Günther Malle, Universität Wien

TESTEN UND SCHÄTZEN MIT HILFE DER BINOMIALVERTEILUNG

1. Vorbemerkungen

Der Unterricht in Stochastik (Wahrscheinlichkeitsrechnung und Statistik) in der Schule zerfällt in drei Teile: Beschreibende Statistik, Wahrscheinlichkeitsrechnung und Beurteilende Statistik (deren Kern das Testen von Hypothesen und Schätzen von Parametern ist). Nach den derzeit gültigen AHS-Lehrplänen wird die Beschreibende Statistik in der Unterstufe bzw. 5. Klasse behandelt, die Wahrscheinlichkeitsrechnung und Beurteilende Statistik sind für die 7. und 8. Klasse vorgesehen. Erfahrungsgemäß wird die Beurteilende Statistik oft weggelassen, doch war diese (als heute wichtigstes Angwendungsgebiet der Wahrscheinlichkeitsrechnung) der Hauptgrund, warum stochastische Inhalte überhaupt in die AHS-Lehrpläne aufgenommen wurden. Wird die Beurteilende Statistik weggelassen, bleibt der ganze Stochastiklehrgang ein Torso. Die Wahrscheinlichkeitsrechnung läuft dabei Gefahr, in eine selbstgefällige "Würfelbudenmathematik" auszuarten, über deren Sinn für die Schule man streiten kann.

Der Weg zur Beurteilenden Statistik sieht in vielen Schul- und Hochschulbüchern so aus: Aufbauend auf Grundlagen der Wahrscheinlichkeitsrechnung wird zunächst die Binomialverteilung, auf diese aufbauend dann die Normalverteilung behandelt. Mit dieser werden dann statistische Test- und Schätzverfahren durchgeführt. Bei diesem Aufbau im Schulunterricht erlaubt es die beschränkte zur Verfügung stehende Zeit zwar, die Wahrscheinlichkeitsrechnung und Binomialverteilung in der 7. Klasse zu behandeln, die Normalverteilung muß jedoch im allgemeinen auf die 8. Klasse verschoben werden. Damit geraten das statistische Testen und Schätzen an das gefährliche Ende der Schulzeit, wo sie dann aus Zeitmangel oft ganz hinausfallen. Es ist also wünschenswert, Möglichkeiten zu finden, die schneller zum statistischen Testen und Schätzen führen. Im folgenden wird eine Möglichkeit vorgestellt, diese Inhalte bereits in der 7. Klasse zu behandeln, lediglich mit der Binomialverteilung allein (also ohne Normalverteilung). In der 8. Klasse können Test- und Schätzverfahren dann natürlich erneut aufgegriffen werden, mit dem eher geringfügigen Unterschied, daß mit der Normalverteilung anstelle der Binomialverteilung gearbeitet wird. Ein solches Vorgehen ist im Sinne des Spiralprinzips durchaus wünschenswert.

Im Mathematikunterricht der Schule kann es nicht darum gehen, möglichst viele Test- und Schätzverfahren kennenzulernen. Es kann nur darum gehen, in exemplarischer Weise einige *grundlegende Ideen* des Testens und Schätzens aufzuzeigen. Leider wird hier vielfach das Vorgehen in der Hochschulliteratur recht unkritisch auf die Schule übertragen - mit zu großem Begriffsaufwand und zu hochgestochener Fachterminologie. Die folgenden Ausführungen sollen jedoch zeigen, daß es auch auf einfachere Weise geht. Dabei beschränke ich mich auf das *Testen bzw. Schätzen von Anteilen.* Dieses hat nämlich gegenüber dem Testen bzw. Schätzen anderer Parameter (z.B. Mittelwerte) den Vorteil, daß weniger mathematische Voraussetzungen erforderlich sind. In der Tat braucht man aus der Wahrscheinlichkeitsrechnung nur die allereinfachsten Grundbegriffe (eine breite Behandlung der Wahrscheinlichkeitsrechnung ist also nicht erforderlich). In Hinblick auf Binomialverteilungen benötigt man den folgenden Satz (vgl. etwa BÜRGER/FISCHER/MALLE 1991, Seite 262):

Satz: Bei einem Zufallsversuch trete ein Ereignis E mit der Wahrscheinlichkeit p ein. Der Versuch werde n-mal unter den gleichen Bedingungen durchgeführt. Ist H die Anzahl der Versuche, in denen das Ereignis E eintritt, dann gilt:

$$P(H=k)= \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} \quad (\text{für } 0\leq k\leq n)$$

Durch die in diesem Satz enthaltene Formel wird jedem möglichen Wert k der Zufallsvariablen H eine Wahrscheinlichkeit P(H=k) zugeordnet. Auf diese Weise erhält man eine **Wahrscheinlichkeitsverteilung** von H. Eine Wahrscheinlichkeitsverteilung, bei der die Wahrscheinlichkeiten P(H=k) durch die im obigen Satz enthaltene Formel festgelegt sind, heißt **Binomialverteilung mit den Parametern n und p**, die Zufallsvariable H nennt man binomialverteilt mit den Parametern n und p. Mit dieser Terminologie kann der obige Satz auch so ausgesprochen werden:

Satz: Bei einem Zufallsversuch trete ein Ereignis E mit der Wahrscheinlichkeit p ein. Der Versuch werde n-mal unter den gleichen Bedingungen durchgeführt. Dann ist die Anzahl H der Versuche, in denen E eintritt, binomialverteilt mit den Parametern n und p.

Ist H binomialverteilt mit den Parametern n und p, dann können Wahrscheinlichkeiten der Form $P(H=k)$, $P(H\leq k)$ bzw. $P(H\geq k)$ aus Tabellen abgelesen werden. (Für $n=10$ und $n=20$ findet man solche Tabellen etwa in BÜRGER/FISCHER/MALLE 1991, Seite 332-334.)

2. Einseitige Anteilstests

Wir gehen von folgender Aufgabe aus:

Ein Geschäftsmann verkauft in seinem Geschäft Lose und behauptet, daß nur 40% der Lose Nieten sind. Ein kritischer Kunde vermutet, daß dieser Anteil höher ist. Er will dies durch eine Stichprobe vom Umfang 10 (d.h. anhand von 10 Losen) überprüfen.

a) Wie viele Nieten sind in der Stichprobe am ehesten zu erwarten, wenn der Geschäftsmann recht hat?

b) Kann der Kunde die Behauptung des Geschäftsmannes verwerfen, wenn er in der Stichprobe mehr Nieten vorfindet?

c) Falls der Geschäftsmann recht hat, mit welcher Wahrscheinlichkeit wird der Kunde in der Stichprobe dann mindestens 6 Nieten vorfinden?

Die Antworten zu a) und b) sind einfach. Wenn der Geschäftsmann recht hat, sind 40% von 10, also 4 Nieten zu erwarten. Der Kunde kann die Behauptung des Geschäftsmannes aber nicht mit Sicherheit verwerfen, wenn er mehr als 4 Nieten vorfindet (nicht einmal dann, wenn er lauter Nieten vorfindet). Denn die Anzahl der Nieten ist von Stichprobe zu Stichprobe zufälligen Schwankungen unterworfen und kann in der vorliegenden Stichprobe zufällig größer als 4 sein.

Zur Beantwortung von c) bezeichnen wird die Nietenanzahl in Stichproben vom Umfang 10 mit H und fassen H als eine Zufallsvariable auf, die die Werte 0,1,2,...,10 annehmen kann. Um die Wahrscheinlichkeitsverteilung von H zu bestimmen, denken wir uns die 10 Lose der Stichprobe der Reihe nach gezogen. Das Ziehen eines Loses fassen wir als Zufallsversuch auf. Bei der Durchführung dieses Zufallsversuches tritt das Ereignis "Niete" mit einer gewissen Wahrscheinlichkeit p ein. Der Zufallsversuch wird 10-mal unter annähernd gleichen Bedingungen wiederholt (denn da die Anzahl der gezogenen Lose sehr klein gegenüber der Grundgesamtheit, d.h. der Menge aller Lose, ist, spielt es keine große Rolle, daß die Lose nicht zurückgelegt werden). Somit können wir H nach dem vorhin erwähnten Satz zumindest als annähernd binomialverteilt mit den Parametern n=10 und p ansehen. Der Wert von p ist an sich nicht bekannt, da wir nicht wissen, wie viele Nieten in der Grundgesamtheit vorhanden sind. In der Fragestellung c) ist jedoch vorausgesetzt, daß der Geschäftsmann recht hat, d.h. daß p=0,4 ist. Somit können wir H als annähernd binomialverteilt mit n=10 und p=0,4 ansehen.

Aus einer entsprechenden Tabelle können wir die gesuchte Wahrscheinlichkeit ablesen:

$$P(H\geq 6)\approx 0{,}166$$

Diese Wahrscheinlichkeit kann man frequentistisch interpretieren: Falls der Geschäftsmann recht hat, dann würde der Kunde bei sehr häufiger Stichprobenerhebung (vom Umfang 10) in ca. 16,6% aller Fälle mindestens 6 Nieten vorfinden.

Wie in c) kann man nun weiter fragen, mit welcher Wahrscheinlichkeit der Kunde in der Stichprobe mindestens 7,8,9,10 Nieten vorfinden würde, falls der Geschäftsmann recht hat. Man erhält durch analoge Überlegungen:

k	6	7	8	9	10
$P(H \geq k)$	0,166	0,055	0,012	0,002	0,000

Man sieht: Je größer k wird, desto kleiner wird $P(H \geq k)$, also die Wahrscheinlichkeit für mindestens k Nieten in der Stichprobe. Je unwahrscheinlicher k (oder mehr) Nieten in der Stichprobe sind, desto eher kann der Kunde klarerweise die Behauptung des Geschäftsmannes verwerfen, wenn er in der Stichprobe k Nieten vorfindet. Findet er in der Stichprobe 10 oder 9 Nieten vor, ist $P(H \geq k)$ nahezu null und er kann die Behauptung des Geschäftsmannes getrost verwerfen. Kann er dies auch tun, wenn er in der Stichprobe 8 Nieten vorfindet? Oder 7 Nieten? Wo liegt die Grenze? Wie groß darf $P(H \geq k)$ höchstens sein?

Diese Frage kann leider nicht objektiv beantwortet werden, denn ob eine Wahrscheinlichkeit als groß oder klein beurteilt wird, ist immer subjektiv. Die Frage sollte also nicht lauten "Wo liegt die Grenze?", sondern "Wo ziehen wir die Grenze?". In der statistischen Praxis ist es üblich, für $P(H \geq k)$ willkürlich eine obere Schranke α_0 anzugeben. Meist wählt man $\alpha_0=0{,}05$, bei genaueren Untersuchungen $\alpha_0=0{,}01$ (in seltenen Fällen eine noch kleinere Zahl). Wählen wir in unserem Beispiel etwa $\alpha_0=0{,}05$, dann sehen wir aus der obigen Tabelle: Der Kunde kann die Behauptung des Geschäftsmannes verwerfen, wenn er in der Stichprobe mindestens 8 Nieten vorfindet, aber nicht mehr, wenn er nur 7 Nieten vorfindet.

Wir beschreiben nun das Vorgehen zur Beantwortung der Fragestellung c) *allgemein* und führen dabei vorsichtig einige der üblichen Termini ein:

Ausgangspunkt ist eine Behauptung über die Größe eines bestimmten Anteils p in einer Grundgesamtheit (Behauptung des Geschäftsmannes). Man bezeichnet diese Behauptung als **Nullhypothese H_0**. Demgegenüber liegt eine Vermutung vor, daß dieser Anteil größer ist (Vermutung des Kunden). Diese Vermutung bezeichnet man als **Alternativhypothese H_1**. Wir schreiben kurz:

$$H_0\!: p=p_0 \qquad\qquad H_1\!: p>p_0$$

Man will die Nullhypothese durch eine Stichprobe vom Umfang n prüfen. Vor der Erhebung der Stichprobe legt man eine Zahl α_0 fest, die man als **Signifikanzzahl** bezeichnet. Dann wird die Stichprobe erhoben. Wir nehmen an, daß sich für die entsprechende Häufigkeit H in der Stichprobe H=k ergibt ($0 \leq k \leq n$).

Nun folgt die entscheidende Überlegung: Man nimmt an, daß die Nullhypothese H_0 gilt und berechnet unter dieser Annahme die Wahrscheinlichkeit $P(H \geq k)$. (Man beachte, daß H annähernd binomialverteilt mit den Parametern n und p_0 ist - letzteres wegen der Annahme von H_0.) Ist $P(H \geq k) \leq \alpha_0$, kann die Nullhypothese verworfen werden.

Man bezeichnet ein solches Vorgehen als **einseitigen Anteilstest mit der Signifikanz α_0** (bzw. auf dem α_0-Signifikanzniveau). Die Wahrscheinlichkeit $P(H \geq k)$ kann so gedeutet werden: Gilt die Nullhypothese H_0, wird sie jedoch bei einem Stichprobenergebnis $H \geq k$ verworfen, so begeht man einen Irrtum; die Wahrscheinlichkeit für einen solchen Irrtum ist gerade $P(H \geq k)$. Man bezeichnet diese Wahrscheinlichkeit daher als **Irrtumswahrscheinlichkeit**. Aber Achtung: $P(H \geq k)$ bedeutet nicht die Wahrscheinlichkeit für einen Irrtum schlechthin, sondern nur die Wahrscheinlichkeit für einen Irrtum, falls die Nullhypothese H_0 gilt. (Ein Irrtum ist auch möglich, wenn H_0 nicht gilt: mann kann H_0 ja irrtümlich annehmen.) Die obere Schranke α_0 für $P(H \geq k)$ kann als **maximal zugelassene Irrtumswahrscheinlichkeit** gedeutet werden.

Einen analogen Test kann man durchführen, wenn die Alternativhypothese nicht in der Form $p > p_0$, sondern in der Form $p < p_0$ formuliert wird. Man berechnet in diesem Fall die Irrtumswahrscheinlichkeit $P(H \leq k)$ unter der Annahme von H_0. Ist $P(H \leq k) \leq \alpha_0$, kann die Nullhypothese verworfen werden.

Was bedeutet es, wenn die Nullhypothese H_0 verworfen wird? Heißt dies, daß die Nullhypothese nicht gilt? Mit Sicherheit kann man dies leider nicht sagen. Es könnte ja sein, daß die Nullhypothese durch den vorgenommenen Test irrtümlich verworfen wurde. Die Wahrscheinlichkeit für einen solchen Irrtum ist aber höchsten α_0. Deutet man diese Wahrscheinlichkeit frequentistisch, ergibt sich - etwa für $\alpha_0 = 0{,}05$ - folgende Interpretation: Würde man sehr oft Stichproben vom Umfang n erheben und jedesmal die Nullhypothese bei einem Stichprobenergebnis $H \geq k$ verwerfen, so würde man höchstens in ca. 5% aller Fälle die Nullhypothese irrtümlich verwerfen. Ob dies bei dem vorliegenden Test der Fall ist, kann man allerdings nicht sagen.

Was bedeutet es, wenn die Nullhypothese H_0 nicht verworfen werden kann? In diesem Fall kann leider über die Gültigkeit von H_0 nichts ausgesagt werden. Man kann nur eine neue, eventuell umfangreichere Stichprobe erheben und einen neuerlichen Test durchführen - in der Hoffnung, H_0 verwerfen zu können.

3. Zweiseitige Anteilstests

Bei einem zweiseitigen Anteilstest wird die Alternativhypothese weder in der Form $p<p_0$ noch in der Form $p>p_0$ formuliert, sondern in der Form $p \neq p_0$. Man führt einen zweiseitigen Anteilstest durch, wenn vermutet wird, daß ein bestimmter relativer Anteil in einer Grundgesamtheit von einem behaupteten Wert p_0 zwar abweicht, doch nicht vermutet wird, daß dieser Anteil kleiner als p_0 ist bzw. größer als p_0 ist.

Man geht ähnlich vor wie bei einem einseitigen Anteilstest. Man formuliert zuerst eine Null- und eine Alternativhypothese:

$$H_0: p=p_0 \qquad H_1: p \neq p_0$$

Die Nullhypothese soll durch eine Stichprobe vom Umfang n geprüft werden. Vor der Erhebung der Stichprobe legt man eine Signifikanzzahl α_0 fest. Dann wird die Stichprobe erhoben. Wir nehmen an, daß sich für die entsprechende Häufigkeit H in der Stichprobe $H=k$ ergibt $(0 \le k \le n)$.

Man nimmt nun an, daß die Nullhypothese H_0 gilt und berechnet unter dieser Annahme die Wahrscheinlichkeiten $P(H \le k)$ und $P(H \ge k)$. (Man beachte, daß H annähernd binomialverteilt mit den Parametern n und p_0 ist - letzteres weil H_0 als gültig angenommen wird.) Die Nullhypothese H_0 wird man verwerfen, wenn der in der Stichprobe beobachtete Wert k von H unwahrscheinlich klein oder unwahrscheinlich groß ist, d.h. wenn $P(H \le k)$ sehr klein oder $P(H \ge k)$ sehr klein ist. Naheliegenderweise teilt man die maximal zugelassene Irrtumswahrscheinlichkeit α_0 je zur Hälfte auf die beiden Enden der Wahrscheinlichkeitsverteilung von H auf. Die Nullhypothese kann somit genau dann verworfen werden, wenn gilt:

$$P(H \le k) \le \frac{\alpha_0}{2} \vee P(H \ge k) \le \frac{\alpha_0}{2}$$

Betrachten wir ein Beispiel. Nehmen wir an, daß der vorhin betrachtete Kunde nur vermutet, daß der Anteil der Lose von 40% abweicht, aber nicht vermutet, daß dieser Anteil höher ist, und nicht vermutet, daß er geringer ist. Er findet in einer Stichprobe vom Umfang 10 genau 6 Nieten vor. Kann er die Behauptung des Geschäftsmannes verwerfen?

Wir formulieren eine Nullhypothese und eine Alternativhypothese:

$$H_0: p=0{,}4 \qquad H_1: p \neq 0{,}4$$

Wir wählen $\alpha_0=0{,}05$. In der Stichprobe hat sich $H=6$ ergeben. Unter der Annahme von H_0 ist H annähernd binomialverteilt mit $n=10$ und $p=0{,}4$. Aus einer Tabelle erhalten wir: $P(H \le 6) \approx 0{,}945$, $P(H \ge 6) \approx 0{,}166$. Die Bedingung

$$P(H \le 6) \le 0{,}025 \vee P(H \ge 6) \le 0{,}025$$

ist also nicht erfüllt. Die Nullhypothese H_0 kann somit nicht verworfen werden.

4. Schätzen von Anteilen

Bei einem Anteilstest wird eine bestimmte Behauptung über einen relativen Anteil in einer Grundgesamtheit geprüft. Vielfach liegt jedoch eine solche Behauptung nicht vor. Es soll vielmehr aufgrund der Daten einer Stichprobe eine Schätzung über den unbekannten relativen Anteil in der Grundgesamtheit abgegeben werden. Dies sei an folgender Aufgabe illustriert:

> Vor den Wahlen möchte jemand den relativen Anteil der Wähler einer bestimmten Partei feststellen. Er erhebt eine Stichprobe von 20 Wahlberechtigten und stellt fest, daß 6 von ihnen diese Partei wählen wollen. Schätze daraus den relativen Anteil der Wähler dieser Partei unter allen Wahlberechtigten!

Als Schätzwert kann der relative Anteil der Wähler dieser Partei in der Stichprobe genommen werden, also 6/20=0,3=30%. Eine solche Schätzung nennt man *Punktschätzung*, weil ein "punktueller Wert", d.h. eine Zahl, als Schätzung angegeben wird. Dieser Schätzwert kann jedoch vom "wahren" Wert in der Grundgesamtheit stark abweichen, da der Anteil der Wähler der betrachteten Partei von Stichprobe zu Stichprobe schwankt. In der Praxis bemüht man sich daher eher, eine ***Bereichsschätzung*** vorzunehmen, d.h. ein Intervall als Schätzung für den unbekannten relativen Anteil in der Grundgesamtheit anzugeben. Dabei werden auch gewisse vom Punktschätzwert abweichende Werte als Schätzwerte akzeptiert.

Welche Werte kann man als sinnvolle Schätzwerte für den unbekannten relativen Anteil in der Grundgesamtheit ansehen? Es liegt nahe, alle jene Werte p_0 als Schätzwerte für den unbekannten relativen Anteil in der Grundgesamtheit anzusehen, die durch einen zweiseitigen Anteilstest nicht verworfen werden können (genauer: für die die Nullhypothese $p=p_0$ durch einen zweiseitigen Anteilstest mit einer bestimmten Signifikanz α_0 aufgrund des Stichprobenergebnisses H=k nicht verworfen werden kann). Die Menge dieser Werte p_0 bezeichnet man als **Konfidenzintervall mit der Signifikanz** α_0 oder auch als **$(1-\alpha_0)$-Konfidenzintervall** für den unbekannten relativen Anteil in der Grundgesamtheit. Die Zahl $(1-\alpha_0)$ wird meist in Prozenten angegeben und als **Sicherheit** des Konfidenzintervalles bezeichnet. Z.B. spricht man für $\alpha_0=0{,}05$ von einem Konfidenzintervall mit der Sicherheit 0,95=95% bzw. einem 95%-Konfidenzintervall.

Wie kann man ein $(1-\alpha_0)$-Konfidenzintervall bestimmen? Wir setzen für die folgenden Überlegungen eine bestimmte Signifikanzzahl α_0 voraus und nehmen an, daß sich in einer Stichprobe vom Umfang n das Ergebnis H=k ergeben hat. Um das gesuchte $(1-\alpha_0)$-Konfidenzintervall zu bestimmen, müssen wir jene Werte p_0 bestimmen, für die die Nullhypothese H_0:$p=p_0$ durch einen zweiseitigen Anteilstest mit der Signifikanz α_0 aufgrund des Stichprobenergebnisses H=k nicht verworfen werden kann. Wir wissen aus dem vorigen Abschnitt, daß die Nullhypothese H_0 genau dann nicht verworfen werden kann, wenn gilt:

$$P(H\leq k)>\frac{\alpha_0}{2} \wedge P(H\geq k)>\frac{\alpha_0}{2}$$

Anhand einer entsprechenden Tabelle ist es recht einfach, jene p_0 zu bestimmen, für die diese Bedingung erfüllt ist. (Eine genauere Erklärung des Vorgehens findet man in BÜRGER/FISCHER/MALLE 1991, Seite 279-283.)

Als Beispiel bestimmen wir das 95%-Konfidenzintervall für den unbekannten relativen Anteil der Wähler der in der obigen Aufgabe betrachteten Partei. Es ist α_0=0,05, n=20 und k=6. Anhand einer geeigneten Tabelle stellen wir fest, daß die Bedingung

$$P(H\leq 6)>0{,}025 \wedge P(H\geq 6)>0{,}025$$

für alle p_0 mit $p_0 \in [0{,}15;0{,}50]$ erfüllt ist. (Die Genauigkeit der Grenzen hängt von der Genauigkeit der verwendeten Tabelle ab.) Somit ist dieses Intervall das gesuchte Konfidenzintervall.

Was bedeutet das Konfidenzintervall [0,15;0,50]? Heißt dies, daß der unbekannte relative Anteil der Wähler der betrachteten Partei in diesem Intervall liegt (d.h. 15% bis 50% beträgt)? Mit Sicherheit kann man dies leider nicht sagen. Es kann nämlich sein, daß der "wahre" Wert dieses relativen Anteils außerhalb dieses Intervalls liegt und durch den vorgenommenen zweiseitigen Anteilstest irrtümlich verworfen wurde. Die Wahrscheinlichkeit hiefür ist aber höchstens α_0=0,05=5%. Deutet man diese Wahrscheinlichkeit frequentistisch, erhält man folgende Interpretation des Konfidenzintervalles: Würde man sehr oft Stichproben vom Umfang 20 erheben und jedesmal auf die beschriebene Weise ein 95%-Konfidenzintervall ermitteln, so würde dieses Intervall in mindestens ca. 95% aller Fälle den unbekannten "wahren" Wert des relativen Anteils in der Grundgesamtheit überdecken. Ob dies allerdings im vorliegenden Fall erfüllt ist, kann man nicht sagen.

Literatur

BÜRGER/FISCHER/MALLE (1991): Mathematik Oberstufe 3. Hölder-Pichler-Tempsky, Wien.