1 Einleitung

Angenommen es soll eine Erhebung über das aktuelle Einkommen der Landwirte, die einen Hof bestimmter Größe bewirtschaften, durchgeführt werden. Eine geeignete Stichprobe wird gezogen, und als Resultat ist unter anderem das durchschnittliche aktuelle Einkommen der zufällig ausgewählten und befragten Landwirte zu erhalten. Gewünscht ist aber eine Aussage über das durchschnittliche Einkommen aller Landwirte, also auch über das Einkommen jener Landwirte, die gar nicht befragt worden sind. Hierin liegt eine wesentliche Aufgabe der bewertenden Statistik (induktive oder schließende Statistik, beziehungsweise Inferenzstatistik).

Beim Testen wird zunächst eine statistische Hypothese über einen interessierenden Sachverhalt formuliert. Eine geeignete Zufallsstichprobe[1] wird gezogen; sie liefert Daten, die den interessierenden Sachverhalt betreffen und aufhellen sollen. Dies erfolgt unter Verwendung eines adäquaten wahrscheinlichkeitstheoretischen Modells. Im Rahmen dieses Modells wird die eingangs formulierte Hypothese über den Sachverhalt durch die Anwendung einer Entscheidungsregel verworfen oder gestützt. Der statistische Befund der Zufallsstichprobe wird somit entweder als nicht in Widerspruch zur Hypothese stehend bewertet oder als Widerlegung der Hypothese angesehen.

Der statistische Befund ist nunmehr insofern verallgemeinert, als er sich nicht mehr ausschließlich auf die Merkmalsträger[2] der Zufallsstichprobe bezieht. Mit der Annahme oder Zurückweisung der zu prüfenden Hypothese kann ein Fehler begangen werden. Von Vorteil ist jedoch, dass die Wahrscheinlichkeit für den Fehler beziffert werden kann.

Beispielsweise wird behauptet, dass weniger als die Hälfte der BürgerInnen die Steuerpolitik der gegenwärtigen Bundesregierung gutheißen. Dies ist die Hypothese, die zu prüfen ist. In einer für die BürgerInnen repräsentativen, durch den Zufall gesteuerten Befragung könnte sich nun als statistischer Befund ergeben, dass 53% der Befragten die Steuerpolitik der Regierung unterstützen. Zu entscheiden ist, ob sich der Befund in der Stichprobe signifikant – das heißt wesentlich – von der zu prüfenden Hypothese unterscheidet, oder ob der Unterschied als zufällig zustande gekommen beurteilt werden kann. Als Entscheidungshilfe wird ein der Problemstellung und der Datenlage adäquates statistisches Modell gewählt. Möglicherweise ergibt sich als formale Lösung im Modell, dass die Hypothese nicht zu verwerfen ist. Dies bedeutet, dass die in der Befragung festgestell-

[1] Eine Zufallsstichprobe ist ein spezielles Auswahlverfahren, bei dem jeder Merkmalsträger die gleiche Chance erhält, in die Stichprobe zu gelangen.

[2] Eine Beobachtungseinheit (Merkmalsträger) ist die kleinste Einheit, an der interessierende Eigenschaften direkt beobachtete werden.

te Unterstützung von 53% der BürgerInnen nicht als unvereinbar mit der Behauptung angesehen werden muss, dass tatsächlich weniger als die Hälfte aller BürgerInnen die Steuerpolitik der gegenwärtigen Bundesregierung gutheißen. In diesem Fall kann der Fehler begangen werden, die Hypothese nicht zu verwerfen, obwohl sie eine unzutreffende Kennzeichnung der Stimmungslage aller BürgerInnen zum Ausdruck bringt. Die Wahrscheinlichkeit für diesen Fehler kann häufig im Rahmen des verwendeten statistischen Modells bestimmt werden.

1.1 Statistisches Schätzen versus statistischem Testen

Im Rahmen des Testens wird eine Aussage über einen interessierenden Sachverhalt dadurch getroffen, dass unter Berücksichtigung von Daten der Stichprobe entschieden wird, ob eine Hypothese über diesen Sachverhalt zu stützen oder zu verwerfen ist. Die Aussage erfolgt somit auf einem indirekten Weg.

Ein direkter Weg wird beim Schätzen eingeschlagen. Hier besteht die Ausgangslage im statistischen Befund der Zufallsstichprobe, die den interessierenden Sachverhalt betrifft. Es wird versucht, den wahren und nicht bekannten Zustand des Sachverhalts möglichst gut abzuschätzen. Hierzu muss ein adäquates Verfahren für die praktische Durchführung der Schätzung gewählt werden. Des Weiteren muss natürlich über die Qualität der Schätzung Klarheit bestehen.

Der statistische Befund der Stichprobe wird beim Schätzen nunmehr insofern verallgemeinert, als er der bestmöglichen – numerischen – Aufhellung des interessierenden wahren Sachverhaltes dient. Auch hier sind keine sicheren Aussagen möglich. Von Vorteil ist jedoch, dass es einen Weg gibt, das Vertrauen in die Güte der Abschätzung zum Ausdruck zu bringen.

Hierzu soll erneut das Beispiel einer für die BürgerInnen repräsentativen, durch den Zufall gesteuerten Befragung bezüglich der Steuerpolitik der gegenwärtigen Regierung betrachtet werden, nunmehr jedoch mit der Absicht einer Schätzung. Die Stichprobe ergibt, dass 53% der Befragten die Steuerpolitik unterstützen. Ein nahe liegendes und gutes Schätzverfahren besteht darin, diesen Anteilswert der Stichprobe auch für jene Personen als gültig anzusehen, die gar nicht befragt worden sind (Punktschätzung). Selbstverständlich besteht die Möglichkeit, dass die nicht befragten Personen, beispielsweise nur zu weniger als 51% der Steuerpolitik der Regierung gewogen sind. Dies kann etwa durch die folgende ergänzende, die Sicherheit der Schätzung quantifizierende Aussage mitgeteilt werden (Intervallschätzung): „Zwischen 50% und 56% der BürgerInnen unterstützen die Steuerpolitik der Regierung. Die Sicherheit dieser Aussage beträgt 95%". Selbstverständlich ist die letzte Aussage keine Erfindung, sondern das Resultat einer wahrscheinlichkeitstheoretischen Berechnung.

Liegen Grundgesamtheiten[3] oder Nichtzufallsstichproben vor, so werden nur Methoden der beschreibenden Statistik[4] angewendet. Dies gilt auch für sehr große Zufallsstichproben und häufig auch dann, wenn neue Resultate mit alten zu vergleichen sind, sowie auch dann, wenn die Ereignisse eine weitergehende Analyse anhand von Methoden der schließenden Statistik weder wünschenswert noch notwendig erscheinen lassen. Die bewertende Statistik wird in folgende Bereiche gegliedert (Abbildung 1):

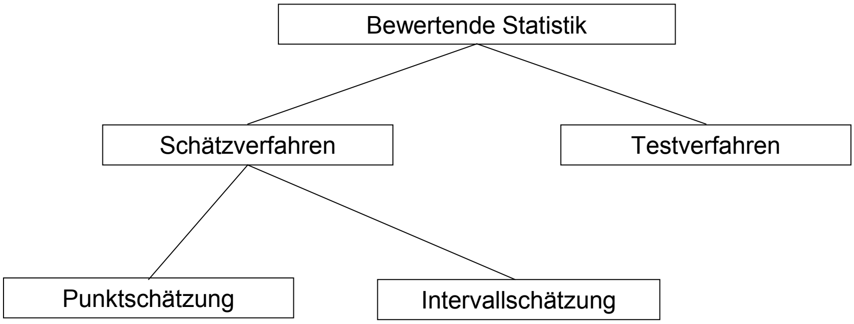

Abbildung 1: Einteilung der bewertenden Statistik

Die in der Statistik wichtige Standardisierung wird im Folgenden einleitend betrachtet, da dieses Konzept sehr oft bei Test- und Schätzverfahren zur Anwendung kommt.

Wenn von einem/einer SchülerIn zum Beispiel ein Testwert von 3 bekannt ist, kann mit dieser Information wenig angefangen werden. Es ist nicht bekannt, ob die Testperson auf das Ergebnis stolz sein kann oder nicht. Aussagekräftig wird diese Zahl erst dann, wenn sie in Relation zu anderen gesetzt wird: wenn bekannt ist, welchen Wert die übrigen SchülerInnen bekommen haben oder wenn diese 3 mit dem Mittelwert $\bar{x}$ (dem arithmetischen Mittel) verglichen wird. Dann kann gesagt werden, ob der/die SchülerIn eine durchschnittliche oder aber eine überdurchschnittliche Leistung erbracht hat. In diesem Fall kann der/die SchülerIn mit einer

[3] Die Grundgesamtheit ist eine Menge von Beobachtungseinheiten, für die vom Untersuchungsziel her eine Frage geklärt werden soll. Die Grundgesamtheit muss zeitlich, räumlich und sachlich eindeutig abgegrenzt werden.

[4] Mit der deskriptiven (beschreibenden) Statistik wird versucht, die in einem Datensatz enthaltene Information durch bestimmte Kennzahlen und graphische Darstellungen zu veranschaulichen.

durchschnittlichen Leistung aufwarten (siehe Tabelle 1) – der Wert entspricht genau dem Durchschnitt:

SchülerIn	Testergebnis
1	1
2	3
3	5
4	4
5	2
	$\bar{x} = 15 : 5 = 3$
	$s = \sqrt{\frac{10}{4}} = 1{,}58$

Tabelle 1: Testergebnis von fünf SchülerInnen

Bezeichnen $x_1, x_2, \ldots, x_n$ die n Beobachtungswerte bezüglich eines metrisch skalierten Merkmals, so ist das arithmetische Mittel $\bar{x}$ gegeben durch:

$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ und die Varianz s^2 durch

$$s^2 = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots (x_n - \bar{x})^2\right].$$

Die Standardabweichung s ist durch die Quadratwurzel der Stichprobenvarianz gegeben.

Eine Möglichkeit, die relative Lage einer Maßzahl in einer Verteilung anzugeben, erlaubt die Transformation der Zahlen in z-Werte. Der z-Wert gibt die Abweichung der Maßzahlen vom arithmetischen Mittel in Einheiten der Standardabweichung an. Der z-Wert von SchülerIn 3 zum Beispiel beträgt 1,27.

$$z_3 = \frac{x_3 - \bar{x}}{s} = \frac{5-3}{1{,}58} = \frac{2}{1{,}58} = 1{,}27$$

Wie vorteilhaft dieser z-Wert ist, wird in Tabelle 2 besonders deutlich: hier sind die Ergebnisse eines weiteren Tests angeführt. Aufgrund der einzelnen Testwerte kann schwer bestimmt werden, ob SchülerIn 3 besser bei Test 1 oder bei Test 2 abgeschlossen hat. Die Punktzahlen (Test 2: 22 Punkte, Test 1: 5 Punkte) legen zunächst die Vermutung nahe, dass Test 2 positiver für ihn ausgefallen ist. Aber trotzdem geht aus der Tabelle 2 hervor, dass er besser bei Test 1 abgeschnitten haben muss; hier hat er den höchsten aller beobachteten Punktwerte erreicht. Beim Vergleich der einzelnen Werte sollten also die Testergebnisse der ganzen Gruppe berücksichtigt werden.

Schüler	Test 1	Test 2	z_1	z_2
1	1	30	-1,27	1,5
2	3	15	0,00	-0,75
3	5	22	1,27	0,30
4	4	20	0,63	0,00
5	2	13	-0,63	-1,05
Summe	15	100	0	0
	$\bar{x} = 3$	$\bar{y} = 20$		
	$s_x = 1{,}58$	$s_y = 6{,}67$		

Tabelle 2: Testergebnisse für Test 1 und Test 2

Die Distanz zum arithmetischen Mittel allein kann nicht ausschlaggebend sein. In beiden Fällen liegt der/die SchülerIn um zwei Punkte über dem arithmetischen Mittel, aber relativ zur gesamten Schülergruppe liegt er besser bei Test 1. Werden die Testergebnisse standardisiert, geht aus den Werten eindeutig hervor, dass SchülerIn 3 bei Test 1 erfolgreicher war als bei Test 2.

Auf zwei Charakteristika der z-Werte soll hingewiesen werden:

1 Die Summe aller z-Werte einer Verteilung ist immer Null, daher sind alle arithmetischen Mittel Null (vgl. dazu die Tabelle 2).

2 Die Varianz beziehungsweise Standardabweichung der z-Werte einer Verteilung ist immer $s_z^2 = s_z = 1$