Auswertung und Präsentation der Ergebnisse einer Befragung

Rolf Porst – unter Mitarbeit von Ruth Holthof[1]

Sie lesen jetzt den dreizehnten Beitrag zur Serie „Schüler forschen“. Wir hatten bisher aufgezeigt, wie eine Befragung in schulischen Forschungsprojekten vorbereitet und durchgeführt werden kann und wie die Daten aus den Fragebogen in einen auswertbaren Datensatz eingelesen werden können. Jetzt geht es darum, die Daten auszuwerten und zu präsentieren. Ohne dem Weiteren vorgreifen zu wollen, bereits an dieser Stelle die Empfehlung: Verwenden Sie zur Analyse der Daten einfache Verfahren, die von den Adressaten Ihrer Präsentation auch nachvollzogen und verstanden werden können. Verzichten Sie auf den Einsatz komplexer Analyseverfahren, auch wenn Sie oder Ihre SchülerInnen solche Verfahren beherrschen. Hinterfragen Sie, ob Ihre LeserInnen, ZuhörerInnen oder ZuschauerInnen komplexe Analysen und deren Darstellung verstehen würden. Sie führen ein schulisches, kein akademisches Forschungsprojekt durch, Sie verfassen darüber am Ende eine interne Darstellung, allenfalls einen Bericht im Rahmen eines schulischen Wettbewerbs, aber keine sozialwissenschaftliche Dissertation.

Zur Erinnerung: Skalenniveaus

Die Auswertungsverfahren, die Sie bei der Datenanalyse zum Einsatz bringen können, hängen vom Skalenniveau der einzelnen Fragen ab. Deshalb, zur Erinnerung, hier ein kurzer Blick auf die vier Skalenniveaus (siehe auch Porst 2015: 245f):

Rolf Porst
Markt- und Sozialforscher, Römerberg

Nominalskala: A ≠ B; die Merkmale von Objekten können lediglich im Hinblick auf Gleichheit oder Ungleichheit untersucht werden, z.B. Geschlecht (männlich *oder* weiblich), Zustimmung zu einer Aussage (ja *oder* nein) oder Lieblingssport (Fußball *oder* Handball *oder* Tennis).

Ordinalskala: A<B<C; die Merkmale von Objekten weisen eine Rangordnung auf, z.B. Schulnoten oder differenzierte Zustimmung zu einer Aussage (stimme voll und ganz zu – stimme eher zu – stimme eher nicht zu – stimme überhaupt nicht zu).

Intervallskala: wenn A, B, C, D aufeinanderfolgen, gilt B-A = D-C; die Merkmalsausprägungen weisen nicht nur eine Rangfolge auf, sondern die Abstände zwischen den Merkmalsausprägungen sind gleich, z.B. Temperaturskala oder die folgende Skala:

trifft überhaupt nicht zu ❑ ❑ ❑ ❑ ❑ ❑ ❑ trifft voll und ganz zu

Ratioskala: entspricht einer Intervallskala, hat aber einen *echten Nullpunkt*, z.B. Einkommen, Zahl der Kinder im Haushalt.

Wichtig: Je höher das Skalenniveau, umso mehr und umso komplexere statistische Auswertungsverfahren sind möglich und zulässig.

Betrachten wir jetzt aber eher einfache Analyseverfahren; beginnen wir mit univariaten Analysen, also der Auswertung der Daten zu einzelnen Fragen.

Univariate Analysen – Häufigkeitsverteilungen

Die einfachste Form der univariaten Datendarstellung ist die *Häufigkeitsverteilung*. Wie viele unserer Befragten haben eine bestimmte Frage mit „Ja“ beantwortet, wie viele mit „Nein“? Wie viele unserer Befragten haben als Geburtsjahr 1952 angegeben, 1953, 1954 usw. Sie sehen an diesen Beispielen: Bei der Häufigkeitsverteilung spielt das Skalenniveau keine Rolle.

Nehmen wir zur weiteren Illustration eine Befragung von 400 Personen. Bei der Frage nach der Anzahl der Kinder unter 18 Jahren im Haushalt erhalten wir das folgende Ergebnis, dargestellt in einer einfachen Häufigkeitsverteilung (Tab.1):

Tab. 1: Einfache Häufigkeitsverteilung

Kinder unter 18 Jahren im Haushalt	Häufigkeiten absolut	Häufigkeiten relativ (%)
0	160	40
1	88	22
2	64	16
3	56	14
4	16	4
5	0	0
6	0	0
7	8	2
8	8	2
	Σ = 400	Σ = 100%

Tab. 2: Gruppierte Häufigkeitsverteilung

Kinder unter 18 Jahren im Haushalt	Häufigkeiten absolut	Häufigkeiten relativ (%)
0	160	40
1 - 2	152	38
3 - 5	72	18
6 oder mehr	16	4
	Σ = 400	Σ = 100%

Wir ersehen aus *Tabelle 1*, wie viele der BefragungsteilnehmerInnen angegeben haben, sie hätten 0, 1, 2, 3 usw. Kinder unter 18 Jahren im Haushalt. Die Prozentualisierung erfolgt auf der Basis aller 400 Befragungspersonen.

Bei Variablen mit vielen Ausprägungen ist die Darstellung der Ergebnisse häufig übersichtlicher und das Resultat besser zu erfassen, wenn man statt einfacher Häufigkeitsverteilungen *gruppierte Häufigkeitsverteilungen* zum Einsatz bringt (Tab. 2).

In *Tabelle 2* wird die Anzahl der Kinder in Gruppen zusammengefasst ausgewiesen. Gruppierte Häufigkeitsverteilungen haben natürlich den Nachteil, dass Informationen verloren gehen, in diesem Falle eben die genaue Verteilung der Kinder innerhalb der neuen Kategorien. Aber andererseits: Stellen Sie sich eine Tabelle vor, in der das Geburtsjahr dokumentiert wird: für alle Jahrgänge zwischen, sagen wir, 1925 und 1995, also insgesamt 70 Ausprägungen der Variablen „Geburtsjahr". Dann drängt sich eine Gruppierung geradezu auf.

Noch anschaulicher lassen sich die Daten aus den beiden Tabellen in grafischer Form darstellen. Eine sehr einfache und weithin bekannte Form grafischer Darstellung ist das sogenannte *Balkendiagramm*, hier in seiner einfachsten Form:

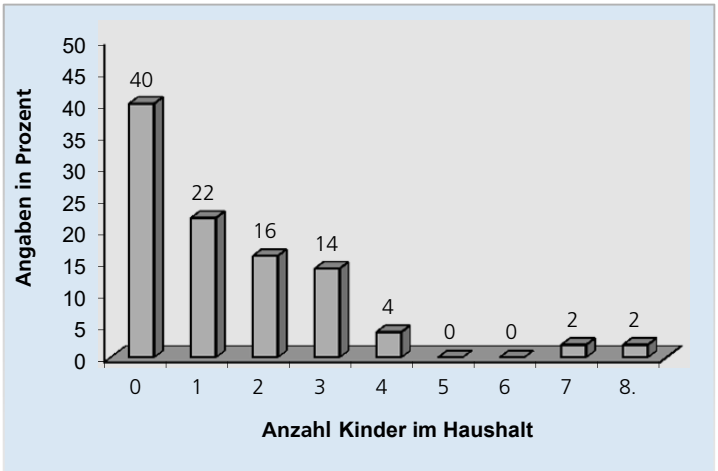

Abb. 1: Balkendiagramm

Was das Layout angeht, gibt es beim Balkendiagramm kaum Grenzen. Microsoft Office bietet derzeit 19 Varianten des Balkendiagramms (dort als „Säulen" bezeichnet) an, darüber hinaus noch 15 Varianten von „Balken", worunter liegende Säulen fallen. Auch Grafstat (vgl. Porst 2016) kennt im Rahmen der „grafischen Auswertung" Balken- und Säulendiagramme.

Neben dem geläufigen Balkendiagramm gibt es eine Vielzahl anderer Darstellungsformen grafischer Auswertung, bei Microsoft Office zum Beispiel Linien, Ringe, Blasen oder Netze. Auch Grafstat bietet eine größere Anzahl vorgefertigter grafischer Auswertungsmöglichkeiten an. Relativ verbreitet ist das Kreisdiagramm (in der Terminologie von Grafstat) oder der Kreis (bei Microsoft Office), umgangssprachlich auch als Kuchendiagramm bezeichnet:

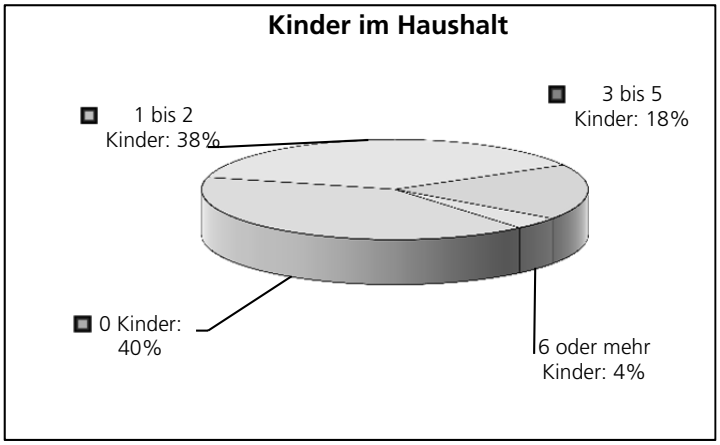

Abb. 2: Kuchendiagramm

Wie beim Balkendiagramm gibt es auch beim „Kuchen" eine Vielzahl von Variationsmöglichkeiten.

Ganz allgemein können wir festhalten: Die Entscheidung für die eine oder andere Form der grafischen Darstellung sollte schlicht davon abhängig gemacht werden, welche am besten geeignet ist, das vorzustellende Ergebnis möglichst optimal und zielgruppengenau zu transportieren.

Aber was heißt „am besten geeignet"? Geben Sie diese Frage doch in den Unterricht. Lassen Sie Ihre SchülerInnen die verschiedenen grafischen Darstellungsformen praktisch, z.B. in Partnerarbeit, erproben, und diskutieren Sie anschließend mit ihnen über die Vor- und Nachteile der jeweiligen Darstellung. Integrieren Sie dabei eine Wiederholung der Methode „Analyse einer Statistik"; auch hierzu benötigen Ihre SchülerInnen ein Gespür für bzw. Wissen über passende und nicht-passende grafische Darstellung von Informationen. Ihre SchülerInnen werden auf diese Weise besser nachvollziehen können, dass die Darstellungsform erheblichen Einfluss auf die Wahrnehmung und Beurteilung von Informationen haben kann.

Während die Entscheidung für die eine oder andere Hauptdiagrammform (Säulen, Balken, Kuchen usw.) also unter dem Gesichtspunkt der optimalen Kommunikation mit den LeserInnen oder ZuhörerInnen erfolgen sollte, ist die Ausarbeitung im Detail eher eine Frage der Ästhetik. Sie können, je nach Fragestellung, in Ihren Texten unterschiedliche Hauptdiagrammformen einsetzen, aber lassen Sie alle Säulendiagramme oder alle Kuchen über den gesamten Text vom Layout her gleich aussehen. Das erleichtert das Lesen und Verstehen der Grafiken ungemein.

Univariate Analysen – Verteilungsmaße

Absolute und relative Häufigkeiten, wie wir sie gerade in den Tabellen dargestellt hatten, sind als Verteilungsmaße für Fragen (im Prinzip) aller Skalenniveaus einsetzbar; bei nominalskalierten Variablen sind sie auch die einzig zulässige, auf jeden Fall aber zulässige „vernünftige" Darstellungsform. Weitergehende Darstellungsformen setzen höheres Skalenniveau voraus.

Reine Prozentuierungen, so meine Empfehlung, verwenden Sie am besten erst dann, wenn die absolute Zahl Ihrer Befragungspersonen 100 überschreitet. Haben Sie weniger als 100 Fälle, würde eine einzelne Person mit mehr als 1 Prozent in die Prozentuierung fallen. Stellen Sie sich vor, Sie hätten nur 38 Befragungspersonen: Dann wären zwei davon 5,3%. Nur diesen Prozentwert darzustellen, wäre verwirrend. Arbeiten Sie bei weniger als 100 Befragungspersonen lieber mit absoluten Häufigkeiten oder immer in der direkten Kombination: 5,3% (also 2 von 38 Befragten) oder nach dieser erstmaligen Anordnung gleich mit 5,3% (2 von 38). Oder so ähnlich.

Bei Fragen, die mindestens ordinal skaliert sind, macht es dagegen häufig viel Sinn, mit den *„Maßen der zentralen Tendenz"* zu arbeiten. Unter diesen Begriff fallen der *Modus*, der *Median* und der *Mittelwert* (auch: arithmetisches Mittel). Streng genommen sollte man die Maße der zentralen Tendenz nicht auf ordinalskalierte Variablen anwenden, aber da wir in unseren Befragungen sehr häufig Variablen genau diesen Skalenniveaus haben, wurde viel Wert darauf gelegt aufzuzeigen, dass man ordinalskalierte Variablen beim Auswerten wie intervallskalierte behandeln kann (z.B. Borgatta & Bohrnstedt, 1980 oder Kampen & Swyngedouw, 2000). Was wir alle auch tun.

Betrachten wir die Maße der zentralen Tendenz nun etwas näher. Beginnen wir mit dem Modus.

Der Modus

Der Modus ist derjenige Wert, der in der Verteilung am häufigsten vorkommt. Betrachten wir das folgende Beispiel, in dem 100 Befragungspersonen irgendeinen Gegenstand auf einer Wichtigkeitsskala bewertet haben; in der oberen Reihe finden wir die Skalenpunkte, in der unteren Reihe die Anzahl der Nennungen für jeden dieser Skalenpunkte.

Überhaupt nicht wichtig	1	2	3	4	5	6	7	Sehr wichtig
Absolute Häufigkeiten	12	10	15	22	16	8	17	N = 100

Der Skalenpunkt mit den häufigsten Nennungen (N = 22) ist der Wert 4; also liegt der Modus dieser Verteilung beim Wert 4.

Den Modus könnten Sie auch für nominalskalierte Variablen ausweisen, aber das erscheint nicht wirklich sinnhaft. Wenn Sie in Ihrer Umfrage z.B. die bekannte Frage nach der Wahlabsicht bei der nächsten Bundestagswahl (die sog. Sonntagsfrage) stellen, können Sie natürlich angeben, dass der Modus bei dieser Frage mit 41 von 100 Nennungen bei der CDU liegt, aber Sie können natürlich genauso gut und verständlicher festhalten, dass die CDU lt. Ihrer Umfrage bei der nächsten Bundestagswahl mit 41 Prozent der gültigen Stimmen rechnen kann.

Der Modus ist in breiteren Öffentlichkeiten nicht bekannt. Darüber hinaus bringt der Modus Probleme bei Verteilungen, die zwei oder gar mehr als zwei Modi haben. Von daher empfehlen wir Ihnen die Verwendung des Modus eher nicht. Da ist der Median schon besser geeignet, um Verteilungen zu beschreiben.

Der Median

Der Median ist das „klassische“ Maß für ordinalskalierte Variablen. Er teilt die Verteilung in zwei gleiche Hälften, ist also derjenige Wert, auf dem sich der N/2 Wert befindet. Um ihn zu erkennen, müssen wir die Werte über die gesamte Skala kumulieren. Nehmen wir wieder das Beispiel von oben:

Überhaupt nicht wichtig	1	2	3	4	5	6	7	Sehr wichtig
Absolute Häufigkeiten	12	10	15	22	16	8	17	N = 100
Kumulierte absolute Häufigkeiten		22	37	59	75	83	100	

In der Zeile „kumulierte absolute Häufigkeiten“ werden die Häufigkeiten der Nennungen für die einzelnen Skalenpunkte von links nach rechts aufaddiert: 12 Nennungen auf Skalenwert 1 plus 10 Nennungen auf Skalenwert 2 ergeben den Wert 22 als ersten Wert dieser Zeile; zu diesem Wert 22 addieren wir die 15 Nennungen für Skalenwert 3 und erhalten so den Wert 37 als zweiten Wert dieser Zeile, usw.

Der Median liegt augenscheinlich auf dem Skalenwert 4, weil der N/2 Wert, also der 50. Fall, auf diesen Skalenwert fällt. Es reicht vollkommen aus, den Median auf diese Weise zu bestimmen, aber es gibt auch eine Formel dazu, die in unserem Beispiel den Median auf den Wert 4,1 präzisieren würde.

Wenn wir oben geschrieben hatten, der Median sei ein „klassischer“ Wert für ordinalskalierte Variablen, dann gilt das nur so lange, bis man nicht auch ordinalskalierte Daten mit dem arithmetischen Mittel beschreibt.

Der Mittelwert (das arithmetische Mittel)

Der Mittelwert ist das präziseste Maß der zentralen Tendenz und zugleich das am weitesten bekannte. Die Formel zur Berechnung des Mittelwertes setzen wir natürlich als bekannt voraus, in unserem Beispiel errechnen wir einen Mittelwert von 4,12.

Aber Vorsicht: Genau davor möchten wir ein wenig warnen. Wenn Sie einen Mittelwert angeben, sollten Sie ihn lieber auf *eine* Stelle hinter dem Komma begrenzen. Mehr als eine Stelle hinter dem Komma verspricht eine Präzision der Daten, die bei den normalerweise niedrigen Fallzahlen bei schulischen Umfragen (und nicht nur dort!) nicht gegeben ist.

Häufig unterschätzt wird beim arithmetischen Mittel das Risiko, dass der Mittelwert die Verteilung nicht wirklich korrekt beschreibt. Betrachten wir die Daten in den beiden folgenden extremen Beispielen:

Extrembeispiel 1:

Überhaupt nicht wichtig	1	2	3	4	5	6	7	Sehr wichtig
Absolute Häufigkeiten	50	0	0	0	0	0	50	N = 100

Extrembeispiel 2:

Überhaupt nicht wichtig	1	2	3	4	5	6	7	Sehr wichtig
Absolute Häufigkeiten	0	0	0	100	0	0	0	N = 100

Für beide Verteilungen ergibt sich ein Mittelwert von 4,0, aber es ist aus den absoluten Häufigkeiten ersichtlich, dass die Verteilungen völlig unterschiedlich sind. Die Angabe des reinen Mittelwertes macht das allerdings nicht sichtbar. „Echte" Daten führen sicherlich nicht so extrem zu diesem Effekt, aber das Risiko, dass Verteilungen auf der Basis des Mittelwertes falsch interpretiert werden, ist vorhanden.

In diesem Zusammenhang könnte es sich anbieten, mit Ihren SchülerInnen die Möglichkeit von Meinungsbeeinflussung, wenn nicht gar von Meinungsmanipulation durch Statistiken zu diskutieren. SchülerInnen (und nicht nur die!) gehen nun mal häufig davon aus, dass Zahlen objektiv und zuverlässig sind.

Aus diesem Grund muss (auf jeden Fall: sollte) man bei der Angabe eines Mittelwertes auch immer dessen *Standardabweichung* ausweisen. Die Standardabweichung ist die Wurzel aus der *Varianz*. Die Varianz wiederum ist die quadrierte Summe aller Abweichungen der einzelnen Messwerte vom gemessenen Mittelwert.

Die Berechnung der Varianz und der Standardabweichung nimmt uns – wie die Berechnung von Modus, Median und Mittelwert selbst – natürlich unser Auswertungsprogramm ab. Aber zum Verständnis doch ganz kurz die Grundlagen:

Der Mittelwert ist ja nichts anderes als die Summe der Multiplikation der Skalenwerte mit der Anzahl der Nennungen pro Skalenwert, dividiert durch die Anzahl der Fälle (hier: der Befragungspersonen). Im Beispiel…

Überhaupt nicht wichtig	1	2	3	4	5	6	7	Sehr wichtig
Absolute Häufigkeiten	12	10	15	22	16	8	17	N = 100

…berechnet sich der Mittelwert also als [(12 *1) + (10 *2) + (15 *3) + (22 +4) + (16 *5) + (6*8) + 17*7)]/100 = (12 + 20 + 45 + 88 + 80 + 48 + 119)/100 = 412/100 = 4,12.

Von diesem gemessenen Mittelwert weichen die Werte für die einzelnen Befragungspersonen mehr oder weniger stark ab; die quadrierte Summe der Abweichungen der einzelnen Werte vom Mittelwert ist die Varianz. Die Varianz nimmt den Wert 0 in dem – fast immer nur theoretischen – Fall an, in dem alle gemessenen Werte identisch sind, wenn also z.B. alle Personen auf der Skala den Wert 4 angeben. Dann beschreibt der Mittelwert die Verteilung absolut genau.

Im Beispiel der extremen Verteilungen oben würde dies dem Extrembeispiel 2 entsprechen. Die Varianz und damit die Standardabweichung belaufen sich auf 0.

Die Varianz und damit die Standardabweichung werden umso höher, je mehr gemessene Werte weit vom Mittelwert entfernt liegen. Je höher die Varianz und damit die Standardabweichung sind, umso ungenauer beschreibt der Mittelwert die Verteilung. Oder anders ausgedrückt: Eine hohe Standardabweichung ist der Beleg dafür, dass der gemessene Mittelwert das Antwortverhalten der befragten Personen nur schlecht wiedergibt.

Im Beispiel der extremen Verteilungen oben würde dies dem Extrembeispiel 1 entsprechen. Die Varianz liegt bei 3,0, die Standardabweichung bei 1,73. Bei 7 Skalenpunkten ist das schon ein ziemlich hoher Wert.

Bivariate Analysen – Kontingenztafeln (Kreuztabellen)

Kontingenztafeln oder *Kreuztabellen* stellen den Zusammenhang zwischen zwei Variablen mit nominalem Skalenniveau oder einer Variablen mit nominalem und einer Variablen mit ordinalem Skalenniveau dar; rein theoretisch könnte man sie auch für Va-

riablen mit höherem Skalenniveau berechnen, aber praktisch macht das wenig Sinn. Hier greift man dann besser auf andere Verfahren zurück.

Auch wenn Ihnen Kreuztabellen sehr wahrscheinlich vertraut sein werden, hier doch noch ein paar Regeln zu ihrer Konstruktion und zu ihrem Verständnis. Im folgenden (fiktiven) Beispiel geht es um die Frage, ob es einen Zusammenhang gibt zwischen Geschlecht und präferierter Jahreszeit. Gefragt wurde: „Welche Jahreszeit gefällt Ihnen alles in allem gesehen am besten – Frühling, Sommer, Herbst oder Winter?" Nach Abzug der Befragten, die sich nicht eindeutig äußern konnten oder die Frage gar nicht beantwortet hatten, verblieben in dem Beispiel insgesamt 2.298 Befragungspersonen, davon 1576 Männer und 722 Frauen.

Tab 3: Präferierte Jahreszeit nach Geschlecht

	Männer	**Frauen**	**gesamt**
Frühling	668	224	892
	(42,4)	(31,0)	(38,8)
Sommer	596	260	856
	(37,8)	(36,0)	(37,3)
Herbst	225	91	316
	(14,3)	(12,6)	(13,8)
Winter	87	147	234
	(5,5)	(20,4)	(10,2)
	1.576	722	2.298
	(100,0)	(100,0)	(100,0)

Wir erkennen in Tabelle 3 zunächst, dass die unabhängige (erklärende) Variable als Spaltenvariable darstellt wird, die abhängige (zu erklärende) Variable als Zeilenvariable. Die Präferenz für eine der Jahreszeiten wird in Abhängigkeit vom Geschlecht der befragten Personen untersucht. Präsentiert werden hier sowohl die absoluten als auch die relativen Häufigkeiten, die auf die Spaltensumme berechnet werden.

Von den 2.298 Befragten präferieren, das sehen wir in der Spalte „Gesamt", 892 oder 38,8% von N = 2.298 den Frühling, 856 (37,3%) den Sommer usw.

Genauso lesen wir die Zahlen für Männer und Frauen. Bei den Männern präferieren 668 oder 42,4% aller 1.576 befragter Männer den Frühling, bei den Frauen sind es 224 oder 31,0 Prozent aller 722 befragter Frauen. Wir prozentuieren also grundsätzlich nach der unabhängigen Variablen, die üblicherweise als Spaltenvariable ausgewiesen wird. Wenn wir die Werte für Männer und Frauen in der jeweiligen Spalte anschauen, erkennen wir, dass Männer den Frühling vor dem Sommer, dem Herbst und dem Winter präferieren – die Prozentwerte bezogen auf alle Männer fallen in dieser Reihenfolge ab. Bei den Frauen liegt der Sommer mit 36,0% vor dem Frühling mit 31,0%, dem Winter mit 20,4% und dem Herbst mit 12,6%. Wir lesen also spaltenweise die Reihenfolge der Präferenz ab.

Nun interessiert uns in dem Beispiel weniger, wie sich Männer und Frauen positionieren, sondern ob und wie sie sich voneinander unterscheiden. Dazu müssen wir die Prozentsätze in den Zeilen vergleichen. Im Beispiel sieht das so aus, dass in der Spalte „Gesamt" für den Frühling 38,8% aller Befragungspersonen stimmen. Wenn wir uns jetzt die „Frühlings"-Werte für Männer und Frauen ansehen, also die Prozentwerte in der Zeile „Frühling", erkennen wir, dass die Männer mit 42,4% oberhalb des Wertes für „Gesamt" rangieren, die Frauen mit 31,0% unterhalb dieses Wertes: Männer tendieren also eher zur Präferenz für den Frühling als Frauen. Für Sommer und Herbst

gibt es kaum Unterschiede, d.h. die Prozentwerte der Frauen unterscheiden sich von den Prozentwerten der Männer nicht stark, beide sind nahe dem Wert für Gesamt. Besonders auffällig ist die Bewertung des Winters: Nur 5,5% der Männer, aber 20,4% der Frauen präferieren den Winter.

Das Problem bei dieser wie bei jeder anderen Kreuztabelle auch ist die Tatsache, dass wir nicht erkennen, ob die Verteilung und das daraus resultierende Ergebnis für den Vergleich (im Beispiel von Männern und Frauen) *statistisch signifikant* sind. Oder anders ausgedrückt: Ob die gemessenen Zusammenhänge zwischen der unabhängigen Variablen (hier: Geschlecht) und der abhängigen Variablen (hier: Präferenz für eine Jahreszeit) nicht vielleicht zufällig zustande gekommen sind. Deshalb müssen wir bei Kreuztabellen deren statistische Signifikanz ermitteln (Aber keine Bange, auch das wird vom Auswertungsprogramm geleistet; ob Signifikanz besteht oder nicht, wird Ihnen durch das Programm ausgewiesen.).

Um aber zu verstehen, was hinter dem Begriff „statistische Signifikanz" steckt, müssen wir uns kurz mit dem Konzept der *statistischen Abhängigkeit* bzw. *Unabhängigkeit* beschäftigen.

Von statistischer *Unabhängigkeit* sprechen wir dann, wenn die bedingten Verteilungen gleich sind. Unterscheiden sie sich hingegen, sprechen wir von statistischer *Abhängigkeit*. Oder anders ausgedrückt: Statistische Abhängigkeit liegt vor, wenn es einen nachweisbaren Zusammenhang gibt zwischen abhängiger und unabhängiger Variable. Wir veranschaulichen das an einem Beispiel, der Einfachheit halber am Beispiel einer 2x2-Tabelle; es geht darum, ob sich Männer und Frauen bei ihrer Antwort auf eine bestimmte Ja/Nein-Frage unterscheiden:

a) Vollkommene Unabhängigkeit

	weiblich	**männlich**
ja	50	50
nein	50	50

$\phi = 0$

b) Schwacher Zusammenhang

	weiblich	**männlich**
ja	60	40
nein	40	60

$\phi = 0{,}2$

c) Sehr starker Zusammenhang

	weiblich	**männlich**
ja	80	20
nein	20	80

$\phi = 0{,}6$

d) Vollkommene Abhängigkeit

	weiblich	**männlich**
ja	100	0
nein	0	100

$\phi = 1{,}0$

Die beiden Variablen sind dann vollkommen *unabhängig* voneinander, wenn – das wäre das Bild a) – die Zellenbesetzungen vollkommen gleich sind; für das Antwortverhalten mit „ja" oder „nein" spielt das Geschlecht dann absolut keine Rolle. Der Zusammenhang zwischen den Variablen wäre *perfekt*, vollkommene Abhängigkeit bestünde, wenn – das wäre das Bild d) – alle Frauen mit „ja", alle Männer mit „nein" antworten würden. Zwischen diesen beiden Extremen finden wir abstufende Darstellungen, – das wären die Bilder b) und c) – die schwache oder starke Zusammenhänge anzeigen.

Der Zusammenhang zwischen den Variablen wird durch bestimme Maßzahlen beschrieben, in unserem Beispiel durch die Maßzahl ϕ, die besonders geeignet ist für 2x2-Tabellen. ϕ bewegt sich zwischen 0 und 1, es wird 0 bei vollkommener Unabhängigkeit und 1 bei vollkommener Abhängigkeit. Die Maßzahl ϕ basiert auf der grundlegenderen

Maßzahl Chi². Ohne näher darauf einzugehen: Chi² vergleicht für jede Zelle einer Tabelle die bei statistischer Unabhängigkeit zu erwartenden Werte mit den gemessenen Werten, ermittelt die Summe aller Abweichungen und vergleicht diese Summe mit dem Wert, der bei vollkommener Unabhängigkeit zwischen den Variablen einer entsprechenden Tabelle zu erwarten wäre.

Neben ϕ und Chi² gibt es noch eine ganze Reihe weiterer Assoziationsmaße, je nach Skalenniveau der Variablen, deren Zusammenhang überprüft werden soll. „Kluge" Auswertungsprogramme entscheiden aufgrund dieser Information, welche Assoziationsmaße überhaupt berechnet werden. Manche davon geben Ihnen nur an, dass ein statistischer Zusammenhang zwischen zwei Variablen besteht und dass dieser Zusammenhang „schwach", „stark" oder „sehr stark" ist, oft nur durch Symbole gekennzeichnet, z. B. schwach durch „+", stark durch „++" oder sehr stark durch „+++". Das Programm Grafstat liefert, zumindest nach unserer Kenntnis, keine Assoziationsmaße.

Bivariate Analysen – Korrelationsanalyse

In der *Korrelationsanalyse* wird die Stärke des Zusammenhangs zwischen zwei *mindestens ordinal* skalierten Variablen (eigentlich: *mindestens intervallskalierten* Variablen, siehe Abschnitt „Univariate Analysen – Verteilungsmaße") gemessen. Damit können Sie z.B. Ihre „je-desto-Hypothesen" (s. Porst 2014: 226ff) prüfen, also Hypothesen wie „Je älter eine Person ist (Variable „Alter"), umso stärker stimmt sie der Aussage zu, die in Deutschland lebenden Ausländer sollten sich ihre Ehepartner unter ihren eigenen Landsleuten aussuchen" (Variable „Diskriminierungsbereitschaft gegen Ausländer im Privatbereich"). Oder: „Je höher der formale Bildungsgrad einer Person ist (Variable „Schulabschluss"), desto geringer die Zustimmung zu der Aussage, man sollte Ausländer wieder in ihr Heimatland zurückschicken, wenn in Deutschland Arbeitsplätze knapp werden" (Variable „Diskriminierungsbereitschaft gegen Ausländer auf dem Arbeitsmarkt").

Üblicherweise – vereinfachend gesagt[2] – verwenden wir bei der Korrelationsanalyse den Korrelationskoeffizienten $r_{xy,}$ im Folgenden nur noch als r bezeichnet. Es gibt unterschiedliche Formeln zur Berechnung von r, wir arbeiten meist mit der folgenden Formel (vgl. Kühnel & Krebs 2001:404):

$$r_{xy} = \frac{n \times \sum x_i \times y_i - \left(\sum x_i\right) \times \left(\sum y_i\right)}{\sqrt{n \times \sum x_i^2 - \left(\sum x_i\right)^2} \times \sqrt{n \times \sum y_i^2 - \left(\sum xy_i\right)^2}}$$

Formel 1: Berechnung des Korrelationskoeffizienten r_{xy}

X und y sind die beiden Variablen, deren Zusammenhang wir prüfen wollen, und n ist die Anzahl der Messungen (Fälle) für beide Variablen.

Sieht komplizierter aus, als es ist. Im Grunde genommen brauchen Sie die Formel gar nicht zu kennen, weil Ihr Auswertungsprogramm Ihnen den Korrelationskoeffizienten r selbstverständlich ausgibt. Für das bessere Verständnis wollen wir dennoch folgendes Beispiel (siehe Tabelle 4) durchgehen. Var x ist die selbstberichtete Körpergröße von BefragungsteilnehmerInnen in cm, und Var y steht für ihr Körpergewicht in kg. Wir haben 9 Personen, die beide Angaben gemacht haben[3].

Tab 4: Daten zur Illustration der Berechnung des Korrelationskoeffizienten r

Fallnummer	Var x	Var y	$(x_i)^2$	$(y_i)^2$	$x_i * y_i$
1	185	80	34225	6400	14800
2	172	77	29584	5929	13244
3	187	86	34969	7396	16082
4	167	76	27889	5776	12692
5	180	88	32400	7744	15840
6	174	86	30276	7396	14964
7	181	84	32761	7056	15204
8	164	81	26896	6561	13284
9	183	89	33489	7921	16287
Σ	1593	747	282489	62179	132397

Setzen wir diese Werte in die obige Formel ein, so ergibt sich:

$$r_{xy} = \frac{9 \times 132397 - 1593 \times 747}{\sqrt{9 \times 282489 - (1593)^2} \times \sqrt{9 \times 62179 - (747)^2}}$$

Rechnen wir die Formel aus, erhalten wir den Wert $r_{xy} = 0{,}58$. Diesen Wert liefert uns das Auswertungsprogramm. Aber wie interpretieren wir den?

Der Korrelationskoeffizient kann Werte zwischen -1 und +1 annehmen. Der Wert -1 steht für einen perfekten *negativen* Zusammenhang zwischen zwei Variablen: Bei allen Befragungspersonen treten hohe Messwerte der einen Variablen stets mit niedrigen Messwerten der anderen Variablen auf. Der Wert +1 steht für einen perfekten *positiven* Zusammenhang zwischen den beiden Variablen: Bei allen Befragungspersonen treten hohe Messwerte der einen Variablen stets mit hohen Messwerten der anderen Variablen auf.

Der Wert 0 bedeutet, dass es keinerlei Zusammenhang zwischen beiden Variablen gibt[4].

Zur Bewertung des Korrelationskoeffizienten greifen wir auf folgende Faustregel zurück (siehe Kühnel & Krebs 2001:404f.):

0 bis .005	zu vernachlässigen
Über .005 bis .20	gering
Über .20 bis .50	mittel
Über .50 bis .70	hoch
Über .70	sehr hoch

Für unser Beispiel mit dem Wert 0,58 würden wir jetzt sagen, der Zusammenhang zwischen der Körpergröße und dem Gewicht unserer Befragungspersonen ist positiv und hoch. Je größer Personen sind, umso schwerer sind sie. Ein Wert von -.58 hätte einem hohen negativen Zusammenhang entsprochen: Je größer Personen sind, umso leichter sind sie (womit natürlich nicht wirklich zu rechnen wäre). Ein Korrelationskoeffizient von 0 hätte bedeutet, dass es keinen Zusammenhang zwischen Körpergröße und Gewicht gibt.

Zum Schluss

Mit diesem Beispiel beenden wir unsere kurze Darstellung der Auswertung und Präsentation der Ergebnisse einer Befragung. Sie können jetzt entscheiden, welches der beschriebenen Auswertungsverfahren für die Daten Ihrer schulischen Befragung angemessen ist. Wenn Sie darüber nachdenken, wie Sie Ihre Ergebnisse am besten präsentieren können, sollten Sie – wie eingangs schon angedeutet – sich grundsätzlich mit der Frage beschäftigen, wer Ihnen eigentlich zuhören oder Ihre Berichte lesen wird. Nur eine zielgruppennahe Darstellung ist eine gute Darstellung. Verzichten Sie auf allzu komplexe Verfahren der Auswertung und Präsentation selbst dann, wenn Sie oder Ihre SchülerInnen diese leisten könnten. Besser eine übersichtliche, optisch ansprechende und verständliche Präsentation als komplexe Modelle und komplexe Präsentationsmuster, die Ihre Leser- und Zuhörerschaft vielleicht gar nicht wirklich nachvollziehen können. Verzichten Sie selbst auf den Anspruch, mit möglichst komplexen, multivariaten Verfahren zu arbeiten; Ihre Daten werden das in den meisten Fällen nicht zulassen – und Ihre LeserInnen und ZuhörerInnen werden Ihnen für einfache, nachvollziehbare und verständliche Darstellungen dankbar sein. Hinzu kommt, dass Ihre SchülerInnen den Wert von sinnvollen grafischen Darstellungen erkennen werden und anzuwenden lernen.

Bei der Auswertung von Umfragedaten und ihrer Präsentation gilt, – zumindest für Arbeiten außerhalb des eigentlichen professionellen Kontextes, also z.B. bei Ihren schulischen Befragungen – dass weniger manchmal mehr ist.

Mit diesem Hinweis beenden wir nicht nur unsere kurze Darstellung der Auswertung und Präsentation der Ergebnisse einer Befragung, sondern in gewisser Weise auch die Serie „Schüler forschen“. Im letzten, noch ausstehenden Beitrag werden wir uns mit der Frage beschäftigen, wie wir selbst mit Umfragedaten umgehen, denen wir alltäglich begegnen. Was macht uns unsicher, wenn wir Umfrageergebnisse zur Kenntnis nehmen? Wie können wir Umfrageergebnisse „bewerten“? Und beim Beantworten dieser und ähnlicher Fragen werden wir auch nicht umhinkommen, Befragungen und ihre Ergebnisse selbst kritisch zu hinterfragen.

Anmerkungen

1 Ruth Holthof ist als Studienrätin im Fach Sozialkunde am Eleonoren-Gymnasium in Worms tätig.

2 Vereinfachend deshalb, weil bei der Verwendung dieses Koeffizienten eigentlich zwei Bedingungen erfüllt sein müssen: Die beiden Variablen, deren Zusammenhang überprüft werden soll, müssen normalverteilt sein und linear zusammenhängen. Wir gehen darauf nicht näher ein, sondern machen das, was man „üblicherweise“ macht: Wir arbeiten mit dem Koeffizienten, ohne die Bedingungen zu überprüfen.

3 Die geringe Fallzahl verbietet es realiter, den Korrelationskoeffizienten dieser Daten zu interpretieren. Es geht hier nur um die beispielhafte Darstellung der Berechnung.

4 Sofern der Zusammenhang als linear definiert worden ist (vgl. Endnote 2); bei nicht-linearem Zusammenhang kann auch ein Wert von 0 für ein Zusammenspiel der beiden Variablen stehen.

Literatur

Borgatta, E.F. & G.W. Bohrnstedt (1980): Level of Measurement. Once Over Again. In: Sociological Methods and Research 9, 2: 147-160

Kampen, J. & M. Swyngedouw (2000): The Ordinal Controversy Revisited. In: Quality & Quantity 34: 87-102

Kühnel, S. & D. Krebs (2001): Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek: rowolths enzyklopädie

Porst, R. (2014): Von der Fragestellung zum Fragebogen – Zentrale Begrifflichkeiten. in: Gesellschaft. Wirtschaft. Politik 63, Heft 2, S. 221-232

Porst, R. (2015): Fragebogenkonstruktion (I) – Grundlagen, Arten von Fragen, Arten von Skalen. in: Gesellschaft. Wirtschaft. Politik 64, Heft 2. S. 239-250

Porst, R. (2016): Datenerfassung und Datenaufbereitung. in: Gesellschaft. Wirtschaft. Politik 64, Heft 4, S. 491-500

Rentenpolitik in Europa

Johannes Varwick
Eberhard Eichenhofer
Jana Windwehr
Manuel Wäschle

Europäisierung der Alterssicherungspolitik

Up-, down- und interload-Prozesse zwischen Mitgliedstaaten und europäischer Ebene

2016. 345 Seiten. Kart.
42,00 € (D), 43,20 € (A)
ISBN 978-3-8474-0510-8
eISBN 978-3-8474-0978-6

Die Rentenpolitik der EU-Mitgliedstaaten ist in den vergangenen Jahren zunehmend in den Blick europäischer Politik geraten. Gerade unter dem Einfluss der Finanz- und Staatsschuldenkrise findet eine fortschreitende Europäisierung der Alterssicherungspolitik statt, die weder in der Forschung noch in der Öffentlichkeit umfassend reflektiert wird.

Die AutorInnen entwerfen ein Modell dieser dynamischen Interaktion zwischen EU-Institutionen und Mitgliedstaaten.

www.shop.budrich-academic.de