt$.

6. Übersicht der provisorischen Stichprobengrösse je Typ (kat=1: aufgrund Schritt «4» voll erhobene Schichten vs. Rest: kat=2) mit $\text{tot_samp} = 13\,500$.

kat	samp_size_pps0	Flag
1: mit $\pi_{pps0}=1$	2585	
2: mit $\pi_{pps0}<1$	5601	tot_prov_kat2
Total	8186	tot_prov

7. Berechnung Erhöhungsfaktor koeff aufgrund der Tatsache, dass $\text{tot_prov}=8186 < 13500$ ist, wie folgt: $\text{koeff} = \frac{\text{tot_samp} - \text{tot_prov} + \text{tot_prov_kat2}}{\text{tot_prov_kat2}} = \frac{13500 - 8186 + 5601}{5601} = 1.94876$.

8. Berechnung definitive pps-Einschlusswahrscheinlichkeit je Schicht:

$\pi_{pps} = \min(\pi_{pps0} \times \text{koeff}, 1)$. Auch hier ist die Begrenzung auf 1 wieder notwendig.

9. Berechnung definitive Stichprobengrösse je Schicht: $\text{samp_size_pps} = \pi_{pps} \times \text{anzunt}$, ergibt im Total: $\sum \text{samp_size_pps} = 12\,057$.

Die Tabelle 9 zeigt einen Auszug des Stichprobenplans der F+E 2021. Die geplante Stichprobengrösse beträgt also 14 104 Unternehmen.

Der Plan wurde für die 22 172 Unternehmen gemäss Tabelle 3 erstellt. Dies ist die Nettopopulation inklusive den 95 Unternehmen für die Panelstudie, nicht berücksichtigt sind aber alle 328 Gruppenmitglieder ausserhalb der Grundgesamtheit. Letztere können bei Bedarf eine Ausgabe erhalten, also aktiviert werden, wenn dies via Gruppenverteilung notwendig wäre.

Tabelle 9 Stichprobenplan F+E 2021

x_15	x_17	x_19	strata_scr	Nh	mean_p_ja	mean_p_nein	Nh x mean_p_ja	π_{pps}	nh_pps
2	2	2	222	8 926	3.08%	96.92%	274.89	0.2999	2 678
0	2	2	022	893	3.43%	96.57%	30.60	0.3338	299
5	2	2	522	497	5.48%	94.52%	27.21	0.5333	266
2	2	0	220	253	6.16%	93.84%	15.58	0.5998	152
2	0	2	202	237	6.63%	93.37%	15.72	0.6458	154
2	5	2	252	382	6.72%	93.28%	25.67	0.6545	251
0	2	0	020	147	6.83%	93.17%	10.04	0.6650	98
0	0	2	002	1 590	7.35%	92.65%	116.84	0.7157	1 138
0	5	2	052	45	7.45%	92.55%	3.35	0.7252	33
2	2	5	225	245	7.59%	92.41%	18.60	0.7396	182
1	2	2	122	514	8.27%	91.73%	42.51	0.8056	415
0	2	5	025	29	8.41%	91.59%	2.44	0.8187	24
5	2	0	888	12	10.68%	89.32%	1.28	1	12
5	0	2	888	19	11.46%	88.54%	2.18	1	19
etc.									
5	1	1	888	85	82.76%	17.24%	70.35	1	85
1	1	1	888	654	88.20%	11.80%	576.82	1	654
Total 1				20 125	13.42%	86.58%	2 701.22		12 057
			999	2 047 apriori Take-All				1	2 047
Total 2				22 172				nh plan:	14 104

Die Variable `strata_scr` wird im Folgenden für die Gewichtung und Hochrechnung als Schichtvariable verwendet. Dabei werden alle Kreuzungen mit $\pi_{pps} = 1$ (aber nicht a priori voll erhoben) zu einer Schicht (888) zusammengefasst, da diese teilweise nur wenige Unternehmen umfassen.

3.5 Ziehung der Stichprobe im SVS

Die Ziehung erfolgte im Stichprobenverwaltungssystem SVS des BFS. Es wird eine Poisson-Stichprobe gezogen, wobei die Koordinierung mit allen anderen BFS-Erhebungen negativ festgelegt wird. Diese Punkte bedeuten einerseits, dass die resultierende Stichprobengrösse nicht

zwingend der geplanten entsprechen muss und mit der Koordination möchte man in den nicht voll erhobenen Schichten die Belastung möglichst gut auf alle Einheiten verteilen. Die effektive Stichprobengrösse der F+E 2021 beträgt so **14 078** Unternehmen.

4 Analyse Screening: Schätzung Anzahl Unternehmen mit Forschungstätigkeit

Die Screening-Antworten, welche im Januar 2022 vorlagen, ermöglichten eine Schätzung der Anzahl Unternehmen, die Forschung betreiben. Dazu muss eine Gewichtung mit Berücksichtigung der Antwortausfälle erstellt werden, was auf verschiedene Arten gemacht werden kann. Im Folgenden beschreiben wir zwei Arten zur Schätzung und geben entsprechende Resultate.

4.1 Grobe Schätzung basierend auf Erfahrungswerten der Vorerhebungen 2015-2019

Wir teilen die Unternehmen auf via der in Tabelle 9 definierten Schicht-Variable `strata_scr` – eine Aggregation aus den Schichten gemäss Plan – die die Unternehmen anhand ihrer Angaben der letzten drei Vorerhebungen 2015-2019 (siehe Abschnitt 3.3) klassifiziert. Wir nehmen an, dass in diesen Klassen die Antwortwahrscheinlichkeit gleich (oder ähnlich) ist für Unternehmen in derselben Klasse. Entsprechend können wir die Antwortwahrscheinlichkeit schätzen, indem wir die Anzahl der antwortenden durch die Anzahl der befragten Unternehmen in der Klasse dividieren.

Für die Schätzung der Anzahl Unternehmen mit Forschung in solch einer Klasse können wir dann die Anzahl Ja-Antwortenden multiplizieren mit der Anzahl Unternehmen im Rahmen geteilt durch die Anzahl Antwortende. Dadurch wird sowohl eine Antwortausfallkorrektur als auch das Stichprobengewicht der Ja-Antwortenden berücksichtigt.

Die Tabelle 10 zeigt die Schätzung der Unternehmen mit Forschung pro Schicht.

Tabelle 10 Schätzung pro Schicht (provisorisch*)

strata_scr	Brutto- rahmen	effekt. Brutto Stichprobe	Anz. Antw. Screening	Screening Anzahl=Ja	gew. Schätzung Anzahl=Ja
002	1 590	1 139	1 059	73	109.6
020	147	101	99	6	8.9
022	893	288	274	10	32.6
025	29	28	22	2	2.6
052	45	31	30	0	0
122	514	416	389	69	91.2
202	237	145	136	3	5.2
220	253	159	152	5	8.3
222	8 926	2 673	2 565	92	320.2
225	245	176	139	10	17.6
252	382	238	219	9	15.7
522	497	270	242	22	45.2
888	6 367	6 367	5 486	2 042	2 369.9
999	2 047	2 047	1 851	826	913.5
998	328	-	1	1	1
Total	22 500	14 078	12 663	3 170	3 941.5
*Bei Auswertung Haupterhebung noch bereinigt (s. Abschnitt 5)					

Die erste Spalte entspricht bis zur Zeile der Schicht 522 der Klassifizierung bezüglich der Schicht mit Werten aus Vorerhebungen 2015-2019, wobei 1=ja, 2=nein, 0= nicht im Rahmen, 5=keine Antwort. Die letzten drei Zeilen sind Aggregationen: 888=Restliche strata_scr zusammengefasst, da voll erhoben gemäss pps-Prozedur und auch als Antworthomogenitätsklassen betrachtet, 999=à priori voll erhobene Schichten (nicht in pps-Prozedur drin), und 998=Mitglieder von Unternehmensgruppen ausserhalb der Population (können indirekt via Muttergesellschaft antworteten). Diese würden mit einem Hochrechnungsgewicht von Eins hinzugefügt. Die Schätzung der Anzahl Unternehmen die Forschung betreiben, ergibt auf diese Weise 3942.

In Bezug auf die klassische Schätzmethode der Erhebungsmethodologie ist die obige Rechnungsart äquivalent zu folgender Formel:

$$\hat{Y} = \sum_{k \in J} \frac{1}{\pi_k} \cdot \frac{1}{p_k}, \quad (1)$$

wobei π_k die pps-Ziehungswahrscheinlichkeit ist, $J \subseteq S$ die Menge der Ja-Antwortenden, und p_k die Antwortwahrscheinlichkeit des Unternehmens. Da diese innerhalb einer Schicht als konstant angenommen wurde, schätzt man sie als Anzahl Antwortende der entsprechenden Schicht dividiert durch Anzahl gezogene Unternehmen in der Schicht.

4.2 Schätzung pro NOGA x Grössenklassen via logistische Regression

Ein gängiger Weg zur Schätzung der Anzahl Unternehmen, die Forschung betreiben ist folgender. Man teilt die Unternehmen aufgrund einer adäquaten Klassifikation auf und schätzt die Antwortwahrscheinlichkeit in jeder Klasse mittels logistischer Regression ähnlich wie in Abschnitt 3.3.

Die logistische Regression wurde so parametrisiert, dass das Vorgehen der in den Vorerhebungen angewandten Methode entspricht: Antworthomogenitätsklassen basierend auf NOGA X Grössenklasse (grkl2), wobei Letztere aus Tabelle (2) benutzt wurden. Die Gesamtanzahl Unternehmen mit Forschung schätzt man wie in Abschnitt 4.1 (Formel 1), wobei die p_k mittels der logistischen Regression geschätzt werden.

Tabelle 11 Schätzung pro NOGA-Branche (provisorisch*).

NOGA Branche	Brutto- rahmen	effekt. Brutto Stichprobe	Anz. Antw. Screening	Screening Anzahl=Ja	gew. Schätzung Anzahl=Ja
1	1 520	926	848	98	115.0
2	611	440	397	181	200.1
3	141	114	96	65	76.2
4	1 968	1 145	1 073	214	252.8
5	1 348	1 025	942	467	499.1
6	543	434	365	200	241.0
7	325	270	234	128	149.1
8	2 308	1 704	1 561	418	496.5
9	1 965	1 590	1 238	559	742.9
99	11 771	6 430	5 909	840	1 058.6
Total	22 500	14 078	12 663	3 170	3 831.1
*Bei Auswertung Haupterhebung noch bereinigt (s. Abschnitt 5)					

In der Tabelle 11 haben wir in den letzten zwei Spalten die Anzahl Unternehmen, die mit Ja geantwortet haben, und die daraus geschätzte Anzahl Unternehmen die Forschung betreiben. Diese

Werte sind hier pro NOGA-Branche und insgesamt angegeben, obwohl die Berechnung auch auf der Aufteilung pro Grössenklasse basiert. Das Total der Anzahl Unternehmen mit Forschung schätzen wir so auf 3831, also etwas tiefer als mit der ersten Methode.

Für die weiteren Analysen und für die zweiphasige Gewichtung des Screenings plus der Haupterhebung wurde die zweite Variante gewählt. Eine Besprechung hierzu und auch ein Vergleich mit anderen Alternativen folgt in Abschnitt 7.

5 Analyse Haupterhebung

Anfangs September 2022 lagen die definitiven, plausibilisierten und eingesetzten Daten der Haupterhebung zur Verfügung. Im Vergleich zum Stand der Screening-Antworten gab es während der Erhebung – wie bereits in den Vorerhebungen üblich – noch Änderungen bezüglich Gruppenmitglieder und auch betreffend der Information, ob ein Unternehmen im Erhebungsjahr wirklich Forschung betrieben hat. Es gibt Unternehmen, die im Screening mitgeteilt haben, dass sie bis Ende Jahr noch Forschung betreiben würden, es dann aber – gemäss Antwort in der Haupterhebung – doch nicht taten. Bei solchen Fällen wurde, unter der Annahme, dass das Nein korrekt ist, die Screening-Antwort Ja nachträglich auf ein Nein korrigiert. Es ist nicht auszuschliessen, dass sowohl einige dieser Nein, wie auch ein Teil der übrigen Nein-Antworten im Screening auf Unternehmen zurückzuführen sind, welche den Erhebungsaufwand vermeiden wollten. Beide Typen können ein Biasrisiko verursachen, d.h. es besteht die Gefahr, dass die Totalen F+E-Ausgaben unterschätzt werden, was hier aber nur schwierig zu überprüfen ist. Die Tabelle 12 enthält die Übersicht mehrerer Merkmale nach Branchen, Sektor, Grössenklassen und global (1. Zeile). Diese Merkmale sind folgende:

Anzahl BUR Anzahl aktive Unternehmen im ganzen Betriebs- und Unternehmensregister BUR

F+E Rahmen Anzahl Unternehmen im Rahmen der F+E 2021 (Nettorahmen plus antwortende Gruppenmitglieder des Brutorahmens)

Brutto Screen Anzahl Unternehmen in der Brutto-Stichprobe des Screenings

Netto Screen Anzahl antwortende Unternehmen in der Netto-Stichprobe des Screenings

AntRate Screen Antwortrate Screening in %

Anzahl Scr_nein Anzahl UNT mit Antwort Nein im Screening

Anteil Scr_nein Anteil UNT mit Antwort Nein im Screening

Brutto Erheb Bruttostichprobe für Haupterhebung

Netto Erheb Anzahl Antwortende bei Haupterhebung

Anzahl Antwtot Anzahl verwendbare Antworten für kombinierte Analyse Screening plus Haupterhebung: Dies sind alle Unt mit Antwort Nein im Screening plus Antwortende in Haupterhebung (Anzahl Scr_nein + Netto Erheb)

AntRate Scr+Erh Antwortrate Screening plus Haupterhebung in %: Anzahl Antwtot/Brutto Screen

AntRate Erheb Antwortrate (nur) Haupterhebung in %: Netto Erheb/Brutto Erheb

Tabelle 12 Antwortverhalten F+E 2021 Screening und Haupterhebung.

NOGA Branche	Sektor	Grössenklasse	Anzahl BUR	F+E Rahmen	Brutto Screen	Netto Screen	AntRate Screen	Anzahl Scr_nein	Anteil Scr_nein	Brutto Erheb	Netto Erheb	Anzahl Antwort	AntRate Scr+Erh	AntRate Erheb
			572 589	*22 413	*13 999	*12 759	91%	10 064	79%	2 695	1 648	11 712	84%	61%
		1.9	510 723	1 807	1 338	1 033	77%	676	65%	357	235	911	68%	66%
		10.49	50 859	16 401	9 440	8 741	93%	7 515	86%	1 226	737	8 252	87%	60%
		50.249	9 321	3 545	2 594	2 397	92%	1 566	65%	831	492	2 058	79%	59%
		250.499	1 011	405	391	369	94%	209	57%	160	106	315	81%	66%
		500	675	255	236	219	93%	98	45%	121	78	176	75%	64%
	2		92 304	11 446	7 026	6 455	92%	4 986	77%	1 469	872	5 858	83%	59%
	3		480 285	10 967	6 973	6 304	90%	5 078	81%	1 226	776	5 854	84%	63%
	2	10.49	14 219	8 580	4 761	4 364	92%	3 772	86%	592	336	4 108	86%	57%
	2	50.249	2 832	2 360	1 798	1 653	92%	1 018	62%	635	378	1 396	78%	60%
	2	250.499	311	294	292	273	93%	136	50%	137	91	227	78%	66%
	2	500	190	180	174	164	94%	60	37%	104	66	126	72%	63%
	3	1.9	510 723	1 807	1 338	1 033	77%	676	65%	357	235	911	68%	66%
	3	10.49	36 640	7 821	4 679	4 377	94%	3 743	86%	634	401	4 144	89%	63%
	3	50.249	6 489	1 185	796	744	93%	548	74%	196	114	662	83%	58%
	3	250.499	700	111	99	96	97%	73	76%	23	15	88	89%	65%
	3	500	485	75	62	55	89%	38	69%	17	12	50	81%	71%
1			4 785	1 519	925	855	92%	775	91%	80	46	821	89%	58%
2			1 544	610	439	409	93%	247	60%	162	100	347	79%	62%
3			296	141	114	100	88%	43	43%	57	34	77	68%	60%
4			7 605	1 966	1 143	1 088	95%	902	83%	186	109	1 011	88%	59%
5			3 271	1 341	1 018	968	95%	538	56%	430	262	800	79%	61%
6			1 263	543	434	385	89%	203	53%	182	101	304	70%	55%
7			806	323	268	235	88%	119	51%	116	65	184	69%	56%
8			21 776	2 290	1 686	1 567	93%	1 228	78%	339	198	1 426	85%	58%
9			1 964	1 964	1 589	1 251	79%	763	61%	488	323	1 086	68%	66%
99			529 277	11 716	6 383	5 901	92%	5 246	89%	655	410	5 656	89%	63%

Anmerkung zu den Werten in obiger Tabelle:

Es gibt einige Werte (mit * markiert), die sich im Vergleich zu den Tabellen 10 und 11 unterscheiden. Dies kommt daher, dass während der Haupterhebung Antworten des Screenings noch bereinigt oder geändert wurden. Die adaptierte Bruttostichprobe (Brutto Screen) von 13 999 lässt sich wie folgt herleiten bzw. erklären: Gezogene Einheiten Bruttostichprobe (14 078) minus Paneleinheiten (95), welche nicht aktiviert wurden, plus zusätzliche Antworten von Gruppenmitgliedern (16): $13999 = 14078 - 95 + 16$. Die Werte in der Tabelle 12 sind die definitiv gültigen und für die Schätzung der Resultate verwendeten Angaben.

Es ist ersichtlich, dass die sehr gute Antwortrate des Screenings (global 91%) kombiniert mit einer mittelmässigen, aber im Vergleich zu den Vorjahren durchaus üblichen Rate der Haupterhebung (61%) einen Datensatz von 11 712 Einheiten (oder 84% der Bruttostichprobe) liefert, welcher für die Weiterverarbeitung (wie Einsetzung, Gewichtung und Schätzung der Resultate) verwendet wird. Diese Verarbeitung wird in den nächsten Abschnitten erläutert.

6 Behandlung Merkmalsausfälle

In jeder Erhebung kann es einzelne Unternehmen geben, welche den Fragebogen nur teilweise ausfüllen. Man spricht in diesen Fällen von Merkmalsausfällen bzw. Item-Nonresponse (INR). Für die Auswertung des finalen Datensatzes ist es praktischer, wenn für jedes Unternehmen alle Antworten vollständig vorliegen, da so ein einziges Hochrechnungsgewicht für alle Erhebungsvariablen angewandt werden kann. Es gibt meistens zwei Möglichkeiten diese INR zu behandeln. Entweder

- man deklariert den Fragebogen als unbrauchbar, z.B. falls die Antworten zu den wesentlichen Fragen fehlen, oder
- man setzt die fehlenden Angaben mit einer geeigneten Methode ein.

Für die Erhebung F+E werden beide Optionen verwendet. Ein Fragebogen wird nur als brauchbar bezeichnet, wenn mindestens die Hauptfrage zu den Intramuros-Ausgaben (B0245) beantwortet wurde. Zudem werden seit mehreren Erhebungen vor allem robuste Methoden für die Einsetzung von Merkmalsausfällen verwendet, da vielfach extreme Werte in den Daten existieren. Die

in Bericht (Kilchmann, 2006) beschriebenen Methoden wurden ab 2008 optimiert und erweitert, respektive auf neue Fragen im Fragebogen angepasst. Grundsätzlich haben die Methoden nicht stark geändert, daher werden diese hier nicht im Detail beschrieben. Im Folgenden wird aber ein kurzer Überblick über die sequentielle Abfolge der verwendeten Einsetzungen gegeben, wobei die Bedeutung der Variablen auch im Fragebogen (siehe Anhang A) ersichtlich ist.

Zur Vereinfachung der Notation bedeuten im Folgenden die Variablennamen mit einem Hochkomma «'» am Schluss, entweder den originalen (falls vorhanden) oder den eingesetzten Wert.

Total F+E-Personal in der Schweiz Variable: H600T. Schätzung einer linearen Regression zwischen dem Beschäftigungstotal im Rahmen (Betot) und H600T anhand der antwortenden Einheiten nach Branche h . Pro Branche wird für die Konstruktion der Einsetzungen jeweils der resultierende Steigungskoeffizient $\hat{\beta}_h$ sowie der mad_h (median absolute deviation) der Residuen obiger Regression benutzt. Es wird $H600T_i = \hat{\beta}_h \cdot betot_i + \hat{\sigma}_h \cdot \epsilon_i$ verwendet, mit $\hat{\sigma}_h = 1.4826 \cdot mad_h(r_h)$, $r_{hi} = (H600T_i - \hat{\beta}_h \cdot betot_i)$ und $\epsilon_i \sim N(0, 1)$ normalverteilte Zufallsvariable. Anschliessend folgt eine Plausibilitätsprüfung der Einsetzung unter Bezug der F+E-Löhne (Variable B0210), bzw. dem Quotienten $\log(H600T'/B0210)$, die in seltenen Fällen eine Anpassung zur Folge hat.

Einsetzung für F+E-Ausgaben (1) Variablen: B0210, B0220, B0244.

Einsetzung basierend auf dem Modell $B02xx_i = (med_h(B02xx/B0245) + \hat{\sigma}_h \cdot \epsilon_i) \cdot B0245_i$ (mit $xx \in 10, 20, 44$). Dabei bezeichnet $med_h(B02xx/B0245)$ den Median des Quotienten $B02xx/B0245$ der antwortenden Unternehmen in der Branche h . $\hat{\sigma}_h$ und ϵ_i sind analog zur Einsetzung von H600T definiert, wobei $med_h(..../..)$ die Rolle des Regressionskoeffizienten übernimmt. Die eingesetzten Werte werden nun noch so angepasst, dass $B0245 = B0210' + B0220' + B0244'$. Meistens ist dies nach einer Einsetzung nicht der Fall und es wird folgende Prozedur angewandt:

Sei $I(B02xx)$ eine Indikatorfunktion, die den Wert 1 hat, wenn der Wert von $B02xx$ eingesetzt wurde, sonst 0. Seien

$sumVarB0245 = B0210' \cdot I(B0210) + B0220' \cdot I(B0220) + B0244' \cdot I(B0244)$ und

$sumDiffB0245 = B0245 - B0210' \cdot (1 - I(B0210)) - B0220' \cdot (1 - I(B0220)) - B0244' \cdot (1 - I(B0244))$. In dem Fall, dass keine Werte eingesetzt wurden, dann ist $sumVarB0245 = 0$ und $sumDiffB0245 = 0$. In dem Fall, dass alle Werte eingesetzt wurden, gilt $sumVarB0245 > 0$, in der Regel ungleich $B0245$, und $sumDiffB0245 = B0245$. Es gibt auch Fälle, bei denen nur ein oder zwei Werte eingesetzt wurden. Bei der Überprüfung der Summe wird in dem Fall, in dem $B02xx$ eingesetzt wurde, also $I(B02xx) = 1$, $B02xx$ durch $B02xx \cdot sumDiffB0245/sumVarB0245$ ersetzt.

Einsetzung für F+E-Ausgaben (2) Variablen: B0246, B0248, B0249.

Mit $B02xx_i = (med_h(B02xx/B0244') + \hat{\sigma}_h \cdot \epsilon_i) \cdot B0244'_i$. Anschliessend Anpassung Summe analog oben, wobei $B0244'$ die Rolle von $B0245$ übernimmt, so dass $B0244' = B0246' + B0248' + B0249'$.

Andere Personal Variablen gemäss Fragebogen Nr 6. (dabei ist: T= Anzahl Total, P= Total in VZÄ, E= Anzahl Ausländische, F= Anzahl Frauen). Es wird obige Idee fürs Modell angewandt:

$H600z_i = (med_h(H600z/H600T') + \hat{\sigma}_h \cdot \epsilon_i) \cdot H600T'_i$ (mit $z \in P, E, F$). Danach

$H640z_i = (med_h(H640z/H600z) + \hat{\sigma}_h \cdot \epsilon_i) \cdot H600z_i$ (mit $z \in P, E, F$).

$H665z_i = (med_h(H665z/H600z) + \hat{\sigma}_h \cdot \epsilon_i) \cdot H600z_i$ (mit $z \in P, E, F$).

$H699z_i = (med_h(H699z/H600z) + \hat{\sigma}_h \cdot \epsilon_i) \cdot H600z_i$ (mit $z \in P, E, F$).

Auch hier müssen wiederum die Summen kontrolliert und angepasst werden, so dass $H600z' = H640z' + H665z' + H699z'$.

Andere Variablen Alle anderen Variablen wurden bei Bedarf ebenfalls mit einer der oben beschriebenen Methoden und unter Berücksichtigung von Nebenbedingungen (einzuhaltende Summen) eingesetzt, wobei bei Anteilsvariablen die einzuhaltende Summe dann eben 100% ist.

7 Gewichtung

Die Schätzung der F+E-Ausgaben basiert auf Gewichten, die die finalen Antworten auf die Gesamtpopulation hochrechnen. Die Berechnung der Gewichte beinhaltet mehrere Schritte:

- Initialgewichtung, Antwortkorrektur fürs Screening und Antwortkorrektur für die Haupterhebung.
- Robustifizierung und Imputationen für Unternehmen, die in der Vergangenheit (und evtl. Gegenwart) hohe Forschungsausgaben deklariert haben.

Für die Antwortkorrektur des Screenings und die Antwortkorrektur der Haupterhebung testen, dokumentieren, und vergleichen wir die folgenden 3 Methoden:

- M1 Die bisherige Methode, basierend auf den beobachteten Antwortraten pro Branche $\times$ Grössenklasse.
- M2 Eine neue Methode, basierend auf dem CHAID-Algorithmus (CHi-squared Automatic Interaction Detection).
- M3 Eine Methode, basierend auf einer logistischen Regression, bei der die signifikanten Variablen mittels Variablenselektion in beiden Antwortkorrekturen ausgewählt wurden.

Mehr Details zu den drei Methoden M1-M3 werden im Abschnitt [7.1](#) erläutert.

7.1 Vergleich dreier Methoden zur Antwortkorrektur

Im Folgenden geben wir zuerst Details zu den benutzten Methoden, danach vergleichen wir Statistiken die zur Evaluation der Modelle der Antwortkorrektur nützlich sein können, analysieren die Auswirkung auf verschiedene wichtige Schätzungen der Ausgaben der Unternehmen und diskutieren die Schlussfolgerungen der Beurteilung dieser Varianten.

7.1.1 Detaillierter Beschrieb der Methoden zur Antwortkorrektur

- M1 Bei der bisher benutzten Methode werden sowohl beim Screening als auch bei der Haupterhebung pro Branche (BRA_NOGA_08Pond) $\times$ Grössenklasse (grkl_cal_erh) Antwortshomogenitätsklassen (comb) gebildet und man geht davon aus, dass die Antwortwahrscheinlichkeit innerhalb dieser Klassen konstant ist und anhand der beobachteten Antwortraten geschätzt werden kann.
Zur Berechnung einiger Kennzahlen – zum Vergleich mit den Methoden M2 und M3 – hat es sich als nützlich erwiesen, dass dieselben Werte auch durch eine Schätzung einer logistischen Regression mit comb als Einflussvariable berechnet werden können.
- M2 Die Methode basiert auf dem CHAID-Algorithmus und wählt die signifikanten Variablen und deren Modalitäten in einem Baum aus. Grenzen zum Aufbau solcher Bäume werden als Parameter geben, wodurch «overfitting» verhindert werden kann. Wir haben die folgenden Variablen für CHAID benutzt: NOGA, Grössenklasse, Vorerhebungsvariablen

(x_15, x_17, x_19). Die Variablennamen im SAS-Programm sind B_NOGA, x_15-x_19, grkl, grkl_cal_erh. Das benutzte Makro für den CHAID-Algorithmus wurde für die Antwortausfallkorrektur entwickelt und entsprechend sind u.a. zwei Grössen anzugeben: Die minimale (Brutto-)Grösse der Gruppen G_{min} im Baum und die minimale Antwortwahrscheinlichkeit p_{min} pro Gruppe. Im Screening ist die Antwortwahrscheinlichkeit bei 91% und 12 759 Unternehmen haben geantwortet. Wir haben $p_{min} = 0.8$ und $G_{min} = 100$ gesetzt. In der Haupterhebung war die Antwortwahrscheinlichkeit bei 61%, und 1648 von 2695 Unternehmen haben geantwortet. Wir haben $p_{min} = 0.4$ und $G_{min} = 50$ gesetzt.

- M3 Bei der Methode basierend auf einer logistischen Regression haben wir die signifikanten Variablen mittels den Selektionsmethoden (selection = forward, backward, stepwise) in beiden Antwortkorrekturen (Screening und Haupterhebung) ausgewählt. Die drei Selektionsmethoden ergaben ähnliche Resultate, so einigte man sich auf folgende erklärende Variablen (ohne Selektion):
- Screening: NOGA, Grössenklasse, und Vorerhebungsvariablen (x_15, x_17, x_19). Die Variablennamen sind wie bei M2 (CHAID).
 - Haupterhebung: NOGA und Vorerhebungsvariablen (x_15, x_17, x_19). Die Grössenklassen wurden weggelassen, da sonst zu kleine Zellen resultieren würden. Die Variablennamen sind wie bei M2 (CHAID).

7.1.2 Statistiken zur Evaluation der Antwortkorrektur

Die Statistiken, die wir vergleichen, basieren alle auf der logistischen Regression. Für M2 ist es relativ einfach, die erhaltene Aufspaltung des Baumes in eine logistische Regression so zu parametrisieren, dass sie die gleichen Wahrscheinlichkeitsschätzungen liefert wie der CHAID-Algorithmus. Für M1 mussten die Variablen Grössenklasse und Branche gekreuzt werden. Die resultierenden Kategorien werden in der logistischen Regression als erklärende Variablen verwendet, was zu den Wahrscheinlichkeiten und Gewichten der alten Methode (M1) führte. Wir vergleichen die folgenden Werte je separat pro Antwortkorrektur:

- percent concordant: Dies gibt den prozentualen Anteil der Paare $((r_i, \hat{p}_i), (r_j, \hat{p}_j))$ bei denen entweder Unternehmen i geantwortet hat ($r_i = 1$), Unternehmen j nicht geantwortet hat ($r_j = 0$), und $\hat{p}_i > \hat{p}_j$, oder i hat nicht geantwortet, j hat geantwortet und $\hat{p}_i < \hat{p}_j$ in der Menge der Paare an, bei denen ein Unternehmen geantwortet hat und das andere nicht. Dabei bezeichnet $\hat{p}_i$ die geschätzte Antwortwahrscheinlichkeit von i .
- percent discordant: Dies gibt den prozentualen Anteil der Paare an bei denen die Wahrscheinlichkeiten umgekehrt sind in der Menge der Paare bei denen ein Unternehmen geantwortet hat und das andere nicht (analog aber umgekehrt wie in percent concordant), also Paare i, j mit $\hat{p}_i > \hat{p}_j$ und $r_i < r_j$.

Die Details zu obigen Kennzahlen sind hier beschrieben: (SAS Institute Inc., 2018, The LOGISTIC Procedure, Option: lackfit, p. 5845 und 5857).

Die folgenden Tabellen enthalten die Resultate für das Screening (Tabelle 13) bzw. die Haupterhebung (Tabelle 14).

Tabelle 13 Resultate für das Screening.

1.Nonresponse	concordant (%)	discordant (%)
M1	62.2	28.6
M2	65.0	25.6
M3	71.2	27.5

Tabelle 14 Resultate für die Erhebung.

2.Nonresponse	concordant (%)	discordant (%)
M1	56.3	37.7
M2	49.5	27.7
M3	61.4	37.1

Insgesamt können wir beobachten, dass die logistische Regression (M3) in beiden Antwortkorrekturen am besten (bezüglich «concordant») abschneidet. Mit den gesetzten Parametern hat die CHAID-Methode in beiden Antwortkorrekturen am wenigsten «falsche» Paare im Sinne, dass die «discordant»-Werte am tiefsten sind. Dabei ist aber zu beachten, dass im Rahmen der Erhebung bei M2 die Summe von «concordant» und «discordant» nur auf 77.2% führt, was auf einen recht hohen Anteil «ties» (Fälle mit $\hat{p}_i = \hat{p}_j$) zurückzuführen ist.

Es wurde weiter für alle Methoden der p-Wert des Hosmer-Lemeshow-Tests analysiert. Dieser Test überprüft die Anpassungsgüte des entsprechenden Modells der logistischen Regression. In allen drei Fällen ergab sich ein hoher Wert > 0.7 , was grundsätzlich auf ein gutes Modell hinweist.

Zudem wurden noch die Verteilungen der Faktoren der Antwortkorrektur pro Etappe (Screening: $r1_scr$, Haupterhebung: $r2_erh$) und jene der Gesamtgewichte ($weights$) analysiert (siehe Tabelle 15). Insgesamt sind die Verteilungen der drei Variablen sehr ähnlich, enthalten also keine allzu grossen Unterschiede, weshalb sie hier auch nicht weiter diskutiert werden müssen.

Tabelle 15 F+E 2021: Verteilung Antwortwahrscheinlichkeiten und Gewichtung, Methoden M1-M3

Verteilung Percentile	M1			M2			M3		
	$r1_scr$	$r2_erh$	$weights$	$r1_scr$	$r2_erh$	$weights$	$r1_scr$	$r2_erh$	$weights$
100%	1.000	1.000	7.496	0.976	1.000	7.380	1.000	1.000	8.555
99%	1.000	1.000	5.764	0.976	1.000	6.873	1.000	1.000	6.800
95%	0.953	1.000	2.456	0.976	0.681	2.502	0.973	0.735	2.816
90%	0.953	0.667	2.155	0.976	0.681	2.225	0.972	0.698	2.190
75%	0.951	0.653	1.957	0.953	0.681	1.930	0.959	0.673	1.910
50%	0.929	0.600	1.611	0.931	0.568	1.570	0.936	0.599	1.554
25%	0.892	0.554	1.088	0.891	0.497	1.085	0.891	0.535	1.099
10%	0.851	0.513	1.052	0.804	0.497	1.070	0.827	0.482	1.042
5%	0.770	0.452	1.000	0.804	0.497	1.033	0.773	0.421	1.023
1%	0.770	0.421	1.000	0.803	0.444	1.000	0.519	0.340	1.000
0%	0.770	0.333	1.000	0.803	0.444	1.000	0.322	0.200	1.000

Wie in Tabelle 15 ersichtlich, führt der Ansatz M3 – basierend auf der Variablenselektion mittels logistischer Regression – zu teilweise recht kleinen geschätzten Antwortwahrscheinlichkeiten, was wiederum hohe Hochrechnungsgewichte impliziert. Aus Stabilitätsgründen hat es sich in der F+E erfahrungsgemäss als zweckmässig erwiesen, dass die Gewichte in einem moderaten Bereich liegen und über die Zeit eine gewisse Konstanz aufweisen.

Mit der Methode M1 kann dies annähernd berücksichtigt werden, solange Unternehmen die Schicht nicht wechseln.

7.1.3 Auswirkungen der Modelle der Antwortkorrektur auf die Schätzungen

Die Tabelle 16 zeigt den Vergleich der hochgerechneten Resultate der Ausgaben pro NOGA-Branche. Die Methoden sind wie in Abschnitt 7 benannt als M1, M2 und M3, wobei «Big-Units» und Unternehmensgruppen wie in 7.2 erklärt, einbezogen wurden. Die Tabelle 17 vergleicht die Resultate pro Methode für die Hauptvariablen.

Tabelle 16 Schätzungen Intramurosausgaben B0245 nach NOGA-Branche (in 1000 CHF). Dabei steht «Sum» für Totalschätzung und «cv» für den Variationskoeffizienten

Branche	n	M1:Sum	M1:cv	M2:Sum	M2:cv	M3:Sum	M3:cv
1	46	64 111.1	8.92%	59 077.0	8.51%	63 654.4	9.53%
2	100	435 451.7	7.02%	420 787.9	5.98%	424 052.0	6.02%
3	34	6 246 729.0	4.10%	5 954 133.3	2.57%	5 980 698.1	2.66%
4	109	298 988.5	4.87%	288 071.5	4.43%	291 398.9	4.50%
5	262	1 513 093.1	6.79%	1 509 459.0	6.42%	1 518 221.7	6.43%
6	101	841 938.4	8.21%	819 536.8	8.21%	871 720.6	8.67%
7	65	1 314 589.6	6.18%	1 312 794.2	6.25%	1 295 380.6	5.80%
8	198	1 056 470.2	8.97%	990 118.5	9.89%	1 035 415.6	10.24%
9	323	2 166 891.3	6.85%	2 149 520.2	6.52%	2 099 348.9	6.34%
99	410	2 841 592.8	4.23%	2 810 752.4	4.34%	2 762 964.7	4.27%
Total	1 648	16 779 855.7	2.15%	16 314 250.7	1.79%	16 342 855.4	1.80%

Tabelle 17 Resultate Hauptvariablen (Schätzungen global, in 1000 CHF).

Variable		M1: Sum	M1: cv	M2: Sum	M2: cv	M3: Sum	M3: cv
Anzahl Unt mit F+E		3 206.2	1.45%	3 201.3	1.53%	3 228.3	1.55%
Intramuros	B0245	16 779 855.7	2.15%	16 314 250.7	1.79%	16 342 855.4	1.80%
Extramuros	G0295	8 101 013.0	4.16%	8 062 624.2	3.76%	8 049 243.2	3.76%
Aufwendungen Ausland	J0205	20 315 785.8	0.85%	20 388 793.3	0.90%	20 486 848.8	1.02%
F+E Personal Frauen	H600F	15 461.1	2.50%	15 092.1	2.14%	15 123.4	2.15%
F+E Personal Total	H600T	61 879.3	2.31%	60 128.4	2.19%	60 372.4	2.19%
F+E Personal VZÄ	H600P	54 008.6	2.37%	52 409.3	2.24%	52 722.1	2.24%

Kommentar zu obigen Tabellen:

Anhand der Tabelle 16 ist zu erkennen, dass die für die NR-Korrektur eingesetzte Methodik die Resultate beeinflussen kann. Die Schätzungen und die Variationskoeffizienten sind in der Regel jedoch in der gleichen Grössenordnung. Eine wichtige Grösse sind die Intramurosausgaben (B0245), welche je nach Methode zwischen 16.3 (mit M2 und M3) und 16.8 (mit M1) Milliarden Franken variieren. Diese Unterschiede zwischen den drei Totalschätzungen relativieren sich, wenn zugleich die Genauigkeiten (Variationskoeffizienten oder Vertrauensintervalle) berücksichtigt werden.

Aus Gründen der Vergleichbarkeit der Resultate mit den Vorjahren wurde darum entschieden, dem bisherigen Vorgehen (M1) den Vorzug zu geben, obwohl die Qualitätsmasse in Tabellen 13 und 14 eher für die Variante 3 sprachen. Nebst der Vergleichbarkeit der F+E 2021 mit der Vorgängererhebung, spielte dabei die Anwendbarkeit der Methode in der Zukunft eine wichtige Rolle: Die Methode 1 basiert nur auf Hilfsvariablen, welche aus dem Betriebs- und Unternehmensregister stammen und somit in Zukunft verfügbar sein werden. Die in M3 verwendeten Informationen basieren auf Information vergangener Vollerhebungen. Da 2021 die Vollerhebung durch eine Stichprobenerhebung ersetzt wurde, wird diese Information voraussichtlich nicht mehr in dieser Form zur Verfügung stehen. Ein Ausblick auf die zukünftigen Erhebungen ist im Abschnitt 9 beschrieben.

7.2 Behandlung einflussreiche Beobachtungen und Gruppen

Bei jeder Erhebung haben ein paar wenige «Big-Units» nicht geantwortet und mussten daher behandelt werden. Im Rahmen der F+E 2019 wurden für die Behandlung von fehlenden Antworten bei den grossen Unternehmen verschiedene Einsetzungsalternativen untersucht, welche die Ausgaben 19 mittels einem Regressionsmodell basierend auf den Werten der vorherigen Erhebung schätzen. Aufgrund der vielen extremen Werte bei den für die Modellierung verwendeten grossen Unternehmen, resultierten hieraus teilweise sehr instabile Einsetzungen, welche auch

sehr sensibel auf den Einsatz von Robustifizierungsmethoden reagierten. Dabei ist zu sagen, dass gerade die Stabilität (insbesondere in Bezug auf die resultierenden Evolutionsschätzungen) die Sonderbehandlung (d.h. Sonderschicht, Einsetzung basierend auf unternehmensspezifischer Information statt Kompensation durch Gewichtung anderer Unternehmen) dieser als einzigartig betrachteten Unternehmen motivierte. Weitere Problematiken, welche sich im Rahmen betrachteter Regressionseinsetzungen ergaben, waren

- Inkonsistenzen zwischen den eingesetzten, untereinander nicht unabhängigen Variablen
- Inkompatibilität der betrachteten Regressionsmodelle mit der Tatsache, dass einige Zielvariablen oft den Wert Null annehmen

Vor diesem Hintergrund und der Tatsache, dass es sich nur um eine kleine Anzahl zu behandelnden Unternehmen handelte (7), wurde entschieden, auf den letzten beobachteten Erhebungswert zurückzugreifen. Diese Einsetzungsmethode ist auch bekannt unter dem Namen «last observation carry-forward» (LOCF). Eine Rolle für diese Wahl spielten auch empirische Analysen, welche zeigten, dass dieses Vorgehen den gewünschten stabilisierenden Effekt zeigte. Hingegen liegt auf der Hand, dass das Vorgehen ein erhöhtes Biasrisiko mit sich bringt und es darum wichtig ist, dass bei den «Big-Units» eine Antwortrate von 100% angestrebt wird. Die Auswirkung auf die Varianzschätzung wird in Abschnitt (8.2) beschrieben.

Nach Behandlung der Antwortausfälle mittels dieser Einsetzung sind für alle «Big-Units» Werte vorhanden. Entsprechend werden diese im Rahmen der Hochrechnung mit 1 gewichtet. Dasselbe Vorgehen wurde auch für die F+E 2021 angewandt.

Ein bereits in einer Vorerhebung eingesetzter Wert wird während maximal zwei Erhebungen weiterverwendet, d.h. antwortet ein Unternehmen in drei aufeinanderfolgenden Erhebungen nicht, wird es nicht mehr eingesetzt, sondern via Gewichtung kompensiert. Diese Prozedur soll verhindern, dass ein zu alter Wert (hier aus 2015 oder früher) in die Berechnung der Ausgaben von 2021 einfließen.

Für die Gewichtung auch in eine separate Schicht gelegt wurden alle zu einer Gruppe gehörenden antwortenden Unternehmen, unabhängig ob diese im Initialrahmen sind oder nachträglich hinzugefügt wurden. Also erhalten auch diese ein Hochrechnungsgewicht von 1.

7.3 Resultierende Gewichtung

Sei R_1 die Menge der antwortenden Unternehmen im Screening, wobei R_1^y jene sind die Ja geantwortet haben, R_1^n sind jene die Nein antworteten. R_2 ist die Menge der antwortenden Einheiten in der Haupterhebung.

Das Hochrechnungsgewicht w_k eines antwortenden Unternehmens k in der Schicht h (strata_scr) wird berechnet, indem das aus der inversen Einschlusswahrscheinlichkeit $\pi_{pps,h}$ (gemäß Abschnitt 3.4) berechnete Initialgewicht $w_{ini_{k,h}} = 1/\pi_{pps,h}$ mit den inversen Antwortwahrscheinlichkeiten von Screening und Haupterhebung (je Antworthomogenitätsklasse A) multipliziert werden. Dabei wird Letzteres nur für die Unternehmen, die angegeben haben Forschung zu betreiben, gemacht. Die Ausnahme sind die antwortenden Gruppen G und «Big-Units» B^* , deren Gewicht auf 1 gesetzt wird.

$$w_k = \begin{cases} 1 & \text{falls } k \in R_1^y \cap R_2 \cap (G \cup B^*) \\ w_ini_{k,h} \times \frac{1}{\widehat{resp1}_A} \times \frac{1}{\widehat{resp2}_A} & \text{falls } k \in R_1^y \cap R_2 \setminus (G \cup B^*) \\ w_ini_{k,h} \times \frac{1}{\widehat{resp1}_A} & \text{falls } k \in R_1^n \setminus (G \cup B^*) \\ 1 & \text{falls } k \in R_1^n \cap (G \cup B^*) \end{cases}$$

wobei $\widehat{resp1}_A = n_{1,A}/n'_{1,A}$ die geschätzte Antwortwahrscheinlichkeit des Screenings und $\widehat{resp2}_A = n_{2,A}/n'_{2,A}$ jene der Haupterhebung in der Homogenitätsklasse A ist. Dabei bezeichnen $n_{\cdot,A}$ die Netto- und $n'_{\cdot,A}$ die Bruttogrößen in der jeweiligen Phase.

8 Schätzung der Resultate mit Varianzschätzung

8.1 Varianzschätzung ohne Berücksichtigung Einsetzung

Die Resultate der F+E-Erhebung wurden mit der Prozedur `surveymeans` der Software SAS berechnet. Diese führt zu folgenden Berechnungen:

Ein Total einer Variablen im Untersuchungsbereich D (für Domain wie z.B. nach Branche oder Grössenklasse) wird geschätzt mit

$$\hat{Y}_D = \sum_h \sum_{k \in D} w_k y_k \quad (2)$$

wobei y_k das Ausgabentotal eines Unternehmens k ist. Für die Varianz des Totalschätzers ergibt sich

$$\begin{aligned} \widehat{\text{Var}}(\hat{Y}_D) &= \sum_h n_h \left(1 - \frac{n_h}{N_h}\right) s_{y,D,h}^2 \\ &= \sum_h n_h \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h - 1} \sum_{k \in h} \left(w_k y_{k,D} - \frac{\sum_{k \in h} w_k y_{k,D}}{n_h} \right)^2 \end{aligned} \quad (3)$$

mit $y_{k,D} = y_k$, falls k im Domain D und 0 sonst, N_h dem Populationstotal und n_h den Anzahl Antwortenden (Anzahl mit «Screening: Nein» plus Anzahl «Screening: Ja und mit Ausgaben in Erhebung» inkl. Eingesetzten) in der Schicht h . Die letzten beiden Variablen findet man aggregiert nach Branche unter «F+E Rahmen» und «Anzahl Antworttot» in Tabelle 12.

Obige Formel behandeln eingesetzte Werte wie echte Antworten, d.h. die Varianzschätzung ist im Prinzip zu tief. Für die Erhebung 2019 wurde zu Testzwecken erstmals eine Formel angewandt, welche dieses Manko behebt, d.h. zusätzlich die Einsetzungsvarianz berücksichtigt. Die Formeln werden in folgendem Abschnitt erläutert.

8.2 Varianzschätzung: Einsetzungsvarianz «last observation carry-forward»

Bei der F+E gibt es eine Liste von grossen Unternehmen «Big-Units». Sofern diese nicht antworten, werden diese mit dem Vorerhebungswert ohne Berücksichtigung einer zusätzlichen Entwicklung eingesetzt. Diese Methode wird auch als «last observation carry-forward» (LOCF) bezeichnet. Es soll die Einsetzungsvarianz geschätzt werden, wozu folgende Notation eingeführt wird. Seien

- k Identifikator für die Unternehmen.
- B die Menge der 10 Branchen.
- B^* die Menge der «Big-Units» (voll erhoben) in allen Branchen.
- b eine Branche aus B
- b^* die Menge der «Big-Units» in b
- N_{b^*} die Anzahl «Big-Units» in b
- r_k der Indikator, ob die Einheit k in der Erhebung t geantwortet hat ($r_k = 1$), oder nicht ($r_k = 0$).
- b_r^* die Menge der antwortenden «Big-Units» in b vor Einsetzung.
- m_{b^*} die Anzahl der antwortenden «Big-Units» in b vor Einsetzung.
- y_k der Wert (die Ausgaben für F+E) aus der aktuellen Erhebung t .
- x_k der entsprechende Wert aus der letzten Erhebung $t - 1$.
- D ein interessierender Untersuchungsbereich
- w_k das Hochrechnungsgewicht der Einheit k .

Zudem ist $y_{k,D} = y_k$ und $x_{k,D} = x_k$ falls k im Domain D und 0 sonst. Die Ausgaben der grossen Einheiten im Untersuchungsbereich D werden mit der LOCF-Methode wie folgt geschätzt:

$$\begin{aligned}
 \hat{Y}_{D \cap B^*} &= \sum_{k \in B^*} w_k \left(r_k y_{k,D} + (1 - r_k) x_{k,D} \right) \\
 &= \sum_{k \in B^*} w_k \left(r_k (y_{k,D} - x_{k,D}) + x_{k,D} \right) \\
 &= \sum_{b \in B} \sum_{k \in b^*} r_k (y_{k,D} - x_{k,D}) + x_{k,D}
 \end{aligned} \tag{4}$$

da $w_k = 1$ (für $k \in B^*$). Im Folgenden wird die Annahme gemacht, dass r_k unabhängige Bernoulli-Variablen sind, d.h. $r_k \sim \text{BIN}(1, \theta_k)$, wobei θ_k die Antwortwahrscheinlichkeit von k bezeichnet, die mittels $\hat{\theta}_k = m_{b^*} / N_{b^*}$ geschätzt wird. Unter diesen Voraussetzungen gilt für die Varianz von $\hat{Y}_{D \cap B^*}$

$$\begin{aligned}
 \text{Var}(\hat{Y}_{D \cap B^*}) &= \text{Var} \left(\sum_{k \in B^*} r_k (y_{k,D} - x_{k,D}) + x_{k,D} \right) \\
 &= \sum_{k \in B^*} \text{Var} \left(r_k \right) (y_{k,D} - x_{k,D})^2 \\
 &= \sum_{k \in B^*} \theta_k (1 - \theta_k) (y_{k,D} - x_{k,D})^2 \\
 &= \sum_{b \in B} \sum_{k \in b^*} \theta_k (1 - \theta_k) (y_{k,D} - x_{k,D})^2
 \end{aligned}$$

Die Schätzung dieser Varianz erfolgt durch

$$\begin{aligned}
 \widehat{\text{Var}}(\hat{Y}_{D \cap B^*}) &= \sum_{b \in B} \sum_{k \in b_r^*} \frac{1}{\hat{\theta}_k} \hat{\theta}_k (1 - \hat{\theta}_k) (y_{k,D} - x_{k,D})^2 \\
 &= \sum_{b \in B} \sum_{k \in b_r^*} (1 - \hat{\theta}_k) (y_{k,D} - x_{k,D})^2 \\
 &= \sum_{b \in B} \sum_{k \in b_r^*} (1 - m_{b^*} / N_{b^*}) (y_{k,D} - x_{k,D})^2 \\
 &= \sum_{b \in B} (1 - m_{b^*} / N_{b^*}) \sum_{k \in b_r^*} (y_{k,D} - x_{k,D})^2
 \end{aligned} \tag{5}$$

Die «Big-Units» werden als separate, voll erhobene Schicht betrachtet, deren durch Antwortausfälle verursachte Varianz in (5) geschätzt wird. In Formel (3) werden die «Big-Units» als voll erhoben, mit Varianz Null behandelt ($n_h = N_h$). Für die Varianzschätzung des Totals wird somit die Varianz der «Big-Units» (5) und jene der übrigen Unternehmen (3) summiert.

$$\widehat{\text{Var}}(\hat{Y}_D)^{\text{mit}} = \widehat{\text{Var}}(\hat{Y}_D)^{\text{ohne}} + \widehat{\text{Var}}(\hat{Y}_{D \cap B^*}) \tag{6}$$

Die Resultate für die Schätzung der Intra- und Extramurosangaben der F+E 2019 und 2021, mit Angabe des Variationskoeffizienten ohne (cv_ohne) und mit (cv_mit) Berücksichtigung der Einsetzung befindet sich in Tabelle 18.

Tabelle 18 F+E 2019 / 2021: Varianzberechnung ohne und mit Berücksichtigung der Einsetzungsvarianz nach Branchen

F+E 2019

NOGA-Branche	Intramuros (B0245)			Extramuros (G0295)			Big-Units (grosse Einheiten)			
	Total	cv_ohne	cv_mit	Total	cv_ohne	cv_mit	Total Rahmen	Anzahl noch aktive	Anzahl Antwortende	Anzahl eingesetzte
1	79 135.9	9.59	10.20	15 177.9	26.69	26.71	4	4	3	1
2	509 700.6	7.96	7.96	16 772.5	34.98	34.98	6	6	6	0
3	5 163 788.4	3.43	3.43	5 573 354.7	2.31	2.31	0	0	0	0
4	380 931.1	3.27	3.27	110 123.6	1.62	1.62	6	6	2	4
5	1 731 488.0	5.65	5.65	165 289.1	13.88	13.88	4	4	4	0
6	913 312.6	11.15	11.69	142 121.8	16.13	25.17	8	8	5	3
7	1 119 490.9	4.40	4.40	88 260.0	10.79	10.79	1	1	0	1
8	639 170.9	5.62	5.62	64 775.1	5.80	5.80	9	4	4	0
9	1 988 012.2	8.22	9.10	1 002 612.7	19.96	22.61	4	4	2	2
99	2 917 933.7	2.42	2.42	1 236 909.6	0.65	5.84	19	16	15	1
	15 442 964.2			8 415 397.2			61	53	41	12

F+E 2021

NOGA-Branche	Intramuros (B0245)			Extramuros (G0295)			Big-Units (grosse Einheiten)			
	Total	cv_ohne	cv_mit	Total	cv_ohne	cv_mit	Total Rahmen	Anzahl noch aktive	Anzahl Antwortende	Anzahl eingesetzte
1	64 111.1	8.92	8.96	1 292.4	10.40	14.66	6	4	2	2
2	435 451.7	7.02	7.02	3 494.3	15.61	15.61	9	7	7	0
3	6 246 729.0	4.10	4.10	4 668 690.5	2.61	2.61	2	2	2	0
4	298 988.5	4.87	4.89	105 290.7	1.01	1.29	7	6	4	2
5	1 513 093.1	6.79	6.85	168 686.9	15.38	15.96	7	7	6	1
6	841 938.4	8.21	8.28	44 158.4	14.86	69.98	9	8	6	2
7	1 314 589.6	6.18	6.18	61 475.0	13.22	13.22	3	3	3	0
8	1 056 470.2	8.97	8.97	240 951.8	18.37	18.37	7	4	4	0
9	2 166 891.3	6.85	6.85	1 286 539.8	21.91	21.91	8	8	8	0
99	2 841 592.8	4.23	4.23	1 520 433.4	8.50	8.50	7	4	4	0
	16 779 855.7			8 101 013.0			65	53	46	7

Bemerkungen zu obiger Tabelle 18:

Je nach Erhebungsvariable und Anzahl Einsetzungen kann der CV mit Berücksichtigung der Einsetzungsvarianz (cv_mit) weniger oder mehr vom CV ohne Berücksichtigung (cv_ohne) abweichen. Wie ersichtlich stimmen in den Branchen, die keine eingesetzten Werte haben (Spalte ganz rechts: «Anzahl eingesetzte»=0), die beiden Variationskoeffizienten überein. Gibt es eine Branche mit Anzahl Antworten = 0 aber eingesetzte Antworten > 0 (wie Branche 7 im 2019), kann keine Varianz mit Einsetzung berechnet werden (hier wurde cv_mit = cv_ohne gesetzt).

Die auf der Webseite des BFS publizierten offiziellen Resultate beinhalten im Zeitpunkt der Publikation dieses Berichtes die Werte «ohne» Berücksichtigung der Einsetzung. Generell kann festgehalten werden, dass die Varianzen bzw. die Genauigkeiten daher eher leicht unterschätzt werden, obwohl wie ersichtlich bei G0295 (Branche 6) diese Differenz auch relativ gross sein kann.

9 Schlussfolgerungen und Ausblick

Dieser Methodenbericht beschreibt die wichtigsten aktuellen methodischen Aspekte der Erhebung Forschung und Entwicklung 2021. Dabei kommt der Neukonzeption des Screenings eine bedeutende Rolle zu. Die Vollerhebung der ersten Phase der Erhebung wurde dabei durch eine Stichprobenerhebung ersetzt, wodurch der Umfang der befragten Unternehmen von rund 22 000 Unternehmen auf 14 000 Unternehmen reduziert wurde. Die Entwicklung des Stichprobenplans basierte auf Erhebungsinformationen der Vorgängererhebungen 2012 bis 2019, welche dafür verwendet wurden, die Wahrscheinlichkeit zu schätzen, dass ein Unternehmen 2021 F+E betreibt. Diese Wahrscheinlichkeiten erlaubten es, einen Stichprobenplan zu entwickeln, mit welchem – trotz Verzicht auf eine Vollerhebung – Unternehmen mit F+E mit grosser Wahrscheinlichkeit befragt wurden. Erfahrungsgemäss ist ein grosser Anteil der F+E-Ausgaben auf einige wenige Unternehmen zurückzuführen. Entsprechend spielt die Behandlung dieser Unternehmen im Rahmen der F+E-Erhebung eine wichtige Rolle. Vor diesem Hintergrund wurde eine Methodik entwickelt, welche es erlaubt, potentiell sehr einflussreiche Unternehmen systematisch zu identifizieren und im Rahmen des Statistischen Datenaufbereitungsprozesses und der Hochrechnung gesondert zu behandeln. Dabei ist es für die Qualität der Erhebungsergebnisse von zentraler Bedeutung, dass bei diesen Unternehmen eine Antwortrate von 100% angestrebt wird.

Für die nächsten F+E-Erhebungen (ab 2023) soll eine Methode entwickelt werden, die der Tatsache Rechnung trägt, dass eine oder mehrere Vorgängererhebungen teilweise auf einer Stichprobe beruhen können. Die Schätzung der Wahrscheinlichkeit, dass ein Unternehmen F+E-Ausgaben tätigt, kann nicht so wie in Abschnitt 3 beschrieben auf alle Unternehmen 1 zu 1 umgesetzt werden, denn im Vergleich zu Tabelle 5 gibt es neu eine Kategorie «6: nicht in der Stichprobe». Um diese neue Situation methodisch korrekt zu handhaben, wird nur für die in der F+E 2021 voll erhobenen Unternehmen die entsprechende Erhebungsinformation verwendet, um die Einschlusswahrscheinlichkeiten für die F+E 2023 zu berechnen. Für die anderen Unternehmen greift man auf die bisherigen Informationen zurück.

Überlegungen, wie die Situation längerfristig gehandhabt werden soll, sind im Gange.

Aus Zeitgründen für die Anpassungen der Erhebung 2021 weggelassen, aber bei einer nächsten Überprüfung durchaus denkbar, sind zudem folgende Aspekte:

- Testen des «random forest»-Algorithmus zur Schätzung von Antwortwahrscheinlichkeiten.

- Optimierung der Berechnung der «Big-Units» und Automatisierung der Nachverfolgung der durch Fusionen tangierten Unternehmen.
- Weitere Analysen zur Beurteilung der Qualität der implementierten Methoden wie zum Beispiel Schätzung von Evolutionsvarianzen.

Literatur

- [Kilchmann 2006] KILCHMANN, Daniel: *Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004*. Bundesamt für Statistik, Neuchâtel, 2006
- [SAS Institute Inc. 2018] SAS INSTITUTE INC.: *SAS/STAT® User's Guide*. Bd. 15.1. Cary, NC: SAS Institute Inc., 2018
- [Särndal, C.E., et al. 1992] SÄRNDAL, C.E., ET AL.: *Model Assisted Survey Sampling*. Springer, New York, 1992

Anhang

A Fragebogen der F+E 2021



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Erhebung online: www.esurvey.admin.ch/eRD

Benutzername:

Passwort:

OID	NOGA	SAB
Bei Rückfragen und Korrespondenz bitte diese Nummer (OID) angeben.		

Monetäre Unternehmensstatistik (MON)

Espace de l'Europe 10
CH-2010 Neuchâtel

Auskünfte und Mitteilungen:

Tel. 0800 20 10 10 (gratis Nummer)
E-Mail: mon.rd@bfs.admin.ch

Identifikation der Unternehmen mit Aufwendungen für Forschung und Entwicklung (F+E) im Jahr 2021



Bitte benutzen Sie einen schwarzen oder blauen Kugelschreiber und Grossbuchstaben.

M U S T E R

Setzen Sie die Markierung (X) genau ins Kästchen ein.



Allgemeine Informationen

Forschung und experimentelle Entwicklung (F+E) umfassen auf systematischer Basis durchgeführte schöpferische Arbeiten mit dem Ziel, den Bestand an Wissen (z.B. über den Menschen, die Kultur und die Gesellschaft) zu erhöhen und dieses Wissen zur Entwicklung neuer Anwendungen zu nutzen (siehe Beispiele auf der Rückseite).

Auskunftsperson in Ihrem Unternehmen für zusätzliche Informationen:

Name / Vorname:

E-Mail:

Telefon: / - -

Forschungs- und Entwicklungsarbeiten im Jahr 2021

Hatte Ihr Unternehmen im Jahr 2021 Forschungs- und Entwicklungsaufwendungen in einer oder mehreren der unten aufgeführten Kategorien getätigt oder geplant, solche zu tätigen?

- Für eigene Zwecke oder im Auftrag Dritter durchgeführte F+E (Intramuros F+E).
- F+E an Dritte vergeben (Extramuros F+E).
- Für eigene Zwecke oder im Auftrag Dritter durchgeführte F+E in Ihren Arbeitsstätten und/oder Zweigunternehmen im Ausland.

☐ Ja ☐ Nein

Wenn ja, in welcher Sprache wünschen Sie den Fragebogen zu F+E?

☐ Deutsch ☐ Französisch ☐ Italienisch ☐ Englisch

Zustelladresse für den Fragebogen zu F+E (wenn von der Adresse abweichend):

Betrieb:

Adresse:

PLZ/Ort:

Vielen Dank für Ihre wertvolle Mitarbeit!

Bitte wenden



Abbildung 1 F+E 2021: Screening Fragebogen, Seite 1

1

1 Intramuros-F+E-Aufwendungen des Unternehmens in der Schweiz, 2021 (in tausend Franken)					
a. F+E-Personalaufwendungen					
Löhne, Gehälter sowie alle dazugehörigen F+E-Personalkosten und Nebenleistungen			210		
b. Andere laufende F+E-Aufwendungen					
umfassen Material, Mieten, Leasing usw. ohne Abschreibungen			220		
c. F+E-Investitionsausgaben (ohne Abschreibungen) (=246 + 248 + 249)					
davon:					
Investitionsausgaben für Gebäude			246		
Investitionsausgaben für Software			248		
Investitionsausgaben für Anlagen und Ausstattung			249		
Total der Intramuros-F+E-Aufwendungen des Unternehmens in der Schweiz, 2021			245		
Siehe Definitionen in der Beilage I, S. 2.		(= Summe der Positionen 210 + 220 + 244)			
2 Verwendung des Resultats der F+E-Aktivitäten des Unternehmens in der Schweiz, 2021 (Schätzung in %)					
In welchem Wirtschaftszweig/welchen Wirtschaftszweigen findet das Resultat der F+E-Aktivitäten des Unternehmens Verwendung?					
Bitte geben Sie die prozentuale Verteilung der Intramuros-F+E-Aufwendungen für die betreffenden Wirtschaftszweige an (in % von Position 245, Rubrik 1)					
Nahrungsmittel		3001	Hochtechnologieinstrumente		3006
Chemie		3002	IKT-Herstellung		3007
Pharma		3003	IKT-Dienstleistungen		3008
Metall		3004	Andere Wirtschaftszweige		3010
Maschinen		3005	Total		
Siehe Definitionen in der Beilage II.					
3 Verteilung der Intramuros-F+E-Aufwendungen des Unternehmens in der Schweiz, 2021 (Schätzung in %)					
3.1 Aufteilung (in %) der Intramuros-F+E-Aufwendungen nach Technologiebereich (in % von Position 245, Rubrik 1)					
Biotechnologie		500	Software		520
Nanotechnologie		510	Andere Forschung		590
Siehe Definitionen in der Beilage III		Total			
3.2 Aufteilung (in %) der Intramuros-F+E-Aufwendungen nach Art der F+E (in % von Position 245, Rubrik 1)					
Grundlagenforschung		331	Experimentelle Entwicklung		340
Angewandte Forschung		335	Total		
Siehe Definitionen und Beispiele in der Beilage I, S. 2 und 3.					
3.3 Aufteilung (in %) der Intramuros-F+E-Aufwendungen nach Ziel der F+E (in % von Position 245, Rubrik 1)					
Gesundheit		365	Industrielle Produktion und Technologie		368
Landwirtschaft		367	Verteidigung		375
Umwelt		364	Andere Ziele		397
Energie		366	Total		
Siehe Definitionen und Beispiele in der Beilage IV.					



Abbildung 3 F+E 2021: Haupterhebung Fragebogen, Seite 2

4 Finanzierung der im Unternehmen durchgeführten Intramuros-F+E in der Schweiz, 2021 (in tausend Franken)			
Von wem wurden die Mittel zur Finanzierung der Intramuros-F+E-Aktivitäten des Unternehmens in der Schweiz im Jahr 2021 aufgebracht?			
a. Finanzierung durch das Unternehmen selbst (einschliesslich auf dem Finanzmarkt u. bei den Banken aufgenommene Mittel)			
			415
b. Finanzierung aus der Schweiz von:			
	F+E-Aufträge	F+E-Beiträge	
anderen Unternehmen derselben Gruppe (Mutterhaus und/oder Zweigunternehmen)			4201
anderen, kapitalmässig nicht verbundenen Privatunternehmen			4251
der öffentlichen Hand (z.B. Innosuisse)			4351
Hochschulen (Universitäten, Fachhochschulen, ETH, eidg. Forschungsanstalten im ETH-Bereich)			4401
anderen Quellen (z.B. private Organisationen ohne Erwerbszweck)			4501
> Total der Aufträge und der Beiträge aus der Schweiz			4051
c. Finanzierung aus dem Ausland von:			
	F+E-Aufträge	F+E-Beiträge	
anderen Unternehmen derselben Gruppe (Mutterhaus und/oder Zweigunternehmen)			4209
anderen, kapitalmässig nicht verbundenen Privatunternehmen			4259
anderen Quellen (z.B. öffentliche Hand, Hochschulen, private Organisationen ohne Erwerbszweck, internationale Organisationen)			4509
der Europäischen Kommission			4609
> Total der Aufträge und der Beiträge aus dem Ausland			4559
Total der Aufträge und Beiträge			4055
		(= Summe der Positionen 4051 + 4559)	
Total der Finanzierung der Intramuros-F+E (= Position 245, Rubrik 1)			405
		(= Summe der Positionen 4051 + 4559)	
Siehe Definitionen in der Beilage I, S. 3.			
5 Extramuros-F+E-Aufwendungen des Unternehmens, 2021 (in tausend Franken)			
Hat das Unternehmen im Jahr 2021 F+E-Aufträge vergeben, F+E-Projekte von Dritten mit F+E-Beiträgen unterstützt?			
Ja ☐ Bitte erfüllen. Nein ☐ Bitte weiter zu Rubrik 6.			
An wen hat das Unternehmen im Jahr 2021 Extramuros-F+E-Aufwendungen vergeben?			
a. Extramuros-F+E-Aufwendungen in der Schweiz vergeben an:			
	F+E-Aufträge	F+E-Beiträge	
andere Unternehmen derselben Gruppe (Mutterhaus und/oder Zweigunternehmen)			2551
andere, kapitalmässig nicht verbundene Privatunternehmen			2501
öffentliche Einrichtungen (z.B. landwirtschaftliche Forschungsanstalten)			2651
Hochschulen in der Schweiz (Universitäten, Fachhochschulen, ETH, Forschungsanstalten des ETH-Bereiches)			2601
übrige Empfänger (z.B. private Organisationen ohne Erwerbszweck)			2751
> Total der Extramuros-F+E-Aufwendungen in der Schweiz			2911
b. Extramuros-F+E-Aufwendungen im Ausland vergeben an:			
	F+E-Aufträge	F+E-Beiträge	
andere Unternehmen derselben Gruppe (Mutterhaus und/oder Zweigunternehmen)			2749
andere, kapitalmässig nicht verbundene Privatunternehmen			2779
übrige Empfänger im Ausland (z.B. öffentliche Einrichtungen, Hochschulen, private Organisationen ohne Erwerbszweck)			2789
> Total der Extramuros-F+E-Aufwendungen im Ausland			2769
Total der Aufträge und der Beiträge			2055
		(= Summe der Positionen 2911 + 2769)	
Total der Extramuros-F+E-Aufwendungen			295
		(= Summe der Positionen 2055 + 2955)	



Abbildung 4 F+E 2021: Haupterhebung Fragebogen, Seite 3

Ausbildung (höchste abgeschlossene Ausbildung)	Total beschäftigte Personen in F+E (Anzahl)	Davon: Frauen (Anzahl)	Davon: ausländische Personen (Anzahl)	Total in Vollzeit-äquivalenten in F+E (VZÄ)	
Tertiärstufe, Hochschulen					640
– davon Doktorat oder gleichwertige Diplome					630
Tertiärstufe, höhere Berufsausbildung					665
Andere Ausbildung					699
Total internes F+E-Personal in der Schweiz					600

Funktion	Total beschäftigte Personen in F+E (Anzahl)	Davon: Frauen (Anzahl)	Davon: ausländische Personen (Anzahl)	Total in Vollzeit-äquivalenten in F+E (VZA)	
Forscher/innen					715
F+E-Techniker/innen					725
F+E-Hilfspersonal (oder nicht feststellbare Funktion)					735
Total internes F+E-Personal in der Schweiz					705

(= Position 600, Rubrik 6)

Besass das Unternehmen 2021 Zweigunternehmen im Ausland, die auf eigene Rechnung oder im Auftrag Dritter F+E-Arbeiten durchgeführt haben?					
Ja	☐	Bitte ausfüllen.	Nein	☐	
Total Beschäftigte in den Zweigunternehmen im Ausland am 31.12.2021 (Anzahl)				▼	115
Total Intramuros-F+E-Aufwendungen der Zweigunternehmen im Ausland im Jahr 2021 (in tausend Franken)				▼ ▼	205
Total internes F+E-Personal der Zweigunternehmen im Ausland im Jahr 2021 (Anzahl)				▼	204

Wieviel Zeit haben Sie für das Ausfüllen benötigt?

Stunden	Minuten

900

Wir danken Ihnen herzlich für Ihre Teilnahme an dieser Erhebung!



Abbildung 5 F+E 2021: Haupterhebung Fragebogen, Seite 4

Methodenberichte der Sektion Statistische Methoden des BFS
Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodology reports published by the FSO's Statistical Methods Section

Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0086

Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS: 338-0085

Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS: 338-0084

Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS: 338-0083

Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS: 338-0082

Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS: 338-0081

Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer: 338-0080

Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015: Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0079

Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS: 338-0078

Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics. FSO number: 338-0077

Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. BFS-Nummer: 338-0076

Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. BFS-Nummer: 338-0073

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009: Statistische Datenaufbereitung und Hochrechnung. BFS-Nummer: 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro OFS: 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009: Stichprobenrahmen und Stichprobenplan. BFS-Nummer: 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007: Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010: Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO number: 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro OFS: 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. BFS-Nummer: 338-0064

- Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0063
- Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0062
- Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro OFS: 338-0061
- Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO number: 338-0060
- Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro OFS: 338-0059
- Graf, E. (2009). Weightings of the Swiss Household Panel: SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. FSO number: 338-0058
- Graf, E. (2009). Pondérations du Panel Suisse de Ménages: PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro OFS: 338-0057-05
- Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro OFS: 338-0056
- Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro OFS: 338-0055
- Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro OFS: 338-0054
- Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires SSalarium". Numéro OFS: 338-0053
- Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro OFS: 338-0052
- Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro OFS: 338-0051
- Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer: 338-0050
- Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro OFS: 338-0049
- Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. BFS-Nummer: 338-0048
- Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. BFS-Nummer: 338-0047
- Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer: 338-0046
- Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro OFS: 338-0045
- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro OFS: 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer: 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. BFS-Nummer: 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS: 338-0041

- Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro OFS: 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS: 338-0039
- Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. FSO number: 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer: 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer: 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer: 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer: 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer: 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS: 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS: 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer: 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer: 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number: 338-0027
- Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. BFS-Nummer: 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS: 338-0025
- Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS: 338-0024
- Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro OFS: 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS: 338-0022
- Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. BFS-Nummer: 338-0021
- Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer: 338-0020.
- Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro OFS: 338-0019
- Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. BFS-Nummer: 338-0018

- Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro OFS: 338-0017
- Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. BFS-Nummer: 338-0016
- Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro OFS: 338-0015
- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. BFS-Nummer: 338-0014
- Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro OFS: 338-0013
- Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro OFS: 338-0012
- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer: 338-0011
- Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro OFS: 338-0010
- Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro OFS: 338-0009
- Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. BFS-Nummer: 338-0008
- Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro OFS: 338-0007
- Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro OFS: 338-0006
- Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro OFS: 338-0005
- Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. BFS-Nummer: 338-0004
- Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro OFS: 338-0003
- Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. BFS-Nummer: 338-0002
- Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro OFS: 338-0001
- Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. BFS-Nummer: 338-0000

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat als zentrale Statistikstelle des Bundes die Aufgabe, statistische Informationen zur Schweiz breiten Benutzerkreisen zur Verfügung zu stellen. Die Verbreitung geschieht gegliedert nach Themenbereichen und mit verschiedenen Informationsmitteln über mehrere Kanäle.

Die statistischen Themenbereiche

- 00 Statistische Grundlagen und Übersichten
- 01 Bevölkerung
- 02 Raum und Umwelt
- 03 Arbeit und Erwerb
- 04 Volkswirtschaft
- 05 Preise
- 06 Industrie und Dienstleistungen
- 07 Land- und Forstwirtschaft
- 08 Energie
- 09 Bau- und Wohnungswesen
- 10 Tourismus
- 11 Mobilität und Verkehr
- 12 Geld, Banken, Versicherungen
- 13 Soziale Sicherheit
- 14 Gesundheit
- 15 Bildung und Wissenschaft
- 16 Kultur, Medien, Informationsgesellschaft, Sport
- 17 Politik
- 18 Öffentliche Verwaltung und Finanzen
- 19 Kriminalität und Strafrecht
- 20 Wirtschaftliche und soziale Situation der Bevölkerung
- 21 Nachhaltige Entwicklung, regionale und internationale Disparitäten

Die zentralen Übersichtspublikationen

Statistisches Jahrbuch der Schweiz



Das vom Bundesamt für Statistik (BFS) herausgegebene Statistische Jahrbuch ist seit 1891 das Standardwerk der Schweizer Statistik. Es fasst die wichtigsten statistischen Ergebnisse zu Bevölkerung, Gesellschaft, Staat, Wirtschaft und Umwelt des Landes zusammen.

Taschenstatistik der Schweiz



Die Taschenstatistik ist eine attraktive, kurzweilige Zusammenfassung der wichtigsten Zahlen eines Jahres. Die Publikation mit 52 Seiten im praktischen A6/5-Format ist gratis und in fünf Sprachen (Deutsch, Französisch, Italienisch, Rätoromanisch und Englisch) erhältlich.

Das BFS im Internet – www.statistik.ch

Das Portal «Statistik Schweiz» bietet Ihnen einen modernen, attraktiven und stets aktuellen Zugang zu allen statistischen Informationen. Gerne weisen wir Sie auf folgende, besonders häufig genutzte Angebote hin.

Publikationsdatenbank – Publikationen zur vertieften Information

Fast alle vom BFS publizierten Dokumente werden auf dem Portal gratis in elektronischer Form zur Verfügung gestellt. Gedruckte Publikationen können bestellt werden unter der Telefonnummer +41 58 463 60 60 oder per Mail an order@bfs.admin.ch.
www.statistik.ch → Statistiken finden → Kataloge und Datenbanken → Publikationen

NewsMail – Immer auf dem neusten Stand



Thematisch differenzierte E-Mail-Abonnemente mit Hinweisen und Informationen zu aktuellen Ergebnissen und Aktivitäten.
www.news-stat.admin.ch

STAT-TAB – Die interaktive Statistikdatenbank



Die interaktive Statistikdatenbank bietet einen einfachen und zugleich individuell anpassbaren Zugang zu den statistischen Ergebnissen mit Downloadmöglichkeit in verschiedenen Formaten.
www.stattab.bfs.admin.ch

Statatlas Schweiz – Regionaldatenbank und interaktive Karten



Mit über 4500 interaktiven thematischen Karten bietet Ihnen der Statistische Atlas der Schweiz einen modernen und permanent verfügbaren Überblick zu spannenden regionalen Fragestellungen aus allen Themenbereichen der Statistik.
www.statatlas-schweiz.admin.ch

Individuelle Auskünfte

Zentrale Statistik Information

+41 58 463 60 11, info@bfs.admin.ch

Die Erhebung des Bundesamtes für Statistik zur Forschung und Entwicklung in der schweizerischen Privatwirtschaft (F+E) wird seit 2015 jedes zweite Jahr durchgeführt.

Sie besteht aus zwei Phasen: einem Screening zur Identifikation der Unternehmen mit F+E-Ausgaben und der Haupterhebung zur Ermittlung der eigentlichen Ausgaben.

Dieser Methodenbericht beschreibt die wichtigsten methodischen Aspekte der beiden Phasen. Ein Hauptaugenmerk wird dabei auf die Neukonzeption des Screenings gelegt, wobei für die Erhebung 2021 die bisherige Vollerhebung durch eine ressourcensparende Alternative basierend auf einer Stichprobe abgelöst wurde.

Aus der Grundgesamtheit (ca. 22 000 vorwiegend grössere Unternehmen in klar definierten Branchen) wurden dabei in der ersten Phase rund 14 000 Unternehmen gezogen. Von den gezogenen Unternehmen haben über 90% geantwortet, wobei aber nur rund jedes fünfte Unternehmen angab, F+E-Ausgaben zu tätigen.

In der zweiten Phase haben etwas mehr als 60% dieser Unternehmen (ca. 1650) Angaben zu den effektiven Ausgaben mitgeteilt.

Die Gewichtung wurde an die Tatsache angepasst, dass es sich bei der ersten Phase der Erhebung nicht mehr um eine Vollerhebung handelte. Es wurden verschiedene Gewichtungsvarianten für die Korrektur von Antwortausfällen getestet, darunter eine beruhend auf einer logistischen Regression. Aus Gründen der Vergleichbarkeit wurde die diesbezügliche Methodik schlussendlich aber nicht verändert.

Im Rahmen der F+E gibt es erfahrungsgemäss einige wenige Unternehmen mit aussergewöhnlich hohen Ausgaben und entsprechend starkem Einfluss auf die Resultate. Um die Vergleichbarkeit der Resultate über die Zeit zu gewährleisten, wurde eine systematische Identifikation dieser Unternehmen basierend auf vorangehenden Erhebungswerten eingeführt. Dies erlaubt es, diese Unternehmen sowohl im Stichprobenplan, als auch in der Datenerhebung und der Hochrechnung gesondert zu behandeln.

Am Ende des Berichts wird die Varianzschätzung beschrieben, welche dazu dient, die Genauigkeit der im Rahmen der F+E erstellten Schätzungen beurteilen zu können und so auch hilft, die Resultate oder Entwicklungen besser interpretieren zu können.

Online

www.statistik.ch

Print

www.statistik.ch
Bundesamt für Statistik
CH-2010 Neuchâtel
order@bfs.admin.ch
Tel. +41 58 463 60 60

BFS-Nummer

338-0086

ISBN

978-3-303-00727-3

Die Informationen in dieser Publikation tragen zur Messung der Ziele für nachhaltige Entwicklung (SDG) bei.



Indikatorensystem MONET 2030

www.statistik.ch → Statistiken finden → Nachhaltige Entwicklung → Das MONET 2030-Indikatorensystem

**Statistik
zählt für Sie.**

www.statistik-zaehlt.ch