

7 Schluss von der Stichprobe auf die Population

Normalerweise ist die bloße **Beschreibung einer Stichprobe** nicht das, was wirklich interessiert. Nehmen Sie manchmal, wenn Sie Spaghetti kochen, eine einzelne Nudel aus dem Topf und probieren, ob sie schon weich genug ist? Und sind Sie daran interessiert zu erfahren, ob diese einzelne Nudel essfertig ist, oder benutzen Sie sie nicht viel mehr als Stichprobe, um auf den Rest der Teigwaren im Topf – die Grundgesamtheit Ihrer Nudeln – zu schließen?

Dabei könnte es geschehen, dass Sie zufälligerweise eine Nudel herausfischen, die als einzige ausreichend gar ist, aus welchen Gründen auch immer. Nehmen wir an, Sie haben die im Topf befindlichen Nudeln immer ausreichend umgerührt, das Wasser im Topf kocht gleichmäßig, die einzelnen Nudeln sind in puncto Zusammensetzung der Teigmasse und hinsichtlich ihrer Stärke vergleichbar usw., dann ist es allerdings äußerst unwahrscheinlich, dass Ihre „Stichprobennudel“ gar ist, die anderen aber noch nicht. Diese eine Nudel wird wohl repräsentativ für die Grundgesamtheit sein.

7.1 Alltags- und statistische Hypothesen

Im Alltagsleben stellen wir mehr oder minder ständig **Hypothesen** über das, was wir erleben und wie wir es erklären, auf, ohne uns dessen bewusst zu sein. „Wenn meine Stichprobennudel gar ist, werden wohl alle anderen auch gar sein“, oder „Wer Armani-Anzüge trägt, verdient viel Geld“, oder „Wer Universitätsprofessor ist, verfügt über einen hohen Intelligenzquotienten“. Beispiele für derartige Alltagshypothesen sind leicht zu finden, und wir suchen und finden Erklärungsmodelle für unsere Entscheidungen und Bewertungen zu meist implizit. Wir legen nicht aufgrund transparenter, nachvollziehbarer Kriterien fest, was und aus welchem Grund wir für dermaßen richtig halten, dass wir unsere Entscheidungen darauf aufbauen.

Auch in der Statistik werden Entscheidungen aufgrund von Hypothesen getroffen. Allerdings unterscheiden sich diese Hypothesen in einigen wesentlichen Punkten gravierend von den eben besprochenen **Alltagshypothesen**. **Statistische Hypothesen** werden stets als „Hypothesenpaar“ formuliert: Die sogenannte „**Nullhypothese**“ steht der „**Alternativhypothese**“ (auch Forschungshypothese genannt) gegenüber, und es ist die Aufgabe der Signifikanztests (Kapitel 8), diese Hypothesen zu überprüfen.

Die Nullhypothese „behauptet“ meistens, dass es zwischen Gruppen oder Variablen keine Zusammenhänge oder Unterschiede gibt, die Alternativhypothese „behauptet“ meistens, dass es Zusammenhänge oder Unterschiede gibt. Deshalb steckt das, was der Untersucher/die Untersucherin glaubt, normalerweise in der Alternativhypothese. Die meisten Alternativhypothesen machen Aussagen über Zusammenhänge, Unterschiede oder Veränderungen.

In unserem Fragebogen werden die Personen u. a. gefragt, wie viele Stunden pro Woche sie sich körperlich betätigen (C1.4) und für wie sportlich sie sich halten (C1.5). Nun könnte man die Forschungshypothese aufstellen, dass körperliche Bewegung und subjektive Einschätzung der eigenen Sportlichkeit zusammenhängen. Das wäre in diesem Fall also die Alternativhypothese. Die dazugehörige Nullhypothese behauptet inhaltlich das Gegenteil: Körperliche Bewegung und subjektive Einschätzung der eigenen Sportlichkeit hängen nicht zusammen.

Wichtig ist: Diese Hypothesen beziehen sich nicht auf die Stichprobe, sondern auf die dahinterstehende **Grundgesamtheit (Population)**, die in der Regel unbekannt ist – denn wären die Verhältnisse in der Grundgesamtheit bekannt, bräuchte man über sie auch keine hypothetischen Überlegungen („Hypothesen“ eben ...) anzustellen! Denken Sie an das einführende Nudelbeispiel: Kontrolliert ein Koch jede einzelne Nudel aus dem Topf (= Population), braucht er keine Hypothesen über diese Population aufzustellen – er kennt sie ja in diesem Fall. Zieht er aber nur eine Stichprobe von sagen wir zwei Nudeln (das ist dann seine Stichprobe) und sind diese zwei Nudeln gar, so weiß er nicht mit Sicherheit, ob alle Nudeln ebenso wie diese zwei gezogenen Nudeln gar sind – hier würde er dann hypothesengeleitet vorgehen, indem er die Alternativhypothese „Die Nudeln der Population sind gar“ versus die Nullhypothese „Die Nudeln der Population sind nicht gar“ aufstellt.

Bisher haben wir uns allerdings noch keine Gedanken darüber gemacht, wie wir die „Garheit“ messen können. Dazu könnten wir eine subjektive Skala von 0 („hart und ungenießbar“) bis 5 („weich und ungenießbar“) aufstellen, wobei ein Wert von 3 optimal wäre. Wir könnten also zehn Nudeln ziehen (Stichprobe), den Mittelwert der „Garheit“ ausrechnen und hätten so die Garheit operationalisiert.

Operationalisierung

Möchte man etwas messbar machen, muss zunächst definiert werden, was gemessen werden soll. Bei der Körpergröße etwa ist das leicht machbar. Aber wie definiert man das Konstrukt „Romantische Liebe“? Man könnte sie beispielsweise als das Bedürfnis definieren, seinem Partner/seiner Partnerin rote Rosen zu schenken (ob das eine sinnvolle Definition ist, sei dahingestellt ...). Aber das allein reicht nicht aus. Es muss auch angegeben werden, durch welche beobachtbaren Ereignisse dieses Konstrukt erfasst, also gemessen werden kann. Etwa: Häufigkeit, mit der man der betreffenden Person rote Rosen schenkt. Hat man sich also auf eine Definition und messbare Ereignisse geeinigt, ist der Begriff „Romantische Liebe“ messbar gemacht, er ist operationalisiert. Der Begriff „Romantische Liebe“ wird operationalisiert durch die Häufigkeit, mit der man rote Rosen schenkt ...

Mittels der **Inferenzstatistik** werden also konkurrierende Hypothesen, die Null- und die Alternativhypothese, geprüft. Die Alternativhypothese kann auf zwei Arten formuliert werden: gerichtet oder ungerichtet. Ungerichtet ist eine Alternativhypothese, wenn keine Richtung des Zusammenhangs oder Unterschieds vorgegeben wird.

SPSS/PASW geht grundsätzlich von der Ungerichtetheit der Hypothesen aus.

Alternativhypothese (H_1): „Männer und Frauen unterscheiden sich hinsichtlich ihres durchschnittlichen wöchentlichen Bewegungspensums.“

Nullhypothese (H_0): „Männer und Frauen unterscheiden sich *nicht* hinsichtlich ihres durchschnittlichen wöchentlichen Bewegungspensums.“

Gerichtet ist eine Alternativhypothese, wenn etwas über die Richtung des erwarteten Zusammenhangs oder Unterschieds ausgesagt wird, wie folgendes Beispiel zeigt.

Alternativhypothese (H_1): „Männer machen durchschnittlich mehr Bewegung als Frauen.“

Nullhypothese (H_0): „Männer machen durchschnittlich weniger oder höchstens gleich viel Bewegung wie Frauen.“

7.2 Statistischer Test

Wenn Hypothesen formuliert sind, muss ein Weg gefunden werden, um zu entscheiden, welche der beiden konkurrierenden Hypothesen für „wahr“ gehalten wird. „Wahr“ heißt: Was wird wohl in der Population gelten? In der Population, die wir nicht kennen! Da wir die Population also nicht kennen, können Aussagen über sie **nicht mit Sicherheit** getroffen werden. Man könnte sagen: Ein Befund, der an einer Stichprobe erhoben wurde, soll auf seine **Allgemeingültigkeit** – über die konkrete Stichprobe hinaus – überprüft werden.

Ein statistischer Test ist das Mittel, um diese **Prüfung auf Allgemeingültigkeit** vorzunehmen. Es gibt mehrere Arten statistischer Tests. Nachdem statistische Hypothesen formuliert und ein Untersuchungsdesign festgelegt wurden, bietet sich folgende Grundstruktur an:

1. Erhebung empirischer Daten, eben z. B. durch Ausfüllen von Fragebogen
2. Berechnung von Statistiken aus diesen Daten (Stichprobenkennwerte), wie beispielsweise Mittelwerte
3. „Verpackung“ dieser Stichprobenkennwerte nach bestimmten Regeln in bestimmte Formeln. Was dabei herauskommt, ist eine sogenannte „Teststatistik“.
4. Berechnung, wie wahrscheinlich diese oder eine extremere Teststatistik ist, unter der Annahme, dass in der Population die Nullhypothese gilt
5. Wenn diese Wahrscheinlichkeit gering ist, „glaubt“ man an die Alternativhypothese (signifikant); wenn diese Wahrscheinlichkeit groß ist, „glaubt“ man weiterhin an die Nullhypothese (insignifikant).

In unserem Fragebogen wird neben dem Geschlecht (C1.1) auch erhoben, wie viele Stunden körperliche Bewegung die Befragten machen (C1.4). Wir vermuten im Vorfeld, dass sich Männer und Frauen durchschnittlich hinsichtlich des Bewegungspensums unterscheiden, und formulieren deshalb ein Hypothesenpaar.

Alternativhypothese: „Männer und Frauen unterscheiden sich hinsichtlich ihres durchschnittlichen Bewegungspensums.“

Nullhypothese: „Männer und Frauen unterscheiden sich *nicht* hinsichtlich ihres durchschnittlichen Bewegungspensums.“

Als Untersuchungsdesign ergibt sich die Fragebogenerhebung.

1. Die Daten werden erhoben, indem Personen gebeten werden, den Bogen auszufüllen.
2. Für die Gruppe der Männer und Frauen werden jeweils Mittelwerte berechnet, was ihr wöchentliches Bewegungspensum betrifft.
3. Aus diesen Mittelwerten wird eine *Teststatistik* berechnet. *Im Wesentlichen* werden die Mittelwerte der beiden Gruppen voneinander subtrahiert, und wenn *in der Population die Nullhypothese gelten sollte* (kein Unterschied zwischen Männern und Frauen), erwarten wir eigentlich, dass bei dieser Subtraktion der Stichprobenmittelwerte der Wert „Null“ herauskommt – Werte weit entfernt von Null wären in diesem Fall doch unwahrscheinlich ... Wichtig ist, dass die so berechnete Teststatistik eine gewisse Auftretenswahrscheinlichkeit hat, die berechenbar ist. Genau das macht das Statistikprogramm. Diese Auftretenswahrscheinlichkeit heißt „Signifikanz“ oder „p-Wert“ (von lat. *probabilitas*: Wahrscheinlichkeit).
4. Dieser p-Wert gibt also die Wahrscheinlichkeit an, diese aus den empirischen Daten berechnete Teststatistik zu erhalten, wenn – und das ist wesentlich! – in der Population, die wir ja nicht kennen, die Nullhypothese gilt. Mit anderen Worten: „Der P-Wert ist die Wahrscheinlichkeit, mit der man sich irrt, wenn man die Nullhypothese ablehnt“ (Sachs, 1999, S. 188).
5. Wenn diese Wahrscheinlichkeit gering ist, entscheidet man sich für die Alternativhypothese. Was heißt aber „gering“? Dafür wurden willkürliche Grenzen festgelegt, man spricht vom Signifikanzniveau „Alpha“ (α). Üblich sind Signifikanzniveaus von 5 %, 1 % und 0,1 % (0,05; 0,01; 0,001).

Wurde also vor der Untersuchung ein **Signifikanzniveau** von 5 % (oder: 0,05) festgelegt und beträgt der ermittelte p-Wert 0,02, so ist das Ergebnis auf dem 5 %-Niveau signifikant, was bedeutet, dass die Alternativhypothese angenommen wird. Mit anderen Worten: Höchstwahrscheinlich wird in der Population die Alternativhypothese gelten! Aber sicher wissen können wir es nicht, weshalb eine statistische Hypothese auch nicht streng bewiesen werden kann. Sicher wissen könnten wir es nur, wenn wir die Population kennen würden, aber dann bräuchten wir auch keine Hypothesen und somit auch keinen statistischen Test zur Überprüfung dieser Hypothesen!

Ermitteln wir einen p-Wert von $p = 0,02$, so ist das Ergebnis auf dem 5 %-Niveau signifikant, auf dem 1 %-Niveau nicht mehr.

Ermitteln wir einen p-Wert von $p = 0,06$, so ist das Ergebnis weder auf dem 5 %- noch auf dem 1 %- noch auf dem 0,1 %-Niveau signifikant (signifikant wäre es auf dem 10 %-Niveau – ein derart hohes Signifikanzniveau wird allerdings selten gewählt, da hier die Wahrscheinlichkeit, einen Fehler erster Art zu begehen, zu groß wäre).