

J Statistik

Statistische Verfahren ermöglichen es, ein Maximum an Informationen bei vergleichsweise geringem Aufwand und daher kostengünstig zu gewinnen. Man untergliedert derartige Verfahren grob in zwei Hauptbereiche: Verfahren zum quantitativen Beschreiben, Erfassen bzw. Kennzeichnen von Erscheinungen (Bereich der *beschreibenden Statistik*; vgl. Kapitel H) und Verfahren zum Beurteilen, Prüfen und Testen von Vermutungen bzw. Hypothesen. Der letztgenannte Bereich der *beurteilenden Statistik* oder auch *Prüfstatistik* ist Gegenstand des vorliegenden Kapitels. Ein Beispiel soll das Anliegen der Prüfstatistik illustrieren:

Noch im Mittelalter hielt sich hartnäckig die Auffassung, dass die Anzahl der Jungengeburten mit der Anzahl der Mädchengeburten übereinstimme – man müsse nur lange genug „Buch führen“. Entsprechende statistische Untersuchungen führte erstmals der schottische Arzt, Mathematiker und Schriftsteller John ARBUTHNOT (1667–1735) durch. Er wertete demografische Daten von mehr als 80 aufeinanderfolgenden Jahren aus und schlussfolgerte: Da in jedem dieser Jahre mehr Jungen als Mädchen geboren wurden, kann das nicht zufällig sein. Die Vermutung vom gleichen Anteil der Jungen- und Mädchengeburten ist als falsch zu betrachten.

In der modernen Prüfstatistik veranschaulicht man demografische Daten z. B. in so genannten Bevölkerungspyramiden (s. Fig. J 1). Aus einer mehr als 100-jährigen Geburtenstatistik ist heute bekannt, dass der Anteil der Jungengeburten 0,514 (Mädchengeburten 0,486) beträgt. Zur fortlaufenden (jährlichen) Aktualisierung der demografischen Daten erfasst beispielsweise jeder Landkreis, jedes Bundesland, wie viele Jungen und Mädchen (auf seinem Territorium) geboren wurden. Werden also in einem Landkreis unter 200 geborenen Kindern 102 Jungen gezählt, so scheint man offenbar „im Trend“ zu liegen. Wie viele Jungen (Mädchen) *müssen* nun aber anteilig (mindestens) unter einer bestimmten Anzahl von Neugeborenen sein, um die Geburtenstatistik „zu bestätigen“? Mit welcher Sicherheit darf man sich für eine Bestätigung entscheiden? (s. Beispiele J 7 und J 13)

Altersaufbau der Bevölkerung in Deutschland

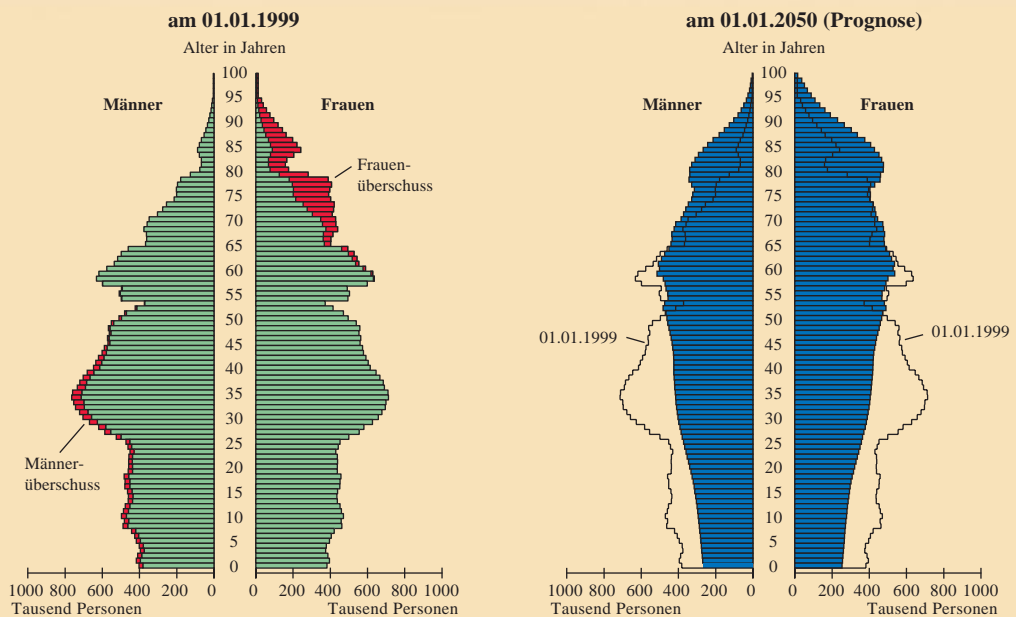


Fig. J 1

J 1 Testen von Hypothesen – Testverfahren

Die historischen Wurzeln (prüf-)statistischer Ermittlungen reichen bis in das Ägypten der Zeit um etwa 3050 v. Chr. zurück. Hier fanden erstmalig Volkszählungen statt, wie sie auch aus dem Römischen Reich (etwa 100 n. Chr.) und dem Inka-Reich (etwa 1200 n. Chr.) überliefert sind. In der Mitte des 17. Jahrhunderts begann man (vornehmlich in England und Deutschland) das Wirtschafts- und Sozialleben mit mathematischen Mitteln zu erfassen, darzustellen und Entwicklungsvorhersagen (Prognosen) abzuleiten. Im 19. Jahrhundert entstanden – aus praktischen Bedürfnissen heraus – insbesondere in Russland (TSCHEBYSCHEW, 1821–1894)¹⁾, England (GALTON, 1822–1911)²⁾ und Deutschland (LEXIS, 1837–1914)³⁾ gewisse statistische Traditionen. Weitestgehend unbeantwortet blieb jedoch hier noch die Frage nach dem Grad der Sicherheit statistischer Analyseergebnisse. Erst in den vierziger Jahren des 20. Jahrhunderts wurden die klassischen statistischen Zielsetzungen (Beschreibung durch Daten und deren Analyse) um gezielte Untersuchungen zur Sicherheit der Analyseergebnisse – dem Niveau der Signifikanz⁴⁾ der Ergebnisse, der Vermutungen bzw. Hypothesen – erweitert. Seitdem haben sich Signifikanztests bewährt, wie sie im Grundsätzlichen von FISHER (1890–1962), NEYMAN (1894–1981) und PEARSON (1895–1980) entwickelt worden sind.⁵⁾

J 1.1 Grundprobleme des Testens von Hypothesen

In der Statistik werden statistische (Daten-)Mengen untersucht und dabei ein interessierender statistischer Zusammenhang durch eine *Zufallsgröße*, z.B. die Zufallsgröße X , beschrieben.

J 1

Definition J 1:

Statistische Mengen sind Gesamtheiten von Ereignissen, Objekten oder Individuen.

Die Menge aller Ereignisse bzw. Objekte oder Individuen, die zu einem klar gekennzeichneten Merkmal (oder einer Merkmalsgruppe) gebildet werden kann, bezeichnet man als **Grundgesamtheit**, insbesondere bei Individuen auch als **Population**.

Beispiele für Grundgesamtheiten (und sich darauf beziehende Zufallsgrößen) wären so die Menge

- aller wahlberechtigten Bürger eines Bundeslandes
(Die Zufallsgröße X könnte die Anzahl der Bürger beschreiben, die Wähler einer bestimmten Partei sind.);
- aller Bäume eines Waldgebietes
(Zufallsgröße X : Anzahl der Bäume, die Schädigungen durch Umwelteinflüsse aufweisen);
- aller Artikel einer bestimmten Sorte aus der Tagesproduktion einer Firma
(Zufallsgröße X : Anzahl der unbrauchbaren Artikel);
- aller Erdbeben im Zeitraum von 100 Jahren in einem bebenintensiven Gebiet
(Zufallsgröße X : Anzahl der Beben ab einer bestimmten Stärke);

¹⁾ Pafnuti Lwowitsch TSCHEBYSCHEW; vielseitiger russischer Mathematiker; veröffentlichte zahlreiche auf praktische Anwendungen orientierte Arbeiten in der Wahrscheinlichkeitstheorie und bewies u.a. eine Verallgemeinerung des Gesetzes der großen Zahlen.

²⁾ Sir Francis GALTON; Vetter von Charles Darwin, gilt als bedeutender Vererbungsforscher des 19. Jahrhunderts und einer der Begründer der empirischen Humangenetik.

³⁾ Wilhelm LEXIS; Volkswirt und Statistiker, entwickelte Theorien zur Untersuchung statistischer Daten.

⁴⁾ Signifikanz – Bedeutsamkeit; hier: signifikant im Sinne von bedeutungsvoll, verallgemeinerungsfähig

⁵⁾ Die englischen Mathematiker Ronald Aylmer FISHER, Jerzy NEYMAN und Egon Sharpe PEARSON erwarben Verdienste in der Entwicklung der statistischen Schätz- und Testtheorie. FISHER stellte als Genetiker und Statistiker insbesondere Praxisbezüge her.

- aller Unfälle im Straßenverkehr innerhalb einer Stadt (Zufallsgröße X: Anzahl der betroffenen Fußgänger).

Bei statistischen Untersuchungen ist es im Allgemeinen aus praktisch-organisatorischen Gründen nicht möglich oder aus Kostengründen nicht erwünscht, eine interessierende Grundgesamtheit vollständig zu untersuchen. Man denke beispielsweise an

- Wahlprognosen, die selbstverständlich nicht die Wahl vorwegnehmen bzw. ersetzen können;
- Qualitätsprüfungen, die nicht zerstörungsfrei bzw. ohne Folgeschäden bleiben (wie Untersuchungen von Materialien auf Elastizität).

Aufgabe der beurteilenden Statistik ist es deshalb vielmehr, aus Eigenschaften von Teilmengen einer Grundgesamtheit (wobei die Wahrscheinlichkeitsverteilung des statistisch interessierenden Merkmals in der Grundgesamtheit unbekannt ist) die Wahrscheinlichkeit für das Auftreten eines bestimmten statistisch interessierenden Merkmals in der *Grundgesamtheit* zu *schätzen* und die Signifikanz des Schätzwertes zu beurteilen.

Definition J 2:

Eine aus einer Grundgesamtheit (i. Allg. zufällig – „auf gut Glück“) ausgewählte (Teil-)Menge mit n Elementen heißt **Stichprobe**. Die Elemente X_1, X_2, \dots, X_n der Stichprobe sind Zahlenwerte der Zufallsgröße X . Die Anzahl n der Elemente gibt den Umfang der Stichprobe an, kurz als **Stichprobenumfang** bezeichnet.

Jedes einzelne Element der Stichprobe heißt **Stichprobenwert**.

J 2

Um aus Eigenschaften der Stichprobe mit einer gewissen Sicherheit auf Eigenschaften der Grundgesamtheit schließen zu können, muss die Stichprobe charakteristisch – man sagt **repräsentativ**¹⁾ – für die Grundgesamtheit sein. Darüber hinaus müssen die interessierenden Eigenschaften der Elemente der Stichprobe quantifizierbar, also zahlenmäßig erfassbar und beschreibbar sein. Das Erfassen und Beschreiben übernimmt die beschreibende Statistik. Die Untersuchung der Stichprobe mithilfe von Schätz- und Testverfahren (einschließlich Entscheidungen und Angaben zu deren Zuverlässigkeit) leistet die beurteilende Statistik.

Der erste wichtige Schritt einer Untersuchung ist die genaue Festlegung bzw. Kennzeichnung der **Grundgesamtheit**. Der zweite Schritt besteht in der Planung der Zusammensetzung der **Stichprobe**. Um **Repräsentativität** zu erreichen, dürfen ihre Zusammensetzung und ihr Umfang nicht dem Zufall überlassen bleiben; das Ermitteln ihrer einzelnen Elemente erfolgt dagegen zufällig. Für einen hinreichend großen Stichprobenumfang gibt der so genannte **Auswahlsatz a** eine Orientierung.

$$\text{Auswahlsatz } a = \frac{\text{Umfang der Stichprobe } n}{\text{Umfang der Grundgesamtheit } N} \cdot 100 \%$$

(Der Umfang der Grundgesamtheit N muss ggf. geschätzt werden.)

Für den Auswahlsatz a existieren empirisch gewonnene Erfahrungswerte. Diese Werte variieren z.B. in Abhängigkeit von der Zusammensetzung einer Stichprobe sowie der Art des Sachgebietes der Grundgesamtheit. Als grober Richtwert kann $a = 10 \%$ angesehen werden.²⁾

¹⁾ Eine Stichprobe gilt als repräsentativ, wenn sie annähernd so wie die Grundgesamtheit zusammengesetzt und der Stichprobenumfang hinreichend groß ist.

²⁾ In der statistischen Praxis sind sowohl erheblich kleinere a -Werte (z.B. bei Wahlprognosen; $a < 1 \%$) als auch erheblich größere Werte (z.B. bei Qualitätskontrollen; $a > 20 \%$) zu finden. Dies begründet sich aus entsprechenden jahrzehntelangen Erfahrungen (Wahlprognosen) oder ständig wechselnder Spezifik und daher fehlender Erfahrung (Qualitätskontrollen) bei der Zusammensetzung von Stichproben.

Vorausgesetzt, eine Stichprobe ist geeignet zusammengesetzt, dann gilt: Je größer der Auswahlatz, desto sicherer die Repräsentativität der Stichprobe.

In der Stichprobe werden n -mal wiederholte Beobachtungen ein- und derselben Zufallsgröße zusammengefasst. Variieren die Beobachtungsergebnisse in nicht vorhersagbarer Weise (Zufälligkeit der Beobachtungsergebnisse) und beeinflussen sie einander nicht (Unabhängigkeit der Beobachtungsergebnisse¹⁾), so hebt man dies gelegentlich auch durch die Verwendung des Begriffes **Zufallsstichprobe** besonders hervor.²⁾ Als häufige Auswahlformen von (Zufalls-)Stichproben seien die **Klumpenstichprobe** und die (proportional) **geschichtete Stichprobe** genannt. Eine *Klumpenstichprobe* setzt sich stets aus allen Elementen von mindestens zwei Klumpen zusammen. Der Begriff „Klumpen“ wird hier im Sinne von Teilmengen (Ziehen mehrerer Elemente auf einen Griff, Ziehen als „Klumpen“) gebraucht. Beim Untersuchen einer Klumpenstichprobe untersucht man alle Elemente aller Klumpen.

Beispiele für Klumpenstichproben:

- Die Qualität eines bestimmten Produktes soll geprüft werden (z.B.: Diskette fehlerfrei oder fehlerhaft). Ist das Produkt etwa in Großpackungen mit je 100 Kleinpackungen zu je 10 Produkten verpackt, könnten über einen längeren Zeitraum aus jeder Tagesproduktion zwei Großpackungen zufällig ausgewählt und diesen wiederum jeweils drei Kleinpackungen (ein Klumpen) zur Prüfung zufällig entnommen werden. Die im Laufe eines bestimmten Beobachtungszeitraumes (z.B. ein Monat) entnommenen („gezogenen“) Kleinpackungen bilden eine Klumpenstichprobe.

- Der Gebisszustand 13-Jähriger soll beurteilt werden.

Zur Beurteilung können an einigen zufällig ausgewählten Schulen die (13-jährigen) Schülerinnen und Schüler zufällig ausgewählter 7. Klassen (jede Klasse ist ein Klumpen) zahnärztlich untersucht werden. Die untersuchten Schülerinnen und Schüler der einbezogenen Schulen bilden eine Klumpenstichprobe.

Eine geschichtete Stichprobe weist in voller Absicht dieselbe Zusammensetzung (Grundstruktur) wie die Grundgesamtheit auf. Man spricht daher auch von einem verkleinerten Abbild oder einer Mikrokopie der Grundgesamtheit.

Beispiele für (proportional) geschichtete Stichproben:

- Ein Wahlausgang soll (z.B. durch ein Meinungsforschungsinstitut) prognostiziert werden.

Meinungsforschungsinstitute verfügen i. Allg. aus jahrelanger Erfahrung über detaillierte Informationen zur Zusammensetzung der Grundgesamtheit. Wahlprognosen entstehen auf der Basis umfassender Kenntnisse über das Wahlverhalten der verschiedenen (wahlberechtigten) Bevölkerungsgruppen. Unterschiede in der Grundgesamtheit wie regionale Unterschiede, Unterschiede nach Geschlecht, Alter, Beruf, ... werden adäquat in der Stichprobe berücksichtigt: Es wird z.B. darauf geachtet, dass die verschiedenen Bevölkerungsgruppen prozentual wie in der Grundgesamtheit vertreten sind. Zufällig bleibt, wen man aus jeder Bevölkerungsgruppe für die Stichprobe auswählt.

- Schülerinnen und Schüler an Gymnasien sollen nach ihrer durchschnittlichen wöchentlichen Arbeitszeit am Computer befragt werden.

Ist die Grundgesamtheit (Gesamtschülerzahl mehrerer Gymnasien) nicht allzu groß, könnten im Interesse einer sicheren Repräsentativität der Befragung alle Schülerinnen und Schüler einbezogen werden. Eine geeignet geschichtete Stichprobe müsste – prozentual der Grundgesamtheit entsprechend – anteilig zufällig ausgewählte Mädchen und Jungen aus den Klassen einer bestimmten Anzahl Gymnasien (ggf. aus Großstädten, Städten mittlerer Größe und Kleinstädten sowie ländlicher Gegend) umfassen.

Insbesondere das letzte Beispiel zeigt, wie wichtig eine geeignete Zusammensetzung der Stichprobe für deren Repräsentativität ist. Würde sich etwa die (zufällige) Auswahl der Befragten auf die Jungen

¹⁾ Das heißt: Unabhängig davon, welche Elemente bereits für die Stichprobe „auf gut Glück“ ausgewählt worden sind, kann jedes Element der Grundgesamtheit mit gleicher Wahrscheinlichkeit ausgewählt werden.

²⁾ Wir verwenden den Begriff *Stichprobe* (Definition J 2) hier stets synonym mit *Zufallsstichprobe*.

zweier achter Klassen eines Gymnasiums beschränken, wäre günstigenfalls eine (vage) Aussage über Jungen dieser Altersgruppe (und nur für dieses Gymnasium) sinnvoll.

In jedem Falle soll eine *Zufallsstichprobe* erzeugt werden. Diese Forderung ist durch die Urnenmodelle *Ziehen mit Zurücklegen* bzw. *Ziehen ohne Zurücklegen* beschreibbar. Beim *Ziehen ohne Zurücklegen* ist jedoch die obige Forderung nach Unabhängigkeit verletzt. Praktisch lässt sich hier Unabhängigkeit aber immer dann gewährleisten, wenn die Anzahl der gezogenen Elemente im Vergleich zur Anzahl der Elemente der Grundgesamtheit (Gesamtzahl der Kugeln jeder Art in der Urne) hinreichend klein bleibt. Bei einer verhältnismäßig großen Anzahl der Elemente der Grundgesamtheit ändert ein *Ziehen ohne Zurücklegen* von vergleichsweise wenigen Elementen die Zusammensetzung der Grundgesamtheit praktisch nicht, so dass die Art des Ziehens für die Unabhängigkeit nicht relevant ist.

Beurteilende Statistik setzt zunächst quantitatives Beschreiben voraus. Am folgenden Beispiel werden die Möglichkeiten einer mathematischen Beschreibung von Stichproben aus einer Grundgesamtheit betrachtet und weitere Grundbegriffe der beurteilenden Statistik erarbeitet. Vollständig bearbeiten werden wir das Beispiel im Abschnitt J 1.2.

Beispiel J 1:

Ein Zulieferbetrieb stellt Federsätze zum „Tieferlegen“ eines bestimmten Fahrzeugtyps her. Aus langjährigen Erfahrungen weiß man, dass aufgrund des komplizierten Fertigungsprozesses 30 % aller Federsätze Ausschuss sind. Durch Verwendung höherwertiger Materialien und Verbesserungen der Fertigungstechnologie will man den Ausschuss drastisch reduzieren. Stichproben von jeweils 20 nach verbesserter Technologie und aus höherwertigem Material gefertigten Federsätzen wiesen höchstens zwei Ausschuss-Federsätze (10 %) auf.

Es interessiert, welche Schlussfolgerungen aus dieser Stichprobe gezogen werden können und wie sicher diese sind. Insbesondere bedeutsam ist, mit welcher Sicherheit aus dem Prüfergebnis (Stichprobe) tatsächlich auf eine Qualitätsverbesserung geschlossen werden kann.

Mögliche mathematische Modellierung:

In einer Urne mit hinreichend vielen weißen und schwarzen Kugeln (keine anderen Kugeln in der Urne) symbolisieren die weißen Kugeln die fehlerfreien, die schwarzen Kugeln die fehlerhaften Federsätze. Der Anteil p der schwarzen Kugeln ist dabei unbekannt. Es werden 20 Kugeln *mit Zurücklegen* gezogen (Stichprobe)¹⁾. Die Anzahl der schwarzen Kugeln unter den 20 Kugeln interessiert.

J 1

Im voranstehenden Beispiel beschreibe die Zufallsgröße X die Anzahl der schwarzen Kugeln unter allen Kugeln in der Urne. Die Zufallsgröße X kann folgendermaßen definiert werden:

$$X = \begin{cases} 1, & \text{wenn eine schwarze Kugel gezogen wird} \\ 0, & \text{wenn keine schwarze Kugel gezogen wird} \end{cases}; \quad X \text{ ist eine BERNOULLI-Größe (vgl. Abschnitt H 5.1)}$$

Das entsprechende BERNOULLI-Experiment (*Ziehen mit Zurücklegen* aus einer Urne) werde n -mal (im Beispiel $n = 20$) durchgeführt. Weil die einzelnen Ziehungen unabhängig voneinander sind, weisen die zugehörigen Zufallsgrößen X_1, X_2, \dots, X_n dieselbe Wahrscheinlichkeitsverteilung wie die Zufallsgröße X auf. Jede Stichprobe vom Umfang n (BERNOULLI-Kette der Länge n) kann somit als ein n -Tupel $(x_1; x_2; \dots; x_n)$ mit $x_i \in \{0; 1\}$ für $i = 1; 2; \dots; n$ aufgefasst werden. Unsere obige Stichprobe (Umfang $n = 20$) ließe sich also beispielsweise durch nachstehende 20-Tupel beschreiben:

(0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 0; 0), (1; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 0) usw.

¹⁾ Praktisch kann die Stichprobe von 20 Federsätzen durchaus mittels *Ziehen ohne Zurücklegen* erzeugt werden.

Entscheidend ist: Unter den 20 gezogenen Kugeln sind genau 18 weiße („0“) und zwei schwarze („1“). Da nicht interessiert, an welcher Stelle die schwarzen Kugeln gezogen worden sind, ist die Untersuchung als 20-elementige Menge (mit 18-mal „0“ und zweimal „1“) $\{0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 0; 0\}$ für die weitere Analysetätigkeit ausreichend.

Die praktische Fragestellung lässt sich „übersetzen“: Aus der Anzahl der schwarzen Kugeln in der Stichprobe soll auf den unbekannten Anteil p der schwarzen Kugeln in der Urne geschlossen werden. Dieser Anteil wird durch die (unbekannte) Wahrscheinlichkeit p beschrieben. Ob es möglich ist, eine Qualitätsverbesserung zu schlussfolgern, hängt davon ab, *mit welcher Sicherheit (Signifikanz)* dieser Anteil unter oder ggf. über dem bisherigen (Ausschuss $p = 0,3$ bzw. $p = 30\%$) liegt. Für die Signifikanz kann durch Vorgabe ein bestimmtes Niveau gefordert werden oder das Signifikanzniveau ist zu ermitteln.

Die Anzahl der schwarzen Kugeln in der Stichprobe ist binomialverteilt, da die Entnahme einer festen Anzahl von n Kugeln aus der Grundgesamtheit zufällig und *mit Zurücklegen* erfolgt. Für die unbekannte Wahrscheinlichkeit p , dass bei einmaligem Ziehen aus der Urne eine schwarze Kugel gezogen wird, lassen sich – unter Beachtung der Stichprobe – **begründete Vermutungen**, man sagt: **Hypothesen**, ableiten. Wegen der Stichprobe (2 von 20, also $\frac{2}{20}$), wäre $p = 0,1$ eine Hypothese.¹⁾ Mit Blick auf den früheren Anteil schwarzer Kugeln (Ausschuss $p = 0,3$) könnte ebenso die Hypothese $p < 0,3$ formuliert werden. Für den vorliegenden konkreten Fall wäre hingegen eine Vermutung $p > 0,3$ i. Allg. nicht begründet (Widerspruch zum Einsatz höherwertigen Materials und verbesserter Technologie) und würde daher als „normale“ Hypothese ausscheiden.²⁾

J 3

Definition J 3:

Die zu überprüfende bzw. zu beurteilende Hypothese heißt **Nullhypothese H_0** ³⁾. Die Verneinung (die Negation, das Gegenteil) der Nullhypothese wird **Alternativhypothese** oder **Gegenhypothese** genannt und mit **H_1** oder auch **\bar{H}** bezeichnet. Nullhypothese und Alternativ- bzw. Gegenhypothese sind konkurrierende (einander ausschließende) Hypothesen.

Auf der Grundlage statistischer Tests (s. Abschnitt J 1.2 und J 1.3) wird entschieden, ob die Nullhypothese abzulehnen (zu verwerfen) ist oder nicht. Im Allgemeinen versucht man (aus historisch gewachsenem „Sicherheitsdenken“ heraus), die Nullhypothese abzulehnen und somit die Gegenhypothese anzunehmen. Die Entscheidung, ob eine Hypothese abzulehnen ist, bleibt stets kompliziert, weil dabei auf der Basis einer *Stichprobenuntersuchung* entschieden wird, während die Hypothese die Verhältnisse in der *Grundgesamtheit* beschreibt. Offenbaren die Untersuchungsergebnisse der Stichprobe extreme Abweichungen von der Nullhypothese, so spricht man von einem *signifikanten Unterschied* zwischen der Nullhypothese und (den Untersuchungsergebnissen) der Stichprobe. Die Nullhypothese *ist abzulehnen* (zu verwerfen). Lassen sich aus der Stichprobe keine signifikanten Abweichungen nachweisen, darf nicht geschlussfolgert werden, dass die Nullhypothese richtig sei.

¹⁾ Hypothesen, die durch genau einen Wert ($p = p_0$) festgelegt sind, nennt man *einfache* Hypothesen im Unterschied zu Hypothesen der Form $p \neq p_0$ (bzw. $p < p_0$; $p > p_0$), die als *zusammengesetzte* Hypothesen bezeichnet werden.

²⁾ Dies heißt jedoch nicht, dass dieser Fall praktisch unmöglich ist. Es wird lediglich festgestellt, dass er nicht zu vermuten ist, weil sein Eintreten unter den gegebenen Bedingungen nicht begründet werden kann.

³⁾ Die üblich gewordene Begriffsbildung „Nullhypothese“ soll verdeutlichen: Die Nullhypothese geht i. Allg. davon aus, dass die unbekannte Wahrscheinlichkeitsverteilung (in der Grundgesamtheit) mit der vermuteten Verteilung (aus der Stichprobe gewonnen) tatsächlich übereinstimmt; zwischen Vermutung und Tatsache besteht dann „die Differenz null“.

Sie steht nur nicht im (offensichtlichen) Widerspruch zu den Untersuchungsergebnissen und *kann* mit Blick darauf lediglich *nicht abgelehnt werden*. Die Entscheidung bleibt also stets mit einer gewissen Unsicherheit behaftet. Daher ist es wichtig, die Wahrscheinlichkeiten für Fehlentscheidungen zu kennen, sie sinnvoll festzulegen, zu berechnen oder abschätzen zu können.

Die Wahrscheinlichkeit, mit der man bereit ist bzw. das Risiko eingeht, eine *in Wirklichkeit wahre* Nullhypothese irrtümlich als falsch abzulehnen, nennt man **Irrtumswahrscheinlichkeit** α . Sie wird meist vor Beginn des Tests festgelegt oder aus Testdaten berechnet und häufig auch als **Signifikanzniveau** α , **α -Fehler** bzw. **Fehler 1. Art** oder **Risiko 1. Art** bezeichnet.¹⁾ Statistische Tests gestatten das Berechnen der Werte der Zufallsgröße X , für die die Nullhypothese abgelehnt wird. Die Menge dieser Werte aus dem Wertebereich der Zufallsgröße X heißt **Ablehnungsbereich** \bar{A} (Verwerfungsbereich oder kritischer Bereich). Die Menge der verbliebenen X -Werte bildet den **Annahmebereich** A . Auch ein weiterer Fehler ist möglich: Nimmt die Zufallsgröße X einen Wert aus dem Annahmebereich an, so dass eine *in Wirklichkeit falsche* Nullhypothese nicht abgelehnt werden kann, begeht man ebenfalls einen Fehler. Dieser Fehler heißt **β -Fehler** bzw. **Fehler 2. Art** oder **Risiko 2. Art**.

Bezüglich der **Fehler beim Testen von Hypothesen** lässt sich damit feststellen: **Fehlentscheidungen** treten auf, wenn

	Hypothese H_0 in Wirklichkeit	
	wahr	falsch
H_0 wird abgelehnt	Entscheidung falsch Fehler 1. Art	Entscheidung richtig
H_0 wird nicht abgelehnt	Entscheidung richtig	Entscheidung falsch Fehler 2. Art

Die Hypothesenprüfung bezieht sich auf Hypothesen, die für die Grundgesamtheit gültig sein sollen, aus der die untersuchte Stichprobe entnommen wird. Solche Hypothesen können sich beziehen auf

- eine unbekannte Wahrscheinlichkeit,
- einen unbekannten Parameter (z. B. p bei der Binomialverteilung $B_{n, p}$ oder μ, σ^2 bei der Normalverteilung N_{μ, σ^2}),
- den Typ der Wahrscheinlichkeitsverteilung (z. B. Normalverteilung),
- die Gleichheit unbekannter Wahrscheinlichkeitsverteilungen oder
- die Unabhängigkeit von Zufallsgrößen.

J 1.2 Testen einer unbekannten Wahrscheinlichkeit; Alternativtests

Verteilungsannahmen (z. B. Hypothesen zu unbekannten Wahrscheinlichkeiten) über Merkmale einer zu untersuchenden Grundgesamtheit werden mithilfe statistischer Tests, sog. Signifikanztests, anhand konkreter Stichproben überprüft. Basis der Überprüfungen ist die Nullhypothese. Der mathematische Aufbau der Signifikanztests erfolgt so, dass genau zwei Prüfergebnisse möglich sind: *die Nullhypothese ist abzulehnen oder die Nullhypothese kann nicht abgelehnt werden*.

Für den Fall, dass die Nullhypothese abzulehnen ist, legt i. Allg. die Alternativhypothese fest, wie das „Nichtgültigsein“ der Nullhypothese zu deuten ist. Sind in einem Test beide Hypothesen einfache Hypothesen (s. Fußnote 1, S. 444), also durch jeweils genau einen konkreten Wert formuliert,

¹⁾ In Abhängigkeit von den konkreten Gegebenheiten und den Ansprüchen an die Sicherheit der Testergebnisse sind Irrtumswahrscheinlichkeiten von 1 % ($\alpha = 0,01$) bis 10 % ($\alpha = 0,1$) allgemein üblich. Sehr häufig wird mit 5 % ($\alpha = 0,05$) gearbeitet.

spricht man von einem *besonderen Signifikanztest*, dem **Alternativtest**, anderenfalls (nur) von einem (normalen) **Signifikanztest**. Der Alternativtest ist sozusagen ein Signifikanztest von Nullhypothese kontra Alternativhypothese. Wegen der eindeutigen Festlegung beider Hypothesen lässt sich für die Signifikanzbeurteilung sowohl der Fehler 1. Art als auch der Fehler 2. Art eindeutig berechnen.¹⁾ Bei einem (normalen) Signifikanztest kann der Fehler 2. Art nicht eindeutig berechnet werden, da (zumindest) die Alternativhypothese nicht eindeutig (nicht durch genau einen Wert) festgelegt ist.²⁾

Ausgehend von Beispiel J 1 sei nun das grundsätzliche Vorgehen bei Alternativtests dargestellt. Zunächst ist zu entscheiden, welche **Testkonstruktion** für die statistische Überprüfung des vorliegenden *konkreten Sachverhalts* am geeignetsten ist. Dies erfordert, insbesondere die folgenden Fragen mit viel Feingefühl und „logischem Geschick“ hinsichtlich der mathematischen Übersetzung des Sachverhaltes und der angestrebten Signifikanz zu beantworten:

- Was wird als Nullhypothese formuliert, was als Alternativhypothese?
- Wo will (muss) man sehr vorsichtig sein, wo kann (darf) man größere Risiken eingehen?

Im Beispiel J 1 sprechen jahrelange Erfahrungen mit alter Technologie und bisherigem Material für einen Ausschuss von 30 % ($p = 0,30$). Es ist anzunehmen, dass bei verbesserter Technologie und höherwertigem Material der Ausschuss *im ungünstigsten Fall* weiterhin 30 % betragen wird. Demgegenüber steht das Ergebnis der Stichprobe, aus dem wohl *im günstigsten Fall* $p = 0,1$ geschlossen werden kann. Richtig ist es, eher vorsichtig zu bleiben, da langjährigen Erfahrungen nur eine relativ kurzfristige (Stichproben-)Erfahrung gegenübersteht: Für das Ansehen des Zulieferbetriebes ist es weniger schädlich, eine in Wirklichkeit *bessere* Qualität zu Unrecht als *schlechter* zu kennzeichnen (Es wird ein *Fehler 1. Art* begangen!). Dagegen ist es in hohem Maße imageschädigend, eine in Wirklichkeit *schlechtere* Qualität als *besser* auszugeben (Es wird ein *Fehler 2. Art* begangen!).

Es werden (genau) zwei miteinander konkurrierende Hypothesen formuliert, nämlich als Nullhypothese $H_0: p_0 = 0,1$ und als Gegenhypothese $H_1: p_1 = 0,3$.³⁾ Die Wahl von $p = 0,1$ (also einer Qualitätsverbesserung) als Nullhypothese steht einerseits für die Vermutung (die erhoffte Qualitätsverbesserung), bringt andererseits aber gerade die gebotene Vorsicht zum Ausdruck: Im Test wird stets versucht, die Nullhypothese abzulehnen und somit die Alternativhypothese anzunehmen.

H_0 und H_1 sind einfache Hypothesen. Für das im Beispiel J 1 erarbeitete Urnenmodell heißt dies: Der Anteil der schwarzen Kugeln (fehlerhafte Federsätze) in der Urne beträgt *entweder* 10 % *oder* 30 %. Auf der Basis einer Stichprobe (es kann die bereits vorliegende oder auch eine neue sein) ist nach n-maligem Ziehen einer Kugel (mit Zurücklegen) zu entscheiden, ob der Anteil 10 % oder 30 % beträgt.

Bei einer Stichprobe mit dem Umfang $n = 20$ erhält man X schwarze Kugeln. Die zufällige Anzahl X ist binomialverteilt mit den Parametern $n = 20$ und unbekanntem p . Welcher Wert p zutrifft, muss *alternativ* zwischen den beiden (einfachen) Hypothesen H_0 und H_1 – abgesichert durch einen Signifikanztest – entschieden werden.

Mit Bezug auf die (angenommene) Art der Wahrscheinlichkeitsverteilung der Zufallsgröße X (hier Binomialverteilung) schreibt man die Hypothesen häufig auch in einer ausführlichen Form.

¹⁾ Dies charakterisiert die Besonderheit des Alternativtests gegenüber dem (normalen) Signifikanztest.

²⁾ Man sagt auch: Bei einem (normalen) Signifikanztest ist die *Alternativhypothese nicht spezifiziert*. Die fehlende Spezifikation der Alternativhypothese gilt bei statistischen Tests als der *Normalfall*.

³⁾ Prinzipiell ist es bei einem Alternativtest (aus „reinmathematischer“ Sicht) gleichgültig, welche der beiden Hypothesen als Nullhypothese gewählt wird, da man sowohl den Fehler 1. Art als auch den Fehler 2. Art eindeutig berechnen kann.

$H_0: X \sim B_{20; 0,1}; H_1: X \sim B_{20; 0,3}$ (gesprochen: X ist binomialverteilt mit den Parametern $n = 20$ und $p_0 = 0,1$ bzw. X ist binomialverteilt mit den Parametern $n = 20$ und $p_1 = 0,3$).

Definition J 4:

Ein statistischer Test auf signifikante Unterschiede (Signifikanztest), bei dem zwischen zwei einfachen Hypothesen alternativ (für den einen oder den anderen konkreten Wert) entschieden wird, heißt „klassischer“ **Alternativtest**, kurz: **Alternativtest**.

J 4

Für die weitere Testkonstruktion gibt es nun zwei prinzipielle Möglichkeiten:

- (1) Man legt sinnvoll fest, ab welcher Anzahl defekter Federsätze in der Stichprobe (Anzahl X der gezogenen schwarzen Kugeln) die Nullhypothese abgelehnt werden soll (Kennzeichnung ihres Annahme- bzw. Ablehnungsbereiches) und ermittelt daraus das zugehörige Signifikanzniveau (Fehler 1. Art) sowie den Fehler 2. Art.
- (2) Man geht von einem bestimmten vorgegebenen oder gewählten Signifikanzniveau aus und ermittelt daraus den zugehörigen Annahme- bzw. Ablehnungsbereich für die Nullhypothese sowie den Fehler 2. Art.

Bei *Möglichkeit (1)* wird man sich für H_0 (gleichbedeutend mit einer Qualitätssteigerung) entscheiden, wenn die Anzahl X der schwarzen Kugeln (Anzahl der defekten Federsätze) sehr klein bleibt. Beispielsweise könnte festgelegt werden:

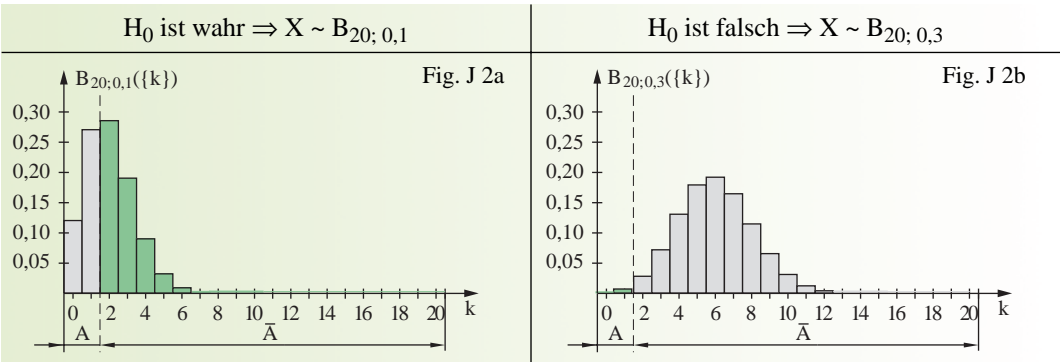
Die „kritische Anzahl“ X , der „Grenzwert“ für X , sei 2, also Annahmebereich $A = \{0; 1\}$. Das heißt: H_0 kann nicht abgelehnt werden, wenn $X < 2$ gilt. Es erfolgt eine Entscheidung für H_0 (also gegen die Alternativhypothese). Ablehnungsbereich: $\bar{A} = \{2; 3; \dots; 20\}$. Das heißt: H_0 ist abzulehnen, wenn $X \geq 2$ gilt. In diesem Fall erfolgt eine Entscheidung gegen H_0 (also für die Alternativhypothese).

Aus diesen Festlegungen oder Entscheidungsregeln ist sofort zu schlussfolgern: Da $X = 2$ (siehe Stichprobe) zum Ablehnungsbereich gehört, ist die Nullhypothese abzulehnen. Mit der Entscheidung gegen die Nullhypothese H_0 haben wir uns automatisch für die Alternativhypothese H_1 entschieden. Die Frage nach einer (signifikanten) Qualitätsverbesserung muss demnach bei einer *derartigen* Testkonstruktion verneint werden. Offen bleibt die Frage, wie sinnvoll diese Testkonstruktion tatsächlich ist, d.h., mit welcher Wahrscheinlichkeit jeweils die beiden möglichen Fehlentscheidungen (H_0 wird abgelehnt, obwohl sie in Wirklichkeit wahr ist bzw. H_0 kann nicht abgelehnt werden, obwohl sie in Wirklichkeit falsch ist) auftreten. Es ist also zu ermitteln, auf welchem Signifikanzniveau α (Fehler 1. Art; α -Fehler) und mit welchem Fehler 2. Art (β -Fehler) wir bei dieser Testkonstruktion entschieden haben.

Überlegungen zum Ermitteln des Fehlers 1. Art und des Fehlers 2. Art:

Die Stichprobe stellt eine BERNOULLI-Kette der Länge $n = 20$ (Stichprobenumfang) dar. Die Wahrscheinlichkeitsverteilung p der Zufallsgröße X ($X = X_1 + X_2 + \dots + X_{20}; 0 \leq X \leq 20$) ist unbekannt. Aufgrund der Stichprobe wird $p = 0,1$ vermutet, die Zufallsgröße X also binomialverteilt mit den Parametern $n = 20$ und $p = 0,1$ angenommen. Da somit für den unbekannten Parameter p der Wert aus der Nullhypothese $p_0 = 0,1$ gesetzt worden ist, gilt bei (in Wirklichkeit) *wahrer* Nullhypothese $X \sim B_{20; 0,1}$. Bei (in Wirklichkeit) *falscher* Nullhypothese würde jedoch $X \sim B_{20; 0,3}$ gelten.¹⁾ Die folgende Übersicht verdeutlicht noch einmal die Zusammenhänge und Berechnungen:

¹⁾ Man bedenke bei allen Überlegungen, dass stets von der Nullhypothese ausgegangen wird. Alle Überlegungen basieren auf der Nullhypothese, mit dem Bestreben, diese Hypothese abzulehnen.



H_0 ist abzulehnen	$P(\bar{A}_{p_0}) = B_{20; 0,1}(\{2; 3; \dots; 20\})$ $P(\bar{A}_{p_0}) = 1 - P(A_{p_0})$ $= 1 - B_{20; 0,1}(\{0; 1\})$ $= 1 - 0,39175 = 0,60825$ $P(\bar{A}_{p_0}) \approx 0,61$	$P(\bar{A}_{p_1}) = B_{20; 0,3}(\{2; 3; \dots; 20\})$ $P(\bar{A}_{p_1}) = 1 - P(A_{p_1})$ $= 1 - B_{20; 0,3}(\{0; 1\})$ $= 1 - 0,00764 = 0,99236$ $P(\bar{A}_{p_1}) \approx 0,99$
Schlussfolgerung	Bei <i>Annahme</i> von H_0 irrt man nicht.	Bei <i>Ablehnung</i> von H_0 irrt man nicht.
Fehler	Man begeht einen Fehler (1. Art), wenn die Nullhypothese (hier mit einer Wahrscheinlichkeit von 0,61) <i>als falsch</i> abgelehnt wird (obwohl sie wahr ist). Fehler 1. Art: $\alpha = 0,61$; $\alpha = 61\%$	Man begeht einen Fehler (2. Art), wenn die Nullhypothese (hier mit einer Wahrscheinlichkeit von 0,01) <i>nicht als falsch</i> abgelehnt wird (obwohl sie falsch ist). Fehler 2. Art: $\beta = 0,01$; $\beta = 1\%$

Zur Ermittlung der Werte $B_{n; p}(\{k\})$ kann die *Tabelle* der summierten (kumulierten) Binomialverteilung (s. Abschnitt H 5.3) verwendet oder mit einem geeigneten Rechner gearbeitet werden. Letzteres ist besonders vorteilhaft, da man so Werte für beliebige (sinnvolle) n und p sofort ermitteln kann, während Tabellen nur ausgewählte n und p erfassen. Fig. J 3 gibt das Schirmbild für den Fall an, dass eine Funktion **bern** für die Binomialverteilung und eine Funktion **kum** für die summierte Binomialverteilung definiert wurde.

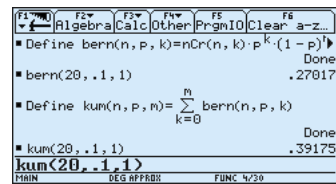


Fig. J 3

Verallgemeinernd lässt sich somit feststellen:

Satz J 1: **Wahrscheinlichkeit für den Fehler 1. Art**

Die summierte Wahrscheinlichkeit des *Ablehnungsbereiches* einer Nullhypothese ($H_0: p = p_0$) unter der Bedingung $X \sim B_{n; p_0}$ ist als Maß dafür anzusehen, wie wahrscheinlich es ist, einen *Fehler 1. Art* (α -Fehler) zu begehen. Mit dieser Wahrscheinlichkeit wird die in Wirklichkeit *wahre Nullhypothese* irrtümlich *abgelehnt*.

Es gilt: $\alpha = P(\bar{A}_{p_0}) = B_{n; p_0}(\bar{A}) = 1 - B_{n; p_0}(A)$

J 1

Satz J 2: **Wahrscheinlichkeit für den Fehler 2. Art**

Die summierte Wahrscheinlichkeit des *Annahmebereiches* einer Nullhypothese ($H_0: p = p_0$) unter der Bedingung $X \sim B_{n; p_1}$ ist als Maß dafür anzusehen, wie wahrscheinlich es ist, einen *Fehler 2. Art* (β -Fehler) zu begehen. Mit dieser Wahrscheinlichkeit wird die in Wirklichkeit *falsche Nullhypothese* irrtümlich *nicht abgelehnt*.

Es gilt: $\beta = P(A_{p_1}) = B_{n; p_1}(A) = 1 - B_{n; p_1}(\bar{A})$

J 2

Wir interpretieren nun die oben berechneten Wahrscheinlichkeiten für die beiden möglichen Fehlentscheidungen mit Blick auf den untersuchten Sachverhalt.

Der Fehler 1. Art (Wert $\alpha = 0,61$ bzw. $\alpha = 61\%$) besagt (verschieden formuliert):

- Der Ablehnungsbereich ist so (sehr groß) gewählt worden, dass die vermutete Qualitätsverbesserung (bei wahrer Nullhypothese H_0) auf einem Signifikanzniveau von $\alpha = 0,61$ (irrtümlich) abzulehnen ist.
- Beim Registrieren von mehr als einem defekten Federsatz (unter 20 untersuchten Federsatzes) spricht man sich mit der Wahrscheinlichkeit $\alpha = 0,61$ bzw. $\alpha = 61\%$ gegen die (in Wirklichkeit erreichte) Qualitätsverbesserung aus.
- In 61 % aller Stichproben (bei einer Wiederholung des Zufallsexperiments „Ziehen einer Stichprobe“ in genügend großer Anzahl) würden die Stichprobenergebnisse gegen die (in Wirklichkeit erreichte) Qualitätsverbesserung sprechen.

Der Fehler 2. Art (Wert $\beta = 0,01$ bzw. $\beta = 1\%$) besagt (verschieden formuliert):

- Der Annahmebereich ist so (sehr klein) gewählt worden, dass die vermutete Qualitätsverbesserung (bei falscher Nullhypothese H_0) mit einer Wahrscheinlichkeit von $\beta = 0,01$ bzw. $\beta = 1\%$ (irrtümlich) nicht abgelehnt wird.
- Beim Registrieren von weniger als zwei defekten Federsätzen (unter 20 untersuchten Federsätzen) spricht man sich mit der Wahrscheinlichkeit von $\beta = 0,01$ bzw. $\beta = 1\%$ nicht gegen die (in Wirklichkeit nicht erreichte) Qualitätsverbesserung aus.
- In 1 % (nur jeder Hundertsten!) aller Stichproben (bei einer Wiederholung des Zufallsexperiments „Ziehen einer Stichprobe“ in genügend großer Anzahl) würden die Stichprobenergebnisse nicht gegen die (in Wirklichkeit nicht erreichte) Qualitätsverbesserung sprechen.

Vom „gesunden Menschenverstand“ her erscheint der Wert für das Signifikanzniveau α unnötigerweise sehr groß. Gleichzeitig ist der β -Fehler (Fehler 2. Art) unnötigerweise sehr klein. Damit würde der Zulieferbetrieb ein wohl doch übertrieben geringes Image-Risiko (bei höherem Kostenrisiko) eingehen. Der Statistiker wird versuchen, „realistischere“ Entscheidungsregeln zu formulieren, also die Testkonstruktion entsprechend zu verändern.

Soll die Wahrscheinlichkeit für den Fehler 1. Art verkleinert werden, so ist dies möglich, indem man den Ablehnungsbereich für die Nullhypothese H_0 verkleinert. Dies bedeutet die gleichzeitige Vergrößerung des Annahmebereiches und somit der Wahrscheinlichkeit des Fehlers 2. Art (vgl. Fig. J 2a und J 2b). Wählt man beispielsweise $A = \{0; 1; \dots; 5\}$ und somit $\bar{A} = \{6; 7; \dots; 20\}$, so kann H_0 nicht abgelehnt werden, da $2 \notin \bar{A}$.

Berechnung des Fehlers 1. Art:

$$\alpha = P(\bar{A}_{p_0}) = B_{20; 0,1}(\bar{A}) = B_{20; 0,1}(\{6; 7; \dots; 20\}) = 1 - B_{20; 0,1}(\{0; 1; \dots; 5\}) = 1 - 0,98875 = 0,01125 \approx 0,01 \text{ (mit Tabelle), also } \alpha \approx 0,01 \text{ bzw. } \alpha \approx 1\%$$

Berechnung des Fehlers 2. Art:

$$\beta = P(A_{p_1}) = B_{20; 0,3}(A) = B_{20; 0,3}(\{0; 1; \dots; 5\}) = 0,41637 \approx 0,42 \text{ (mit Tabelle), also } \beta \approx 0,42 \text{ bzw. } \beta \approx 42\% ^1)$$

Interpretation:

- Für den Fall, dass die Qualitätsverbesserung (tatsächlich) erreicht worden ist, spricht man sich (nur noch) mit der Wahrscheinlichkeit $\alpha = 0,01$ bzw. $\alpha = 1\%$ gegen die Qualitätsverbesserung aus. (Das Signifikanzniveau beträgt $\alpha = 0,01$ bzw. $\alpha = 1\%$.)
- Für den Fall, dass die Qualitätsverbesserung (tatsächlich) nicht erreicht worden ist, spricht man sich (immerhin noch) mit der Wahrscheinlichkeit von $\beta = 0,42$ bzw. $\beta = 42\%$ nicht gegen die (in Wirklichkeit nicht erreichte) Qualitätsverbesserung aus. (Der Fehler 2. Art beträgt $\beta = 0,42$ bzw. $\beta = 42\%$.)

Ob die beiden möglichen Fehlentscheidungen mit diesen Wahrscheinlichkeiten akzeptabel sind, hat allein der Zulieferbetrieb mit Blick auf sein Image und die bei Fehlentscheidungen entstehenden Kosten zu beurteilen. Der (vermuteten) Qualitätsverbesserung auf einem Signifikanzniveau von $\alpha = 1\%$ zuzustimmen (wenn $\alpha = 1\%$, dann $2 \notin \bar{A}$), dürfte eine durchaus akzeptable Entscheidung sein, bedeutet aber auch – im Falle keiner Verbesserung – mit einer Wahrscheinlichkeit von etwa 42 % (Fehler 2. Art) zu irren.

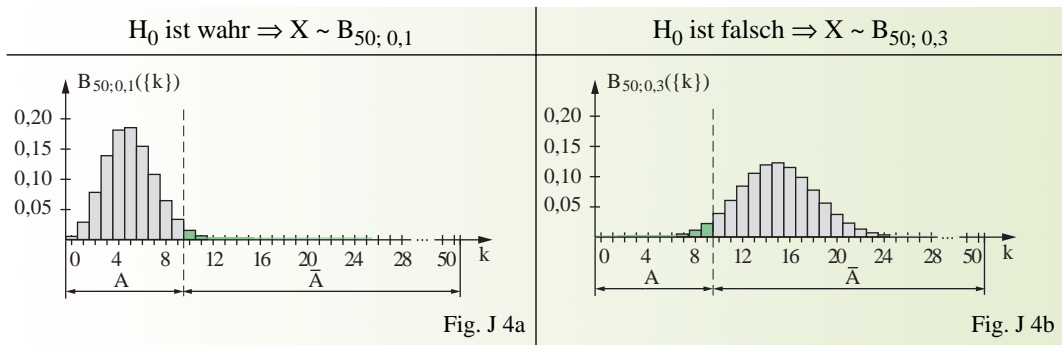
Zusammenfassend lässt sich bei festem Stichprobenumfang n feststellen:

- Je kleiner man den Ablehnungsbereich \bar{A} wählt, desto kleiner wird auch die Wahrscheinlichkeit für den Fehler 1. Art.
- Je kleiner man den Annahmebereich A wählt, desto kleiner wird die Wahrscheinlichkeit für den Fehler 2. Art.

¹⁾ Man erkennt: Bereits kleine Änderungen des Ablehnungsbereiches (Annahmebereiches) können große Änderungen der Wahrscheinlichkeiten des Fehlers 1. Art und des Fehlers 2. Art bewirken.

- Bei festen Werten für p_0 (Nullhypothese) und p_1 (Alternativhypothese) bewirkt jede Verkleinerung der Wahrscheinlichkeit α eine Vergrößerung der Wahrscheinlichkeit β .

In Abhängigkeit vom konkreten Sachverhalt ist abzuwägen, für welchen Fehler die Wahrscheinlichkeit möglichst klein bleiben soll. Müssen möglichst beide Wahrscheinlichkeiten für Fehlentscheidungen klein bleiben, dann ist dies nur mit einer Vergrößerung des Stichprobenumfangs erreichbar. Bezogen auf den in Beispiel J 1 untersuchten Sachverhalt könnte z.B. der Stichprobenumfang von $n = 20$ auf $n = 50$ vergrößert werden. Nehmen wir an, es werden nun in einer Stichprobe (mit $n = 50$) fünf defekte Federsätze festgestellt.¹⁾ Als Ablehnungsbereich wird z.B. $\bar{A} = \{10; 11; \dots; 50\}$ und folglich als Annahmebereich $A = \{0; 1; \dots; 9\}$ gewählt (Fig. J 4a und J 4b). Damit sind höhere Ansprüche gesetzt: Im Vergleich zu dem vorher bei $n = 20$ gewählten Annahmebereich ($A = \{0; 1; \dots; 5\}$) hat sich der Stichprobenumfang auf $n = 50$ mehr als verdoppelt, während der „kritische Wert“ für X von $X = 5$ auf $X = 10$ für den neuen Annahmebereich nur verdoppelt worden ist.



Berechnung des Fehlers 1. Art:

$$\alpha = P(\bar{A}_{p_0}) = B_{50;0,1}(\bar{A}) = B_{50;0,1}(\{10; 11; \dots; 50\}) = 1 - B_{50;0,1}(\{0; 1; \dots; 9\}) = 1 - 0,97546 = 0,02454 \approx 0,02 \text{ (mit Tabelle), also } \alpha \approx 0,02 \text{ bzw. } \alpha \approx 2 \%$$

Berechnung des Fehlers 2. Art:

$$\beta = P(A_{p_1}) = B_{50;0,3}(A) = B_{50;0,3}(\{0; 1; \dots; 9\}) = 0,04023 \approx 0,04 \text{ (mit Tabelle), also } \beta \approx 0,04 \text{ bzw. } \beta \approx 4 \%$$

Damit lässt sich bei *veränderlichem* Stichprobenumfang verallgemeinern:

Vergrößert man den Stichprobenumfang n , so wird die Summe der Wahrscheinlichkeiten für die Fehler 1. und 2. Art verkleinert. Die Sicherheit für die zu treffende Entscheidung wächst.

In der Praxis gilt es jedoch abzuwägen zwischen einer gewissen Mindestsicherheit und dem dafür notwendigen Stichprobenumfang, der aus Zeit- und Kostengründen nicht unnötig groß sein sollte. Wir nehmen an, das Management des Zulieferbetriebes aus Beispiel J 1 habe den Stichprobenumfang $n = 20$ und das Signifikanzniveau $\alpha = 0,05$ ($\alpha = 5 \%$) vorgegeben (s. Möglichkeit (2), S. 447). Es sollen nun der zugehörige Annahme- bzw. Ablehnungsbereich für die Nullhypothese sowie der Fehler 2. Art ermittelt werden. Dazu überlegen wir:

In einer Stichprobe mit dem Umfang n können genau keine (null) oder eine oder zwei oder ... (bis hin zum Extremfall) n schwarze Kugeln (defekte Federsätze) enthalten sein. Für die Zufallsgröße X gilt also $X = \{0; 1; 2; \dots; k-1; k; k+1; \dots; n-1; n\}$. Der Wert k stehe dabei für die kritische Anzahl X , bei der wir in Annahme- und Ablehnungsbereich trennen. Beim einseitigen, *rechtsseitigen* Test

¹⁾ Der Wert $p_0 = 0,1$ für die Nullhypothese soll unverändert bleiben. Der Erwartungswert für $n = 50$ ist dann $EX = n \cdot p = 50 \cdot 0,1 = 5$.

(große Werte von X sprechen gegen die Nullhypothese H_0 und somit für die Alternativhypothese H_1) ist der Ablehnungsbereich $\bar{A} = \{k; k + 1; \dots; n\}$. Beim (einseitigen) *linksseitigen* Test (kleine Werte von X sprechen gegen die Nullhypothese H_0 und somit für die Alternativhypothese H_1) wäre der Ablehnungsbereich $\bar{A} = \{0; 1; \dots; k\}$. Im Vergleich zu dem Vorgehen bei Möglichkeit (1) ist das Vorgehen nun genau umgekehrt.

J 3

Satz J 3: Ermitteln des kritischen Werts $X = k$ bei vorgegebenem Signifikanzniveau α (Einseitiger) *rechtsseitiger* Test:

Bei vorgegebenem α -Wert ist k als diejenige *kleinste* ganze Zahl zu ermitteln, für die gilt:

$$P(\bar{A}_{p_0}) = P(X \geq k) = B_{n; p_0}(\{k; k + 1; \dots; n\}) = 1 - B_{n; p_0}(\{0; 1; \dots; k - 1\}) \leq \alpha.$$

(Im Allgemeinen wird mit der Beziehung $B_{n; p_0}(\{0; 1; \dots; k - 1\}) \geq 1 - \alpha$ gearbeitet.)

(Einseitiger) *linksseitiger* Test:

Bei vorgegebenem α -Wert ist k als diejenige *größte* ganze Zahl zu ermitteln, für die gilt:

$$P(\bar{A}_{p_0}) = P(X \leq k) = B_{n; p_0}(\{0; 1; \dots; k\}) \leq \alpha.$$

Für den Sachverhalt aus Beispiel J 1 ergibt sich (rechtsseitiger Test): $B_{20; 0,1}(\{0; 1; \dots; k - 1\}) \geq 0,95$. Gemäß Tabelle der summierten Binomialverteilung [$B_{20; 0,1}(\{0; 1; \dots; 4\}) = 0,95683 \geq 0,95$] ist diese Ungleichung erstmals für den Wert $k - 1 = 4$, also $k = 5$ erfüllt. Beim Arbeiten mit einem geeigneten Rechner können auch mehrere vermutete k -Werte (hier 2; 3; 4; 5; 6) eingegeben und die angezeigten summierten Wahrscheinlichkeiten mit dem Wert 0,95 verglichen werden.

Als Ablehnungsbereich folgt damit $\bar{A} = \{5; 6; \dots; 20\}$ – der Annahmehereich ist $A = \{0; 1; \dots; 4\}$. Da $2 \notin \bar{A}$ (siehe Stichprobe im Beispiel J 1), darf man sich auf einem Signifikanzniveau von $\alpha = 0,05$ für die Qualitätsverbesserung (Nullhypothese) entscheiden. Der Fehler 2. Art beträgt dann $\beta = P(A_{p_1}) = B_{20; 0,3}(A) = B_{20; 0,3}(\{0; 1; \dots; 4\}) = 0,23751 \approx 0,24$ bzw. $\beta = 24\%$ (mit Tabelle).

Die im vorliegenden Abschnitt J 1.2 zu Alternativtests erarbeiteten Zusammenhänge sollen nun an weiteren Beispielen verdeutlicht werden.

J 2

Beispiel J 2:

Zwei Sorten Saatgetreide mit einer Keimfähigkeit von 90 % bzw. 70 % wurden in verschiedenen Containern angeliefert, die nicht klar bez. der Keimfähigkeit des jeweiligen Inhalts gekennzeichnet waren. Es ist zu prüfen, wie groß die Keimfähigkeit des Saatgetreides in einem bestimmten Container ist. Wir testen die Hypothesen:

$H_0: p = 0,9$ („Die Keimfähigkeit des Getreides im Container beträgt 90 %.“)

$H_1: p = 0,7$ („Die Keimfähigkeit des Getreides im Container beträgt 70 %.“)

Zur Überprüfung werden 50 „auf gut Glück“ ausgewählte Getreidekörner aus einem Container auf Keimfähigkeit untersucht. X sei die zufällige Anzahl der keimenden Körner (unter den 50 ausgewählten). X kann als binomialverteilt mit den Parametern $n = 50$ und p sowie dem Erwartungswert $EX = 50p$ angenommen werden (kleine Stichprobe aus großer Grundgesamtheit). Wir haben nun zwischen den zwei Alternativen H_0 und H_1 zu entscheiden.

Gilt $p = 0,9$, so ergibt sich $EX = 45$. Es ist daher sinnvoll, als Annahmehereich $A = \{45; \dots; 50\}$ und entsprechend als Ablehnungsbereich $\bar{A} = \{0; \dots; 44\}$ zu wählen. Es sind zwei Fehlentscheidungen möglich:

- Hypothese H_0 wird abgelehnt, obwohl sie in Wirklichkeit zutrifft (Fehler 1. Art).
- Hypothese H_0 wird nicht abgelehnt, obwohl sie in Wirklichkeit falsch ist (Fehler 2. Art).

- a) Als Wahrscheinlichkeit für einen Fehler 1. Art erhalten wir dann

$$B_{50; 0,9}(\bar{A}) = B_{50; 0,9}(\{0; \dots; 44\}) = 0,38388.$$

Wir können die Wahrscheinlichkeit für diesen Fehler 1. Art verringern, indem wir den Ablehnungsbereich verkleinern und damit den Annahmehereich vergrößern. Wählen wir z.B. als Ablehnungsbereich $\bar{A} = \{0; \dots; 40\}$ und somit als Annahmehereich $A = \{41; \dots; 50\}$. Daraus erhält man als Wahrscheinlichkeit für einen Fehler 1. Art:

$$B_{50; 0,9}(\bar{A}) = B_{50; 0,9}(\{0; \dots; 40\}) = 0,02454$$

Das heißt: Lehnen wir die Nullhypothese H_0 ab, wenn höchstens 40 Körner keimen, dann beträgt die Wahrscheinlichkeit etwa 0,02, dass wir die in Wirklichkeit wahre Hypothese ablehnen. Auf den vorliegenden Sachverhalt bezogen hieße das also: Bei dem angewandten Prüfverfahren müsste man damit rechnen, in etwa 2% der Fälle einen Container mit Saatgut der Keimfähigkeit 90% irrtümlich als einen Container mit Saatgut der Keimfähigkeit 70% einzustufen.

- b) Wir haben uns in a) für $\bar{A} = \{0; \dots; 40\}$ und $A = \{41; \dots; 50\}$ entschieden. Sind unter den 50 Getreidekörnern weniger als 41 keimfähig, dann spricht das gegen die Hypothese H_0 – also würden wir uns für H_1 entscheiden und H_0 ablehnen.

Als Wahrscheinlichkeit für einen Fehler 2. Art würde sich dann ergeben:

$$B_{50; 0,7}(A) = B_{50; 0,7}(\{41; \dots; 50\}) = 1 - B_{50; 0,7}(\{0; \dots; 40\}) = 1 - 0,95977 = 0,04023$$

Das heißt: Lehnen wir die Nullhypothese H_0 nicht ab, wenn mehr als 40 (mindestens 41) Körner keimen, dann beträgt die Wahrscheinlichkeit etwa 0,04, dass wir uns fälschlich für H_0 entscheiden. Auf den vorliegenden Sachverhalt bezogen hieße das also: Bei dem angewandten Prüfverfahren müsste man damit rechnen, in etwa 4% der Fälle einen Container mit Saatgut der Keimfähigkeit 70% irrtümlich als einen Container mit Saatgut der Keimfähigkeit 90% einzustufen.

Durch Probieren haben wir damit eine Entscheidungsregel gefunden, bei der sowohl die Wahrscheinlichkeit des Fehlers 1. Art als auch die des Fehlers 2. Art unter 0,05 liegt.

Wegen des Zusammenhangs der Wahrscheinlichkeiten des Fehlers 1. Art und des Fehlers 2. Art können die gewünschten bzw. akzeptablen Größenverhältnisse nur mit Bezug auf den konkreten Sachverhalt sinnvoll entschieden werden. Dies illustrieren nachfolgende Beispiele.

Beispiel J 3:

(1)

Ein Pilzsammler hat in einem Waldgebiet mehrere große Körbe Pilze gesammelt, die er zur Sicherheit von einem zuverlässigen Pilzkenner prüfen lässt. Der Pilzkenner weiß aus jahrelanger Erfahrung, dass 2 % aller von Sammlern vorgelegten Pilze irrtümlich gesammelte Giftpilze sind, wobei ihm selbst noch nie ein Prüffehler unterlaufen ist.

Lauten die Hypothesen „Der Pilz ist ungiftig“ (Nullhypothese) und „Der Pilz ist giftig“ (Alternativhypothese), so muss die Wahrscheinlichkeit des Fehlers 2. Art praktisch null sein. Ein größerer Fehler 1. Art schmälert „schlimmstenfalls“ den Umfang der Mahlzeit ...

Der Pilzkenner wird keine Stichprobe ziehen und etwa Berechnungen ausführen (lassen), sondern jeden einzelnen Pilz begutachten. Die vollständige Pilzsammlung ist sozusagen Stichprobe und Grundgesamtheit zugleich. Nur so können (beide möglichen) Fehlentscheidungen (theoretisch) sicher vermieden werden. Dieses Vorgehen mag zwar aufwändig werden, garantiert dafür aber maximale Sicherheit.

(2)

Ein Pharmakonzern lässt ein Präparat testen. Umfangreiche Beobachtungen zeigen, dass das Präparat in 5 % aller Anwendungsfälle zu (nachweisbar) schädlichen Nebenwirkungen führt. Auf der Basis umfangreicher Forschungstätigkeit weiterentwickelt, zeigt es bei erneuten Tests noch in 3 % aller Anwendungen schädliche Effekte.

Eine statistische Wertung (z. B. mittels Alternativtest) ist nur dann sinnvoll, wenn der (ggf. lebensrettende) Nutzen die schädlichen Nebenwirkungen aus medizinischer Sicht als weniger schwerwiegend in den Hintergrund treten lässt. Ist dies nicht der Fall, darf das Präparat nicht auf den Markt kommen. Bei einem Alternativtest ($H_0: p_0 = 0,05$; $H_1: p_1 = 0,03$) müssten dann sowohl der Fehler 1. Art (vorrangig Kostengründe) als auch insbesondere der Fehler 2. Art (vorrangig medizinische Gründe) möglichst klein gehalten werden.

(3)

Laut Angabe eines Pharmakonzerns führt sein hochwirksames, preisgünstiges Präparat P nur in 10 % aller Anwendungen zu Nebenwirkungen. Treten Nebenwirkungen auf, so gilt als sicher, dass diese völlig ungefährlich bleiben, wenn sie mit einem Zusatzpräparat Z behandelt werden. Von Fachärzten wird (z. B.) zur Diskussion gestellt, ob die Angabe 10 %, gemessen an den langjährigen Erfahrungen zu Nebenwirkungen, akzeptabel ist. Ein Arzt stellt fest, dass von 100 mit P behandelten Patienten 40 unter Nebenwirkungen gelitten haben und daher zusätzlich mit dem sehr teuren Präparat Z behandelt werden mussten.

Zur statistischen Wertung wird ein Alternativtest gewählt: $H_0: p_0 = 0,4$; $H_1: p_1 = 0,1$ ¹⁾

Die Zufallsgröße X beschreibe die Anzahl der Patienten mit Nebenwirkungen unter 100 (zufällig) ausgewählten. Da kleine Werte von X gegen die Nullhypothese H_0 (und somit für die Alternativhypothese H_1) sprechen, konstruiert man den Alternativtest als (einseitigen) linksseitigen Test. Für das Signifikanzniveau wird der i. Allg. übliche Wert $\alpha = 0,05$ gewählt.

Wegen $P(\bar{A}_{p_0}) = P(X \leq k) = B_{n; p_0}(\{0; 1; \dots; k\}) \leq \alpha$ erhält man $B_{100; 0,4}(\{0; 1; \dots; k\}) \leq 0,05$.

Gemäß Tabelle der summierten Binomialverteilung $[B_{100; 0,4}(\{0; 1; \dots; 31\}) = 0,03985 \leq 0,05]$ ist diese Ungleichung „letztmals“ für den Wert $k = 31$ erfüllt.

Als Ablehnungsbereich folgt damit $\bar{A} = \{0; 1; \dots; 31\}$ – der Annahmebereich ist $A = \{32; 33; \dots; 100\}$. Da $40 \notin \bar{A}$, kann H_0 (auf einem Signifikanzniveau von $\alpha = 0,05$) nicht abgelehnt werden. Die 10 %-Angabe darf berechtigt angezweifelt werden; die Feststellung des Arztes ($H_0: p_0 = 0,4$) ist demnach als nicht zufällig anzusehen. Für den Fehler 2. Art errechnet man

$\beta = P(A_{p_1}) = B_{100; 0,1}(A) = B_{100; 0,1}(\{32; 33; \dots; 100\}) = 1 - B_{100; 0,1}(\{0; 1; \dots; 31\}) = 1 - 1 = 0$.

Dieser Fehler 2. Art erhärtet, dass praktisch keine Zweifel an der Feststellung des Arztes bestehen.

J 1.3 Testen einer unbekannten Wahrscheinlichkeit; Signifikanztests

Im Abschnitt J 1.2 wurde angemerkt, dass die fehlende Spezifikation der Alternativ- bzw. Gegenhypothese bei statistischen Tests als der Normalfall anzusehen ist. Unter dieser Bedingung wird von einem normalen Signifikanztest gesprochen.

J 5

Definition J 5:

Ein statistischer Test auf signifikante Unterschiede (Signifikanztest), bei dem auf Stichprobenbasis über die Beibehaltung der (einfachen oder zusammengesetzten) Nullhypothese H_0 oder deren Ablehnung entschieden wird, heißt **normaler Signifikanztest**, kurz: **Signifikanztest**.

¹⁾ Auch hier wird (wie aus Sicherheitsgründen zumeist üblich) die Alternative zur eigentlichen Vermutung als Nullhypothese gewählt.

Von einem Signifikanztest spricht man also, wenn im Prinzip *nur eine Hypothese*, die Nullhypothese H_0 untersucht wird. Die Nullhypothese ist zwischen einer (einfachen oder zusammengesetzten) Hypothese und deren Negation zu wählen.

Satz J 4: Nullhypothese bei einem Signifikanztest

Bei einem Signifikanztest wählt man diejenige Hypothese als Nullhypothese, bei der der Fehler 1. Art – *in Abhängigkeit vom konkreten Sachverhalt* – von größerer Bedeutung ist als der (i. Allg. nicht eindeutig zu berechnende) Fehler 2. Art.

J 4

Das Formulieren einer „*zahlenmäßig konkreten*“ Alternativ- bzw. Gegenhypothese ist (z. B. aufgrund fehlender Angaben oder Erfahrungen bzw. aus rein mathematischen Gründen) i. Allg. nicht möglich. Für die Wahrscheinlichkeit des Fehlers 1. Art wird eine möglichst kleine Zahl ($0 < \alpha < 1$) als Höchstwert (obere Schranke) vorgegeben oder gewählt. Diese Zahl heißt **Signifikanzniveau** α oder auch **Irrtumswahrscheinlichkeit** α . Bei praktischen Anwendungen setzt man zumeist $\alpha = 0,05$ oder $\alpha = 0,01$. Soll anhand einer Stichprobe vom Umfang n ¹⁾ entschieden werden, ob die Nullhypothese abgelehnt werden kann oder nicht, und erfolgt die Entscheidung auf dem Signifikanzniveau $\alpha = 0,05$, so spricht man von einem **signifikanten Ergebnis** (Unterschied), und bei einem Signifikanzniveau $\alpha = 0,01$ von einem **hochsignifikanten Ergebnis** (Unterschied).

Beispiel J 4:

Der Altstadtbereich einer Kleinstadt ist als verkehrsberuhigte Zone (zulässige Höchstgeschwindigkeit von Fahrzeugen aller Art: $30 \frac{\text{km}}{\text{h}}$) umgestaltet worden. Beobachtungen in den ersten Monaten nach der Umgestaltung belegen, dass 10 % aller Pkws die zulässige Höchstgeschwindigkeit überschreiten. Nach zwei Jahren haben die Anwohner den Eindruck, dass in der „Zone 30“ immer noch relativ oft zu schnell gefahren wird. Sie möchten daher wissen, ob der alte Erfahrungswert noch zutreffend ist.

J 4

Überlegungen zur Klärung des Problems aus Beispiel J 4:

Als Nullhypothese wird $H_0: p_0 = 10\%$ festgelegt. Da keine weiteren *gesicherten* Informationen (Erfahrungen, Anhaltspunkte) vorliegen, ist eine Spezifikation der Negation (hier als Alternativ- bzw. Gegenhypothese gewählt) nicht sinnvoll. Die Gegenhypothese lautet somit $H_1: p_1 \neq 0,1$. Hierdurch bringt man nur die Vermutung zum Ausdruck, dass der alte Erfahrungswert (10 %) nicht mehr zutreffend ist. Insbesondere (aus Gründen der Vorsicht) wird zunächst keine weitere Vermutung etwa zur Erhöhung oder Senkung des Anteils der Geschwindigkeitsüberschreiter berücksichtigt. Als Stichprobe wählt man insgesamt 100 Beobachtungen mit Geschwindigkeitsmessung (über einen längeren Zeitraum; Klumpenstichprobe, vgl. J 1.1). Beschreibt die Zufallsgröße X die zufällige Anzahl der „Überschreiter“ (in der Stichprobe, $n = 100$), so sprechen sowohl (sehr) große als auch (sehr) kleine Werte von X gegen die Nullhypothese. Es sind also Abweichungen *zweiseitig* – nach links und nach rechts – von Interesse. Der Signifikanztest ist daher als **zweiseitiger Signifikanztest** zu konstruieren. Als Signifikanzniveau α wählen wir $\alpha = 0,05$. Für den Test heißt dies, dass wir höchstens mit einer Wahrscheinlichkeit von $\alpha = 0,05$ bzw. $\alpha = 5\%$ eine wahre Nullhypothese irrtümlich als falsch ablehnen wollen. Anders formuliert: Der Signifikanztest soll – kann die Nullhypothese H_0 abgelehnt werden – eine statistische Sicherheit von $1 - \alpha = 0,95$ bzw. 95 % besitzen.

¹⁾ Die Zufallsgröße X kennzeichnet dann die absolute Häufigkeit, mit der ein interessierendes Merkmal bei n Beobachtungen (Stichprobenumfang) in der Stichprobe auftritt.

J 6

Definition J 6:

Kann bei einem Signifikanztest die Nullhypothese H_0 auf dem Signifikanzniveau α abgelehnt werden, so bezeichnet man die Wahrscheinlichkeit $1 - \alpha$ als **statistische (Mindest-)Sicherheit des Signifikanztests**.

Bei einem zweiseitigen Signifikanztest wird $\bar{A} = \{0; 1; \dots; k_L\} \cup \{k_R; k_R + 1; \dots; n\}$ der **zweiseitige Ablehnungsbereich**. Er setzt sich aus der Vereinigung zweier Mengen (einer „linken“ und einer „rechten“ Teilmenge) zusammen. Man bezeichnet k_L als die linke und k_R als die rechte **Signifikanzgrenze** im Ablehnungsbereich. Der Wert k_L ist der größte, der Wert k_R der kleinste X-Wert im jeweiligen Teilbereich des Ablehnungsbereiches. Die Werte k_L und k_R sind durch das Signifikanzniveau α festgelegt. Ihrer Berechnung liegt folgende Überlegung zugrunde: Da die Gegenhypothese H_1 nicht weiter spezifiziert worden ist (also lediglich „verschieden von“ ausdrückt), muss sowohl für den linken Bereich als auch für den rechten Bereich das Signifikanzniveau gleichwertig eingehen. Diese Gleichwertigkeit erfordert die jeweilige Zuordnung von $\frac{\alpha}{2}$, also das Halbieren von α .¹⁾

J 5

Satz J 5: **Signifikanzgrenze eines zweiseitigen Signifikanztests**

Bei einem zweiseitigen Signifikanztest ist der vorgegebene α -Wert zu halbieren.

Der „linke“ Wert k_L ist als diejenige größte ganze Zahl zu ermitteln, für die gilt:

$$P(\bar{A}_{p_0}) = P(X \leq k_L) = B_{n; p_0}(\{0; 1; \dots; k_L\}) \leq \frac{\alpha}{2}.$$

Der „rechte“ Wert k_R ist als diejenige kleinste ganze Zahl zu ermitteln, für die gilt:

$$P(\bar{A}_{p_0}) = P(X \geq k_R) = B_{n; p_0}(\{k_R; k_R + 1; \dots; n\}) = 1 - B_{n; p_0}(\{0; 1; \dots; k_R - 1\}) \leq \frac{\alpha}{2}.$$

(Im Allgemeinen wird mit der Beziehung $B_{n; p_0}(\{0; 1; \dots; k_R - 1\}) \geq 1 - \frac{\alpha}{2}$ gearbeitet.)

Nun kann der vorgesehene zweiseitige Signifikanztest konstruiert werden.

Nullhypothese $H_0: p_0 = 0,1$; [Gegenhypothese $H_1: p_1 \neq 0,1$]; $n = 100$; $\alpha = 0,05$

X: Anzahl der „Überschreiter“ (bei 100 Beobachtungen); $X \sim B_{100; 0,1}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches \bar{A} :

Linker Teilbereich: $B_{100; 0,1}(\{0; 1; \dots; k_L\}) \leq 0,025$

Gemäß *Tabelle* der summierten (kumulierten) Binomialverteilung [$B_{100; 0,1}(\{0; 1; \dots; 4\}) = 0,02371 \leq 0,025$] ist diese Ungleichung letztmals für den Wert $k_L = 4$ erfüllt.

Rechter Teilbereich: $B_{100; 0,1}(\{0; 1; \dots; k_R - 1\}) \geq 0,975$

Gemäß *Tabelle* der summierten (kumulierten) Binomialverteilung [$B_{100; 0,1}(\{0; 1; \dots; 16\}) = 0,97940 \geq 0,975$] ist diese Ungleichung erstmals für den Wert $k_R - 1 = 16$, also $k_R = 17$ erfüllt.²⁾

Als Ablehnungsbereich folgt damit $\bar{A} = \{0; 1; \dots; 4\} \cup \{17; 18; \dots; 100\}$.

Interpretation:

Wird das Zufallsexperiment 100 Mal durchgeführt (100 Beobachtungen) und werden dabei weniger als fünf oder mehr als 16 Geschwindigkeitsüberschreiter registriert, dann kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von (höchstens) $\alpha = 0,05$ abgelehnt werden. Das heißt, die statistische Sicherheit der (signifikanten) Entscheidung (des Signifikanztests) beträgt (mindestens)

¹⁾ Dies bedeutet aber keineswegs, dass die beiden Teilmengen des Ablehnungsbereiches gleichmächtig sein müssen.

²⁾ Zur Ermittlung der Werte $B_{n; p}(\{k\})$ kann die Tabelle der summierten (kumulierten) Binomialverteilung oder ein geeigneter Rechner verwendet werden (vgl. Abschnitt J 1.2).

$1 - \alpha = 0,95$ bzw. 95 %. Oder: Mit einer (statistischen) Sicherheit von 95 % kann angenommen werden, dass sich der Anteil der „Überschreiter“ gegenüber dem Erfahrungswert 10 % verändert hat.

Gibt es „relativ sichere“ Anzeichen dafür, dass sich der Anteil der Geschwindigkeitsüberschreiter verringert haben dürfte (wenn z.B. ein „Geschwindigkeits-Blitzgerät“ fest installiert worden ist), dann kann auch ein **einseitiger Signifikanztest** gerechtfertigt sein. Beim einseitigen Signifikanztest ist der α -Wert des Signifikanzniveaus nicht zu halbieren. Es wird – in Abhängigkeit vom konkreten Sachverhalt – (einseitig) **links** oder (einseitig) **rechts** getestet, und zwar folgendermaßen:

- Wenn (allein) *große* Werte der Zufallsgröße X gegen die Nullhypothese sprechen, führt man einen (einseitigen) **rechtsseitigen Signifikanztest** mit dem (rechtsseitigen) Ablehnungsbereich $\bar{A} = \{k; k + 1; \dots; n\}$ durch.
- Wenn (allein) *kleine* Werte der Zufallsgröße X gegen die Nullhypothese H_0 sprechen, führt man einen (einseitigen) **linksseitigen Signifikanztest** mit dem (linksseitigen) Ablehnungsbereich $\bar{A} = \{0; 1; \dots; k\}$ durch.

Der kritische Wert k ist jeweils gemäß Satz J 3 zu ermitteln. Für unser obiges Beispiel hieße das:

(1) Aus der Sicht „Der Anteil der Geschwindigkeitsüberschreiter hat sich verringert oder ist gleich geblieben“¹⁾ könnte als (einseitiger) **rechtsseitiger Signifikanztest** konstruiert werden:

Nullhypothese $H_0: p_0 \leq 0,1$; [Gegenhypothese $H_1: p_1 > 0,1$], $n = 100$; $\alpha = 0,05$

X : Anzahl der „Überschreiter“ (bei 100 Beobachtungen); $X \sim B_{100; 0,1}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches: $\bar{A} = \{k; k + 1; \dots; 100\}$

Die Ungleichung $B_{100; 0,1}(\{0; 1; \dots; k - 1\}) \geq 0,95$ ist nach der Tabelle der summierten (kumulierten) Binomialverteilung $[B_{100; 0,1}(\{0; 1; \dots; 15\}) = 0,96011 \geq 0,95]$ *erstmal*s für den Wert $k - 1 = 15$, also $k = 16$ erfüllt. Als Ablehnungsbereich folgt damit $\bar{A} = \{16; 17; \dots; 100\}$.

Interpretation:

Werden bei 100 Beobachtungen mindestens 16 Geschwindigkeitsüberschreiter registriert, dann kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von (höchstens) $\alpha = 0,05$ abgelehnt werden. Mit einer statistischen Sicherheit von (mindestens) $1 - \alpha = 0,95$ bzw. 95 % kann davon ausgegangen werden, dass sich der Anteil der „Überschreiter“ eher vergrößert hat (als dass er gleich geblieben ist).

(2) Gibt es „relativ sichere“ Anzeichen dafür, dass sich der Anteil der Geschwindigkeitsüberschreiter vergrößert haben dürfte (wenn z.B. ein fest installiertes „Geschwindigkeits-Blitzgerät“ entfernt worden ist), dann kann auch ein (einseitiger) **linksseitiger Signifikanztest** sinnvoll sein. Aus der Sicht „Der Anteil der Geschwindigkeitsüberschreiter hat sich vergrößert oder ist gleich geblieben“ könnte konstruiert werden:

Nullhypothese $H_0: p_0 \geq 0,1$; [Gegenhypothese $H_1: p_1 < 0,1$; $n = 100$; $\alpha = 0,05$]

X : Anzahl der „Überschreiter“ (bei 100 Beobachtungen); $X \sim B_{100; 0,1}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches $\bar{A} = \{0; 1; \dots; k\}$:

Die Ungleichung $B_{100; 0,1}(\{0; 1; \dots; k\}) \leq 0,05$ ist nach der Tabelle der summierten (kumulierten) Binomialverteilung $[B_{100; 0,1}(\{0; 1; \dots; 4\}) = 0,02371 \leq 0,05]$ *letztmal*s für den Wert $k = 4$ erfüllt. Als Ablehnungsbereich folgt damit $\bar{A} = \{0; 1; \dots; 4\}$.

Interpretation:

Werden nicht mehr als (höchstens) vier Geschwindigkeitsüberschreiter (bei 100 zufälligen Beobachtungen) registriert, dann kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von (höchstens) $\alpha = 0,05$ abgelehnt werden. Mit einer statistischen Sicherheit von (mindestens)

¹⁾ Man beachte, wie vorsichtig diese Vermutung formuliert ist.

$1 - \alpha = 0,95$ bzw. 95 % kann davon ausgegangen werden, dass sich der Anteil der „Überschreiter“ eher verringert hat (als dass er gleich geblieben ist).

(3) Unter Umständen liegen bereits „relativ sichere“ Informationen – z. B. aus Beobachtungen im Rahmen eines Jahres-Projektes eines Leistungskurses Mathematik – vor, die eine *konkrete* Vermutung gestatten. Kann beispielsweise ein Anteil von 2 % Überschreitern als *neuer* Erfahrungswert vermutet werden, dann ist auch ein **Alternativtest** (mit $H_0: p_0 = 0,2$ und $H_1: p_1 = 0,3$) **als besonderer Signifikanztest** (Abschnitt J 1.2) möglich.

Anmerkung:

Bei einseitigen Signifikanztests werden zumeist beide Hypothesen auch als einseitige Hypothesen (beispielsweise $H_0: p_0 \geq p$; $H_1: p_1 \leq p$) geschrieben, obwohl von der zu untersuchenden Nullhypothese nur der „Grenzwert“ $p_0 = p$ rechnerisch einfließt. Dies ist aber gerade das (prüf-)statistische Problem, denn der wahre Wert von p_0 ist ja jeweils unbekannt. Für alle theoretisch möglichen Werte von p gilt nun aber sowohl bei rechtsseitigen ($p \leq p_0$) als auch bei linksseitigen ($p \geq p_0$) Signifikanztests: $B_{n; p}(\bar{A}) \leq B_{n; p_0}(\bar{A})$. Somit ist gesichert, dass eine Ermittlung des kritischen Wertes $X = k$ anhand von $B_{n; p_0}(\bar{A})$ alle theoretisch möglichen Werte des Parameters p berücksichtigt. Diese Zusammenhänge werden im Abschnitt J 1.4 genauer untersucht.

J 5

Beispiel J 5:

Ein Würfel mit den Augenzahlen „1“ bis „6“ brachte einer Schülerin beim „Mensch-ärgere-dich-nicht“-Spiel mehrmals nacheinander den Sieg, weil sie die Augenzahl „6“ relativ oft würfelte. Es interessiert, ob der benutzte Würfel bezüglich dieser Augenzahl wirklich ideal/regulär ist.

Zur Klärung dieser Frage kann man folgendermaßen vorgehen:

Nullhypothese $H_0: p_0 = \frac{1}{6}$ [Gegenhypothese $H_1: p_1 \neq \frac{1}{6}$]

Als Stichprobe wird 25-mal gewürfelt. Die Zufallsgröße X beschreibe dabei die zufällige Anzahl der Sechsen; $X \sim B_{25; 1/6}$ (bei wahrer H_0). Das Signifikanzniveau (für die Ablehnung von H_0) wird mit $\alpha = 0,05$ festgelegt. Wegen $\alpha = 0,05$ als Höchstwert für die Wahrscheinlichkeit des Fehlers 1. Art ist $B_{25; 1/6} \leq 0,05$ zu setzen.

Wir lehnen H_0 ab, wenn die Anzahl der sich bei 25 Würfeln ergebenden Augenzahl „6“ „zu groß“ oder „zu klein“ ist, d. h. $\bar{A} = \{0; 1; \dots; k_L\} \cup \{k_R; k_R + 1; \dots; 25\}$. (Man erkennt: Die Wahrscheinlichkeit für den Fehler 2. Art kann nicht eindeutig ermittelt werden, weil für $B_{25; p}(\{k_L + 1; \dots; k_R - 1\})$ wegen $p_1 \neq \frac{1}{6}$ kein Wert $p = p_1$ eindeutig bestimmbar ist.) Der Test ist also als **zweiseitiger Signifikanztest** zu führen. Das Signifikanzniveau α ist zu halbieren. Wir erhalten die zwei Ungleichungen $B_{25; 1/6}(\{0; 1; \dots; k_L\}) \leq 0,025$ und $B_{25; 1/6}(\{0; 1; \dots; k_R - 1\}) \geq 0,975$. Aus der *Tabelle* der summierten **Binomialverteilung** (oder mithilfe eines geeigneten Rechners) ermittelt man $k_L = 0$ sowie $k_R - 1 = 8$, also $k_R = 9$ und somit den Ablehnungsbereich $\bar{A} = \{0\} \cup \{9; 10; \dots; 25\}$.

Interpretation:

Führen wir das Zufallsexperiment 25-mal durch und erhalten eine Anzahl der Augenzahl „6“, die im zweiseitigen Ablehnungsbereich \bar{A} liegt, dann lehnen wir die Nullhypothese, dass der Würfel bezüglich der „6“ regulär ist, mit einer Irrtumswahrscheinlichkeit von höchstens 0,05 ab. Mit dem vorliegenden Würfel wurde 25-mal gewürfelt. Trat die Augenzahl „6“ dabei z. B. 10-mal auf, so heißt dies (wegen $10 \in \bar{A}$), dass der Würfel nicht als ideal angesehen werden kann.

Nach diesem Beispiel für einen zweiseitigen Signifikanztest sei ein weiteres (typisches) Beispiel für einen einseitigen Signifikanztest betrachtet.

Beispiel J 6:

Vor Wahlen werden Prognosen über die Chancen der einzelnen politischen Parteien aufgestellt. Um solche Prognosen zu erhärten, wählt man eine bestimmte Anzahl stimmberechtigter Personen nach einem Zufallsprinzip aus und befragt sie unabhängig voneinander (i. Allg. proportional geschichtete Stichprobe; vgl. Abschnitt J 1.1). Die Partei P hatte bei der letzten Wahl einen Stimmenanteil von 40 %. Vor der neuen Wahl möchte diese Partei nun wissen, ob sich ihr Wahlergebnis verändern (verbessern bzw. verschlechtern) wird. Dazu befragt man 100 Personen und prüft die Hypothese „Der Stimmenanteil hat sich verringert oder ist gleich geblieben“. Die Wahrscheinlichkeit (Signifikanzniveau α), dass man irrtümlich auf eine Erhöhung des Stimmenanteils schließt, soll nicht größer als $\alpha = 0,05$ sein.

Variante (1)

X sei die zufällige Anzahl der Personen, welche die Partei P wählen wollen; $X \sim B_{100; p}$.

Nullhypothese: $p_0 \leq 0,4$; [Gegenhypothese: $p_1 > 0,4$]

Die Nullhypothese ist abzulehnen, wenn der Anteil der Personen, welche die Partei P wählen wollen, „groß“ ist. Der Ablehnungsbereich \bar{A} wäre dann „rechtsseitig“ mit $\bar{A} = \{k; k + 1; \dots; 100\}$, der Test als (einseitiger) **rechtsseitiger Signifikanztest** durchzuführen. Für $B_{100; 0,4}(\{k; k + 1; \dots; 100\}) \leq 0,05$ entnimmt man der Tabelle der summierten Binomialverteilung [$B_{100; 0,4}(\{0; 1; \dots; k - 1\}) = B_{100; 0,4}(\{0; 1; \dots; 48\}) = 0,95770 \geq 0,95$] den Wert $k - 1 = 48$, also $k = 49$, woraus $\bar{A} = \{49; 50; \dots; 100\}$ folgt.

Interpretation:

Wollen mindestens 49 befragte Personen die Partei P wählen, dann lehnen wir die Nullhypothese H_0 mit $p \leq 0,4$ auf dem Signifikanzniveau von 0,05 ab.

Variante (2):

X sei die zufällige Anzahl der Personen, welche die Partei P *nicht* wählen wollen; $X \sim B_{100; p}$.

Nullhypothese: $p_0 \leq 0,6$; [Gegenhypothese: $p_1 > 0,6$]

Die Nullhypothese ist abzulehnen, wenn der Anteil der Personen, welche die Partei P *nicht* wählen wollen, „klein“ ist. Der Ablehnungsbereich \bar{A} wäre dann „linksseitig“ mit $\bar{A} = \{0; 1; \dots; k\}$, der Test als (einseitiger) **linksseitiger Signifikanztest** durchzuführen. Für $B_{100; 0,6}(0; 1; \dots; k) \leq 0,05$ entnimmt man der Tabelle der summierten Binomialverteilung [$B_{100; 0,6}(0; 1; \dots; 51) = 0,04230 \leq 0,05$] den Wert $k = 51$, woraus $\bar{A} = \{51; 52; \dots; 100\}$ folgt.

Interpretation:

Wollen mindestens 51 befragte Personen die Partei P nicht wählen, dann lehnen wir die Nullhypothese H_0 mit $p \geq 0,6$ auf dem Signifikanzniveau von 0,05 ab.

Beim Vergleich von Variante (1) und Variante (2) ist festzustellen, dass wir den Ablehnungsbereich für Variante (2) bereits aus dem Resultat von Variante (1) ablesen können: Bei einem Signifikanzniveau von $\alpha = 0,05$ entsprechen mindestens 49 Wähler der Partei P (1) höchstens 51 Nichtwählern (2) der Partei P.

Für größere Stichprobenumfänge n ist die Binomialverteilung in der Regel nicht tabelliert; Berechnungen ohne geeignete Rechner wären äußerst aufwändig. In solchen Fällen kann man unter bestimmten Bedingungen die *Binomialverteilung* (mit hinreichender Genauigkeit) *durch eine (Standard-)Normalverteilung* (s. Abschnitt H 5.8) *approximieren*. Selbst wenn ein geeigneter Rechner

verfügbar ist, zieht man i. Allg. die einfachere, praktisch ausreichend genaue Näherungsrechnung auf der Basis einer angenommenen Normalverteilung vor. Zur Entscheidung, ob eine Näherung zulässig ist oder nicht, werden empirisch gewonnene Kriterien herangezogen.

J 6

Satz J 6: Approximation einer Binomialverteilung durch eine Normalverteilung

Bei großen Stichprobenumfängen n kann i. Allg. eine Binomialverteilung dann durch eine (Standard-)Normalverteilung hinreichend genau approximiert werden, wenn die empirischen Kriterien $\sigma^2 = V(X) > 9$ bzw. $\sigma^2 = V(X) \geq \frac{1}{2} \sqrt{n}$ erfüllt sind.¹⁾

Um den Vergleich einer Berechnung mittels Binomialverteilung und ihrer Näherung durch Normalverteilung zu ermöglichen, wird für den Signifikanztest im nachstehenden Beispiel eine tabellierte Binomialverteilung ($B_{200; 0,5}$) gewählt.

J 7

Beispiel J 7:

Nun soll das einleitend angeführte prüfstatische Problem (s. S. 439) näher untersucht werden. Es wurde vermutet, dass die Geburt von 102 Jungen bei insgesamt 200 Geburten „im Rahmen“ der Geburtenstatistik liegt. Der Fehler 1. Art muss in diesem Fall für die („vorsichtige“) Hypothese „Die Anzahl der Jungengeburten ist höchstens gleich der der Mädchengeburten“ gegenüber dem Fehler 2. Art als bedeutsamer angesehen werden. Man entscheidet sich für das Signifikanzniveau $\alpha = 0,05$ und formuliert:

Nullhypothese $H_0: p_0 \leq 0,5$; [Gegenhypothese $H_1: p_1 > 0,5$]

Die Nullhypothese ist abzulehnen, wenn die Anzahl der Jungen in der Stichprobe zu „groß“ ist. Der Ablehnungsbereich \bar{A} wäre dann „rechtsseitig“ mit $\bar{A} = \{k; k + 1; \dots; 100\}$, der Test als (einseitiger) **rechtsseitiger Signifikanztest** durchzuführen.

Für $B_{200; 0,5}(\{k; k + 1; \dots; 100\}) \leq 0,05$ entnimmt man der Tabelle der summierten Binomialverteilung [$B_{100; 0,4}(\{0; 1; \dots; k - 1\}) = B_{200; 0,5}(\{0; 1; \dots; 112\}) = 0,96158 \geq 0,95$] den Wert $k - 1 = 112$, also $k = 113$, woraus $\bar{A} = \{113; 114; \dots; 200\}$ folgt.

Im vorliegenden Fall darf die Binomialverteilung durch Normalverteilung approximiert werden ($\sigma^2 = V(X) = n \cdot p \cdot (1 - p) = 200 \cdot 0,5 \cdot 0,5 = 50 > 9$ bzw. $\sigma^2 = V(X) = 50 \geq \frac{1}{2} \sqrt{200} \approx 7,07$).

Mit der globalen Näherungsformel von DE MOIVRE-LAPLACE (vgl. Satz H 31)

$P(X \leq k) = B_{n; p}(\{0; 1; \dots; k\}) \approx \Phi\left(\frac{k + 0,5 - \mu}{\sigma}\right)$ folgt bei „rechtsseitigem“ Ablehnungsbereich wegen $P(X \geq k) = B_{n; p}(\bar{A}) = B_{n; p}(\{k; k + 1; \dots; n\}) = 1 - B_{n; p}(\{0; 1; \dots; k - 1\}) \leq \alpha$ die Beziehung $P(X \geq k) \approx 1 - \Phi\left(\frac{k - 1 + 0,5 - \mu}{\sigma}\right) \leq \alpha$, also $\Phi\left(\frac{k - 1 + 0,5 - \mu}{\sigma}\right) \geq 1 - \alpha$.

Mit $\alpha = 0,05$, $\mu = EX = n \cdot p = 100$ und $\sigma = \sqrt{V(X)} = \sqrt{50}$ ergibt sich

$$\Phi\left(\frac{k - 1 + 0,5 - 100}{\sqrt{50}}\right) \geq 1 - 0,05.$$

In der Tabelle der Normalverteilung lesen wir ab: $\Phi(1,64) = 0,95$. Daraus folgt:

$$\frac{k - 1 + 0,5 - 100}{\sqrt{50}} \geq 1,64$$

Durch Umstellen nach k erhalten wir

$$k \geq 1,64 \cdot \sqrt{50} + 99,5 = 111,1 \quad \text{und damit} \quad \bar{A} = \{112; 113; \dots; 200\}.$$

¹⁾ Kriterien bezeichnet man i. Allg. dann als empirisch, wenn sie auf der Grundlage vielfältiger Erfahrungen als „Faustregeln“ formuliert worden sind und sich Entscheidungen auf ihrer Basis in der Praxis bewährt haben.

Interpretation:

Sowohl die Berechnungen nach der Binomialverteilung als auch die nach der (hinreichend genauen) Normalverteilung belegen also: Da $102 \notin \bar{A}$, kann die Nullhypothese (aufgrund der vorliegenden Daten auf dem Signifikanzniveau $\alpha = 0,05$) nicht abgelehnt werden. Das heißt: Die 200 Beobachtungen sprechen im Landkreis nicht gegen die Geburtenstatistik. Zur weiteren Untersuchung könnten umfangreichere Beobachtungen (Stichprobe über längeren Zeitraum, z.B. mit dem Umfang $n = 1000$) durchgeführt werden.

J 1.4 Zur Qualität statistischer Tests; Gütefunktion

Am Beispiel des Signifikanztests (für eine unbekannte Wahrscheinlichkeit bei vorgegebenem Signifikanzniveau α und festem Stichprobenumfang n) werden wir uns nun mit Überlegungen zur praktischen Brauchbarkeit, zur Qualität bzw. Güte statistischer Tests vertraut machen. Solche Überlegungen sind insbesondere bei (normalen) Signifikanztests bedeutsam, da hier i. Allg. ein in den Hypothesen festgelegter „Grenzwert“ der unbekannten Wahrscheinlichkeit p sozusagen stellvertretend für alle theoretisch möglichen Wahrscheinlichkeiten verwendet wird (vgl. Anmerkung S. 458).

Definition J 7:

Die Funktion G , die (bei vorgegebenem Ablehnungsbereich \bar{A} und bekanntem n) jedem Wert von p die Wahrscheinlichkeit für das Ablehnen der Nullhypothese H_0 zuordnet, heißt **Gütefunktion des Tests**: $G(p) = B_{n; p}(\bar{A})$

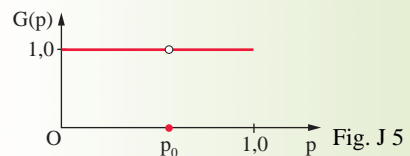
J 7

Überlegungen für einen zweiseitigen Signifikanztest mit $H_0: p = p_0$, $H_1: p \neq p_0$ und dem zweiseitigen Ablehnungsbereich $\bar{A} = \{0; 1; \dots; k_L\} \cup \{k_R; \dots; n\}$: Durch $G(p_0) = B_{n; p_0}(\bar{A})$ wird die Wahrscheinlichkeit für den Fehler 1. Art gegeben.

Die Beziehung $1 - G(p) = B_{n; p}(A)$ erfasst dann für $p \neq p_0$ alle theoretisch möglichen Wahrscheinlichkeit für den Fehler 2. Art.

Ideal wäre:

$$G(p) = \begin{cases} 0, & \text{falls } p = p_0 \\ 1, & \text{falls } p \neq p_0 \end{cases}$$



Man erkennt: $G(p)$ ist die Wahrscheinlichkeit einer Entscheidung *gegen* H_0 , unter der Bedingung, dass p der „wahre“ Wert ist.

Allgemein gilt bei vorgegebenem Signifikanzniveau α : $B_{n; p_0}(\bar{A}) \leq \alpha$. Gesucht wird nun $B_{n; p_0}(\bar{A})$ für $0 < p < 1$. Die notwendigen Überlegungen seien am Beispiel J 5 (aus Abschnitt J 1.3) illustriert:

$$\alpha = 0,05; n = 25; p_0 = \frac{1}{6}; \bar{A} = \{0\} \cup \{9; 10; \dots; 25\}$$

$$G(p) = B_{25; p}(\bar{A}) = 1 - B_{25; p}(\{1; 2; \dots; 8\}) = 1 + B_{25; p}(\{0\}) - B_{25; p}(\{0; 1; 2; \dots; 8\})$$

Unter Verwendung der Tabelle der summierten Binominalverteilung (oder eines geeigneten Rechners) erhält man folgende Wertetabelle

p	G(p)
0,05	0,27739
0,10	0,07225
$\frac{1}{6}$	0,02619
0,20	0,05055
0,25	0,15019
0,30	0,32321
$\frac{1}{3}$	0,46246
0,40	0,72647
0,50	0,94612
0,60	0,99567
$\frac{2}{3}$	0,99958
0,70	0,99990
0,75	0,99999
0,80	1,00000

(gerundet auf 5 Stellen
nach dem Komma)

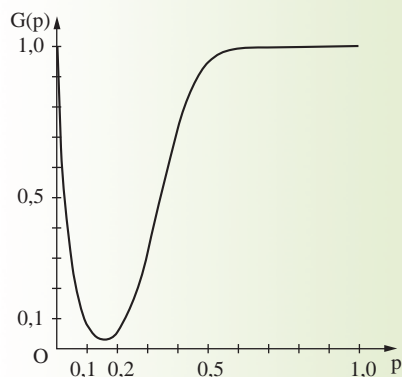


Fig. J 6

Verglichen mit dem Idealfall (Fig. J 5) ist aus Fig. J 6 zu erkennen:

Der Test ist umso besser, je steiler der Graph der Gütefunktion in der Nähe von p_0 verläuft.

Überlegungen für einen einseitigen Signifikanztest mit $H_0: p \leq p_0$; $H_1: p > p_0$

und dem rechtsseitigen Ablehnungsbereich $\bar{A} = \{k; k + 1; \dots; n\}$.

$G(p)$ Wahrscheinlichkeit für Fehler 1. Art, falls $0 \leq p \leq p_0$ (also H_0 wahr)

$1 - G(p)$ Wahrscheinlichkeit für Fehler 2. Art, falls $p_0 < p \leq 1$ (also H_0 falsch)

Ideal wäre:

$$G(p) = \begin{cases} 0, & \text{falls } p \leq p_0 \\ 1, & \text{falls } p > p_0 \end{cases}$$

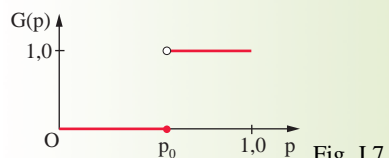


Fig. J 7

Für die weiteren Überlegungen greifen wir das Beispiel J 6 (aus Abschnitt J 1.3) auf:

$\alpha = 0,05$; $p_0 = 0,4$; $\bar{A} = \{49; 50; \dots; 100\}$

$G(p) = B_{100; p}(\{49; 50; \dots; 100\}) = 1 - B_{100; p}(\{0; 1; \dots; 48\})$

p	G(p)
0,25	0,00000
0,30	0,00005
$\frac{1}{3}$	0,00085
0,40	0,04230
0,50	0,61782
0,60	0,98999
$\frac{2}{3}$	0,99991
0,70	1,00000

(gerundet auf 5 Stellen
nach dem Komma)

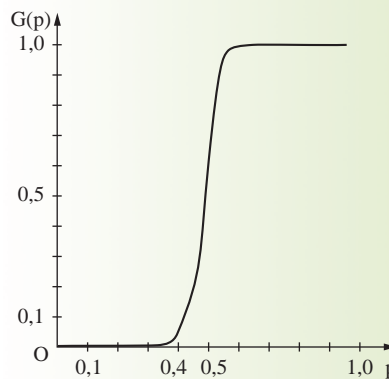


Fig. J 8

Als Verallgemeinerung lässt sich feststellen:

J 7

Satz J 7: Monotonie der Gütefunktion

Die Gütefunktion G ist für einen einseitigen Signifikanztest mit rechtsseitigem (linkseitigem) Ablehnungsbereich streng monoton wachsend (fallend).

Auf den Beweis dieses Satzes wird hier verzichtet. Aus Satz J 7 folgt:

Satz J 8:

Bei einseitigem Signifikanztest mit rechtsseitigem Ablehnungsbereich \bar{A} gilt bei vorgegebenem Signifikanzniveau α : $B_{n; p}(\bar{A}) \leq B_{n; p_0}(\bar{A}) \leq \alpha$ für $p \leq p_0$

Bei einseitigem Signifikanztest mit linksseitigem Ablehnungsbereich \bar{A} gilt bei vorgegebenem Signifikanzniveau α : $B_{n; p}(\bar{A}) \leq B_{n; p_0}(\bar{A}) \leq \alpha$ für $p \geq p_0$

J 8

Anmerkung:

Man kann anstelle der Gütefunktion auch die **Operationscharakteristik (OC-Funktion)** eines Tests betrachten: $OC(p) = B_{n; p}(A) = 1 - G(p)$ für $0 < p < 1$

Die OC-Funktion ordnet bei gegebenem Annahmebereich A jedem Wert von p die Wahrscheinlichkeit einer Entscheidung für die Nullhypothese H_0 zu.

J 2 Anwendungen aus verschiedenen Bereichen

Die in den vorangegangenen Abschnitten erarbeiteten grundlegenden Verfahren zum Beurteilen, Prüfen und Testen von Vermutungen bzw. Hypothesen haben einen Einblick in die Vielfalt prüfstatischer Anwendungen und Praxisbezüge vermittelt. Im vorliegenden Abschnitt soll nun die Möglichkeit gegeben werden, an einigen typischen, komplexeren Beispielen Zusammenhänge zu rekapitulieren sowie erworbene Verfahrenskenntnisse anzuwenden. Dabei ist stets zu beachten, dass die statistische **Hypothesenprüfung** zu keinen „Wahrheiten“ führt, sondern „lediglich“ zu Wahrscheinlichkeitsangaben darüber, wie gut oder schlecht das empirische Ergebnis (der Stichprobenuntersuchung) mit der Nullhypothese H_0 vereinbar ist. Die Entscheidung zu Gunsten der Gegen- bzw. Alternativhypothese H_1 wird also indirekt getroffen, indem man von den beiden konkurrierenden Hypothesen diejenige als falsch ablehnt, die als Erklärung für das empirische Ergebnis praktisch nicht in Frage kommt. Dabei kann ein Signifikanzniveau von 0,05 bzw. 0,01 als hinreichende Absicherung dagegen angesehen werden, dass willkürlich zufallsbedingte und spekulative Entscheidungen getroffen werden. Je kleiner das Signifikanzniveau α ist, desto schärfer ist der Test, aber desto seltener wird man H_0 ablehnen (verwerfen) können. Das entspricht einer Erfahrung des täglichen Lebens: Klare Urteile kann man nur selten fällen, verschwommene Aussagen (d.h. großes Signifikanzniveau α) sind hingegen leichter zu treffen. Man hat festgestellt, dass wir uns bei Alltagsentscheidungen je nach subjektiver Einschätzung mit einer Irrtumswahrscheinlichkeit bis zu 0,20 begnügen.

• Immer wieder Signifikanztests

In den nachfolgenden Beispielen werden wir einige **typische Entscheidungsfragen** untersuchen, zu deren prüfstatischer Absicherung Signifikanztests üblich sind.

Beispiel J 8:

Für eine beliebte Fernsehsendung eines Regionalsenders ist eine Einschaltquote von 60 % ermittelt worden. Aufgrund eines Moderatorenwechsels vermutet man, dass sich die Einschaltquote verändert haben könnte. Bei einer Befragung mittels TED¹⁾ äußern 96 von 200 zufällig ausgewählten Fernsehzuschauern, dass sie die Fernsehsendung regelmäßig sehen.

J 8

¹⁾ **T**eledialog; Name für Computeranlagen, die eine Registrierung und statistische Aufbereitung („Hochrechnung“) telefonischer Stimmabgaben ermöglichen.

Es ist zu untersuchen, ob aus dem Befragungsergebnis (auf dem Signifikanzniveau $\alpha = 0,01$) geschlussfolgert werden darf, dass sich die Einschaltquote verändert hat.

Lösung: (vgl. auch Beispiel J 4)

Da keine weiteren Informationen zur möglichen Veränderung der Einschaltquote vorliegen (höhere oder niedrigere Einschaltquote), wird ein **zweiseitiger Signifikanztest** (s. S. 455) konstruiert. (Sehr große und sehr kleine Werte der Zufallsgröße X sprechen gegen die Nullhypothese.)

Nullhypothese $H_0: p_0 = 0,6$; [Gegenhypothese $H_1: p_1 \neq 0,6$];

Stichprobenumfang: $n = 200$ Signifikanzniveau: $\alpha = 0,01$

X : Anzahl derjenigen Zuschauer, die die Fernsehsendung einschalten (unter 200 Befragten);

$X \sim B_{200; 0,6}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches \bar{A} :

Linker Teilbereich: $B_{200; 0,6}(\{0; 1; \dots; k_L\}) \leq 0,005$

Gemäß *Tabelle* der summierten Binomialverteilung $[B_{200; 0,6}(\{0; 1; \dots; 101\}) = 0,00405 \leq 0,005]$ ist diese Ungleichung „letztmals“ für den Wert $k_L = 101$ erfüllt.

Rechter Teilbereich: $B_{200; 0,6}(\{0; 1; \dots; k_R - 1\}) \geq 0,995$

Gemäß *Tabelle* der summierten Binomialverteilung $[B_{200; 0,6}(\{0; 1; \dots; 138\}) = 0,99662 \geq 0,995]$ ist diese Ungleichung „erstmal“ für den Wert $k_R - 1 = 138$, also $k_R = 139$ erfüllt.

Als Ablehnungsbereich \bar{A} folgt damit $\bar{A} = \{0; 1; \dots; 101\} \cup \{139; 140; \dots; 200\}$.

Interpretation / Schlussfolgerung:

Da $X = 96$ und $96 \in \bar{A}$, kann die Nullhypothese abgelehnt werden. Es darf aus dem Befragungsergebnis eine veränderte Einschaltquote geschlussfolgert werden. Die Veränderung ist hochsignifikant ($\alpha = 0,01$).

Die Wahrscheinlichkeit für den Fehler 2. Art (β -Fehler) ist nicht eindeutig berechenbar, da zur Gegenhypothese unendlich viele verschiedene Wahrscheinlichkeitswerte gehören. Wird (z.B. aufgrund des Befragungsergebnisses) angenommen, dass in Wirklichkeit für die Gegenhypothese H_1 der Wert $p_1 = 0,5$ gilt, lässt sich ein Fehler 2. Art berechnen:

$\beta = B_{200; 0,5}(\{102; 103; \dots; 138\})$;

$\beta = B_{200; 0,5}(\{0; 1; \dots; 138\}) - B_{200; 0,5}(\{0; 1; \dots; 101\}) = 1 - 0,58396 = 0,41604$

Mit dieser Wahrscheinlichkeit ($\beta \approx 0,42$; $\beta \approx 41,6\%$) würde man dann die in Wirklichkeit falsche Nullhypothese irrtümlich nicht ablehnen.

J 9

Beispiel J 9:

Die tägliche Auslastung eines Computer-Arbeitsplatzes einer Firma wird stichprobenartig überprüft. Der Auslastungsrichtwert betrage acht Stunden (täglich). Aus zurückliegenden Überprüfungen ist bekannt, dass der Arbeitsplatz an (höchstens) 2 % aller Arbeitstage weniger als acht Stunden in Betrieb ist. Die neue Stichprobe umfasst 90 Tage, wobei an drei Tagen der Auslastungsrichtwert nicht erreicht worden ist.

Es ist zu untersuchen, ob aus dem Stichprobenergebnis geschlussfolgert werden darf, dass sich die Auslastung verschlechtert hat.

Lösung: (vgl. auch die weiteren Betrachtungen zu Beispiel J 4)

Da eine Verschlechterung vermutet wird, konstruiert man einen (einseitigen) **rechtsseitigen Signifikanztest** (s. S. 457). (Beschreibt die Zufallsgröße X die Tage mit einer zu geringen Auslastung, so sprechen große Werte von X gegen die Nullhypothese.)

Nullhypothese $H_0: p_0 \leq 0,02$; [Gegenhypothese $H_1: p_1 > 0,02$];

Stichprobenumfang: $n = 90$; gewähltes Signifikanzniveau: $\alpha = 0,05$

X: Anzahl derjenigen Tage (von 90 Tagen), an denen der Auslastungsrichtwert unterschritten wird; $X \sim B_{90; 0,02}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches \bar{A} ; $\bar{A} = \{k; k + 1; \dots; 90\}$:

$B_{90; 0,02}(\{0; 1; \dots; k - 1\}) \geq 0,95$

Da i. Allg. für $n = 90$ und $p = 0,02$ keine Tabellierung der summierten Binomialverteilung vorliegt, ist – wie bereits im Abschnitt J 1.2; vgl. Fig. J 3 und Anmerkungen – ein geeigneter Rechner zu verwenden oder die BERNOULLI-Formel wiederholt anzuwenden.

Mit einem GTA ermittelt man

$B_{90; 0,02}(\{0; 1; \dots; 3\}) = 0,89325 \geq 0,95$. Die Ungleichung ist „erst-mals“ für den Wert $k - 1 = 3$, also $k = 4$ erfüllt (s. Fig. J 9).

Als Ablehnungsbereich \bar{A} folgt damit $\bar{A} = \{4; 5; \dots; 90\}$.

Interpretation/Schlussfolgerung:

Wegen $X = 3$ und $3 \notin \bar{A}$ kann die Nullhypothese nicht abgelehnt werden. Das Stichprobenergebnis muss als zufällig angesehen werden. Es ist weiterhin (auf dem Signifikanzniveau $\alpha = 0,05$) davon auszugehen, dass die Auslastung 2 % beträgt.

F1	F2	F3	F4	F5	F6
Algebra	Calc	Other	PrnIO	Clear	a-z
$\text{bin}(90, .02, 2)$.73117 $\text{bin}(90, .02, 3)$.89325 $\text{bin}(90, .02, 4)$.96519 $\text{bin}(90, .02, 4)$					
FUNC 3/20 RAD AUTO					

Fig. J 9

Beispiel J 10:

Ein Supermarkt einer Kleinstadt gibt in einem Werbematerial an, dass mindestens 75 % aller Kunden, die Waschmittel kaufen, die preisgünstige Marke „Tiefenrein“ auswählen. Von einer Mitarbeiterin der Verbraucherschutzzentrale werden 100 Kunden des Supermarktes, die Waschmittel gekauft haben, nach dem Zufallsprinzip befragt. 58 Kunden geben an, die Marke „Tiefenrein“ gekauft zu haben.

Es ist zu untersuchen, ob (auf dem Signifikanzniveau $\alpha = 0,05$) geschlussfolgert werden darf, dass die Angabe im Werbematerial übertrieben hoch ist.

Lösung: (vgl. auch die weiteren Betrachtungen zu Beispiel J 4)

Da man eine zu hohe Prozentangabe vermutet, wird ein (einseitiger) **linksseitiger Signifikanztest** (s. S. 457) konstruiert. (Beschreibt die Zufallsgröße X die Anzahl der Käufer des Waschmittels „Tiefenrein“, so sprechen kleine Werte von X gegen die Nullhypothese.)

Nullhypothese H_0 : $p_0 \geq 0,75$; [Gegenhypothese H_1 : $p_1 < 0,75$];

Stichprobenumfang: $n = 100$; Signifikanzniveau: $\alpha = 0,05$

X: Anzahl derjenigen Kunden (von 100 Waschmittelkäufern), die das Waschmittel „Tiefenrein“ kaufen;

$X \sim B_{100; 0,75}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches \bar{A} ; $\bar{A} = \{0; 1; \dots; k\}$:

$B_{100; 0,75}(\{0; 1; \dots; k\}) \leq 0,05$

Gemäß *Tabelle* der summierten Binomialverteilung [$B_{100; 0,75}(\{0; 1; \dots; 67\}) = 0,04460 \leq 0,05$] ist diese Ungleichung „letztmals“ für den Wert $k = 67$ erfüllt. Als Ablehnungsbereich \bar{A} folgt damit $\bar{A} = \{0; 1; \dots; 67\}$.

Interpretation/Schlussfolgerung:

Da $X = 58$ und $58 \in \bar{A}$, ist die Nullhypothese abzulehnen. Das Stichprobenergebnis ist nicht zufällig. Es weist einen signifikanten Unterschied ($\alpha = 0,05$) zur Aussage im Werbematerial auf. Diese Aussage darf somit berechtigt als übertrieben hoch angezweifelt werden.

• **Ein „Etikettierungsproblem“ – der „klassische“ Alternativtest hilft**

Produkte eines jeden Unternehmens werden insbesondere an ihrer Qualität (Preis-Leistungs-Verhältnis) gemessen. Gute Qualität sichert dem Unternehmen einen guten Ruf und i. Allg. eine gute Auftragslage.

Erzeugen Unternehmen Produkte zur Weiterverarbeitung („Zwischenprodukte“) in anderen Unternehmen ist die Einhaltung strenger Qualitätsvorgaben besonders wichtig. Zumeist werden die Zwischenprodukte in Qualitätskategorien (nach internen Kriterien) eingeteilt. Eine falsche oder fehlende Kennzeichnung der jeweiligen Kategorie kann dem Unternehmen neben hohen finanziellen Verlusten auch erhebliche Imageschäden zufügen.

J 11

Beispiel J 11:

Ein großes Unternehmen stellt Garn für die Fertigung in Teppichfirmen auf zwei verschiedenen Produktionsanlagen I und II – zu gleichen Anteilen – her. Die modernere Anlage I produziert erfahrungsgemäß 5 % Ausschuss, die veraltete Anlage II produziert 10 % Ausschuss. Das Garn wird in Containern zu je 1000 Garnrollen verpackt. Dabei etikettiert ein Automat die Garnrollen der Anlage I als erste Qualität (QI), die Garnrollen der Anlage II als zweite Qualität (QII). Die Container werden nach Sichtkontrolle anschließend von Hand ebenso mit QI bzw. QII etikettiert. Durch zeitweiligen Ausfall des Etikettierungsautomaten ist die Qualität der Garnrollen in einem Container nicht erkennbar. Es lässt sich auch nicht mehr nachvollziehen, von welcher der beiden Produktionsanlagen der Container bestückt worden ist. Ein Wirtschaftsmathematiker des Unternehmens erhält daher den Auftrag, einen Vorschlag zu erarbeiten, als welche Qualität Container mit zunächst nicht etikettierten Garnrollen unter möglichst geringen finanziellen Verlusten an Teppichfirmen verkauft werden können. Er hat bei der Entscheidungsfindung die folgenden Konsequenzen zu beachten:

- (K 0): Ein Verzicht auf Verkauf scheidet aus finanziellen Gründen aus.
- (K 1): Wird ein Container irrtümlich als QI verkauft, obwohl die Garnrollen nur QII sind, so können Schadensersatzforderungen eine Höhe von 7000 € erreichen.
- (K 2): Wird ein Container irrtümlich als QII verkauft, obwohl die Garnrollen QI sind, beträgt der Verlust 2000 €.
- (K 3) Wird eine Einzelprüfung von Garnrollen vorgenommen, so entstehen je geprüfter Garnrolle Kosten in Höhe von 2,5 €.

Lösung:

Nach erstem Abwägen der Konsequenzen könnte man für den Einzelfall eine „Extremalrechnung“ anstellen: Wegen $7000 \text{ € (K 1)} > 2500 \text{ € (K 3)} > 2000 \text{ € (K 2)}$ erscheint es am günstigsten, den Container ohne weitere Prüfung als QII zu verkaufen. Bleiben die Kosten einer Einzelprüfung (deutlich) unter 2000 € (also Stichprobenuntersuchung), so lohnt sich diese offensichtlich nur dann, wenn mit „akzeptabler Sicherheit“ die Qualität auch richtig erkannt wird. Unter dieser Sicht – und bei häufigerem Auftreten derartiger Etikettierungsprobleme – ist die **Entscheidung für einen „klassischen“ Alternativtest** angebracht.

Entweder es liegt QII ($H_0: p_0 = 0,10$) oder es liegt QI ($H_1: p_1 = 0,05$) vor. (Wegen des größten finanziellen Verlustes und ggf. Imageschäden bei einem irrtümlichen Verkauf als QI wird „QII“ als Nullhypothese mit dem Bestreben gewählt, H_0 mit sehr geringer Irrtumswahrscheinlichkeit – Fehler 1. Art – abzulehnen.) Wichtig ist das „Im-Auge-Behalten“ der beiden möglichen Fehlentscheidungen. Begeht man im vorliegenden Fall einen Fehler 1. Art (H_0 ist wahr, wird jedoch irrtümlich abgelehnt), so entsteht ein finanzieller Verlust von 7000 € (K 1). Liegt ein Fehler 2. Art vor (H_0 ist falsch, wird jedoch irrtümlich nicht abge-

lehnt), entsteht im Einzelfall „nur“ ein finanzieller Verlust von 2000 € (K 2).

Es wäre aber auch möglich, dass der Wirtschaftsmathematiker einen Vorschlag zur **Entscheidung ohne Stichprobenuntersuchung** (und Absicherung durch einen Signifikanztest) erarbeitet.

Da Etikettierungsprobleme häufiger auftreten, soll praktisch ohne zusätzlichen Aufwand gewährleistet werden, dass finanzielle Verluste (V) auf lange Sicht möglichst gering bleiben. Das heißt, der Erwartungswert $\mu(V)$ – wobei die Zufallsgröße V den Verlust (in Euro) beschreibt – muss möglichst klein bleiben. Unter der angegebenen Voraussetzung, dass das Garn im „Problem-Container“ mit gleicher Wahrscheinlichkeit der Produktionsanlage I oder II entstammt (beide Anlagen erreichen das gleiche Produktionsvolumen), gilt: $P(H_0) = P(H_1) = 0,5$. Demnach sind zwei generelle Entscheidungen denkbar. Ein „Problem-Container“ wird stets als QII (Entscheidung für H_0) oder stets als QI (Entscheidung für H_1) verkauft.

Wird als QII verkauft, entsteht stets genau dann ein Verlust, wenn der Verkauf als QII ungerechtfertigt ist (hier Fehler 2. Art). Wegen $\mu(V) = P(H_0) \cdot 2000 = 0,5 \cdot 2000 = 1000$ (s. Abschnitt H 4.2) ist dann auf lange Sicht ein durchschnittlicher Verlust von 1000 € zu erwarten.

Wird als QI verkauft, entsteht stets genau dann ein Verlust, wenn der Verkauf als QI ungerechtfertigt ist (hier Fehler 1. Art). Wegen $\mu(V) = P(H_1) \cdot 7000 = 0,5 \cdot 7000 = 3500$ ist dann auf lange Sicht ein durchschnittlicher Verlust von 3500 € zu erwarten.

Offensichtlich ist also die Entscheidung sinnvoll, stets als QII (mit der Konsequenz 2) zu verkaufen – will man auf Stichprobenuntersuchungen verzichten. Diese Entscheidung hat jedoch den Makel, dass von vornherein (wegen $P(H_0) = P(H_1) = 0,5$) auf lange Sicht in 50 % aller Fälle „einseitig“ (da stets QII gewählt ist, kann nie der größere finanzielle Verlust auftreten!) falsch entschieden wird, wenn auch mit erträglichen finanziellen Verlusten und ohne Imageschäden. Also kehrt der Wirtschaftsmathematiker **zurück zur Überlegung „klassischer“ Alternativtest**:

Die Zufallsgröße X beschreibe die Anzahl der Ausschuss-Garnrollen in einer Stichprobe vom Umfang n. Die Zufallsgröße X darf als binomialverteilt angenommen werden; $X \sim B_{n;p}$. Wegen der vergleichsweise hohen Kosten für die Einzelprüfung von Garnrollen (K 3) wird ein relativ geringer Stichprobenumfang von $n = 50$ gewählt. Somit entstehen (je Stichprobe) Fixkosten in Höhe von 125 €.

Testkonstruktion:

Nullhypothese $H_0: p_0 = 0,10$; $[H_1: p_1 = 0,05]$;

Stichprobenumfang: $n = 50$; Gewählte Irrtumswahrscheinlichkeit: $\alpha = 0,05$

Zufallsgröße X: $X \sim B_{50; 0,10}$ (bei wahrer Nullhypothese)

Da (sehr) kleine Werte von X gegen die Nullhypothese H_0 sprechen, wird der **Alternativtest als (einseitiger) linksseitiger Test** (vgl. S. 448) konstruiert;

$P(\bar{A}_{p_0}) = P(X \leq k) = B_{n; p_0}(\{0; 1; \dots; k\}) \leq \alpha$.

Der Tabelle der summierten Binomialverteilung $[B_{50; 0,10}(\{0; 1\}) = 0,03379 \leq 0,05]$ entnimmt man $k = 1$ und erhält als Ablehnungsbereich $\bar{A} = \{0; 1\}$.

Schlussfolgerung/Interpretation:

Enthält die Stichprobe höchstens eine Ausschuss-Garnrolle, so wird der „Problem-Container“ als QI verkauft (Ablehnung von H_0), ansonsten erfolgt der Verkauf als QII (Annahme von H_0).

Durch die bei dem vorliegenden Alternativtest gewählte Irrtumswahrscheinlichkeit α ist der *Höchstwert* für den Fehler 1. Art mit $\alpha = 0,05$ bereits bekannt. Wegen $\alpha = B_{n; p_0}(\bar{A})$ ist sein „genauer Wert“ gemäß Tabelle der summierten Binomialverteilung

$\alpha = B_{50; 0,10}(\{0; 1\}) = 0,03379$.

Für den Fehler 2. Art erhält man

$\beta = B_{n; p_1}(A) = B_{50; 0,05}(\{2; 3; \dots; 50\}) = 1 - B_{50; 0,05}(\{0; 1\}) = 1 - 0,27943 = 0,72057$.

Interpretation / Schlussfolgerung:

Die berechneten Wahrscheinlichkeiten für den Fehler 1. Art ($\alpha = 0,03379$; $\alpha \approx 3,4 \%$) und den Fehler 2. Art ($\beta = 0,72057$; $\beta \approx 72,1 \%$) belegen: Nur in etwa 3,4 % aller Fälle würde man (bei dieser Testkonstruktion) irrtümlich Garnrollen zweiter Qualität als Garnrollen erster Qualität (Konsequenz 1) verkaufen. Umgekehrt würde man in etwa 72,1 % aller Fälle irrtümlich Garnrollen erster Qualität als Garnrollen zweiter Qualität (Konsequenz 2) verkaufen. Die Frage ist nun, wie die finanziellen Verluste bei dieser Irrtumswahrscheinlichkeit auf lange Sicht beeinflusst werden. Es ist zu beachten:

- Jede Einzelprüfung der Garnrollen kostet 2,5 € (K 3).
- Wegen $P(H_0) = P(H_1) = 0,5$ gehen beide Fehlerarten (nach wie vor) zu gleichen Anteilen in die Verlustrechnung ein.
- Die in den Konsequenzen 1 und 2 angegebenen Kosten entstehen anteilig mit der jeweiligen Fehlerwahrscheinlichkeit.

Demnach gilt:

$$\mu(V_{n; k}) = P(H_0) \cdot \alpha \cdot 7000 + P(H_1) \cdot \beta \cdot 2000 + n \cdot 2,5$$

$$\mu(V_{50; 1}) = 0,5 \cdot 0,03379 \cdot 7000 + 0,5 \cdot 0,72057 \cdot 2000 + 50 \cdot 2,5 = 118,265 + 720,57 + 125$$

$$\mu(V_{50; 1}) \approx 963,84$$

Auf lange Sicht ist ein durchschnittlicher finanzieller Verlust von 963,84 € zu erwarten.

Da dieser finanzielle Verlust geringer ist als der ohne Stichprobenuntersuchung entstehende Verlust (1000 €), wird man bei wiederkehrender Entscheidungsnotwendigkeit die Variante mit Stichprobenuntersuchung bevorzugen.

Offen bleibt, ob sich der finanzielle Verlust noch weiter mindern lässt, z.B. durch Akzeptieren eines noch kleineren Fehlers 1. Art (in der Stichprobe dürfte dann keine Ausschuss-Garnrolle festgestellt werden). Eine formale Rechnung zeigt, dass praktisch keine weitere sinnvolle Reduzierungsmöglichkeit besteht.

Mit $\alpha = B_{50; 0,10}(\{0\}) = 0,00515$ und $\beta = B_{50; 0,05}(\{1; 2; \dots; 50\}) = 1 - B_{50; 0,05}(\{0\}) = 1 - 0 = 1$ gilt: $m(V_{50; 0}) = 0,5 \cdot 0,00515 \cdot 7000 + 0,5 \cdot 1 \cdot 2000 + 50 \cdot 2,5 = 18,025 + 1000 + 125$, also $\mu(V_{50; 0}) \approx 1643,03$.

• Hohe Produktqualität – wenig Ärger mit den Käufern

Abgesichert durch kontinuierliche Qualitätskontrollen und die damit verbundenen Erfahrungen zu Ausschuss- bzw. Mängelanteilen bei der Produktion verschiedenster Artikel geben Herstellerfirmen über den Einzelhandel (z.B. Kaufhausketten) i. Allg. auch Garantien an Käufer weiter (z.B. zeitlich begrenztes Umtauschrecht über den gesetzlich vorgeschriebenen Zeitraum hinaus). Eventuelle Reklamationen der Käufer werden (i. Allg. nach Mängelarten) registriert und ermöglichen zusätzlich Rückschlüsse hinsichtlich der erreichten Qualität. Nehmen Reklamationen zu, werden sich Kaufhauskette und Herstellerfirma verständigen und geeignete Maßnahmen einleiten. Das Beispiel J 12 soll verdeutlichen, wie geeignete Maßnahmen prüfstatistisch abgesichert werden können. Neben Hypothesen zu Wahrscheinlichkeiten wird mit Hypothesen zu Erwartungswerten gearbeitet (s. auch Abschnitt J 1.1).

J 12

Beispiel J 12:

Eine Kaufhauskette kauft bei einer Elektronikfirma regelmäßig preisgünstig CD-Player eines bestimmten Typs. Je Einkauf werden 200 dieser CD-Player erworben. Auf der Basis von Qualitätskontrollen gibt die Firma den Ausschussanteil mit (höchstens) 4 % an. Infolge häufiger Reklamationen von Käufern veranlasst die Kaufhauskette zusätzliche Qualitätskontrollen. Diese Qualitätskontrollen sollen durch einen geeigneten **Signifikanztest** abgesichert werden.

Die Zufallsgröße X beschreibe dabei die Anzahl der defekten CD-Player. Bei einer vollständigen Kontrolle eines zufällig ausgewählten Einkaufs (200 CD-Player) wurden 12 defekte Geräte festgestellt. Für die prüfstatistischen Untersuchungen wird zunächst gefordert:

Es sind der Erwartungswert $\mu = EX$ und die Standardabweichung $\sigma = \sqrt{V(X)}$ der Zufallsgröße X zu berechnen.

Es ist zu prüfen, ob die festgestellte Anzahl defekter Geräte als signifikante oder zufällige Abweichung von dem berechneten Erwartungswert angesehen werden darf. (Zum Vergleich werden wir einen Signifikanztest sowohl bei Binomialverteilung als auch bei Normalverteilung der Zufallsgröße X durchführen.)

Um die Herstellerqualität besser überwachen zu können, soll der Stichprobenumfang für Qualitätskontrollen in der Firma neu festgelegt werden: Unter Annahme des bekannten Ausschussanteils von 4 % ist die Mindestanzahl der CD-Player in einer Stichprobe (Stichprobenumfang n) so festzulegen, dass die Stichprobe mit höchstens 10 % Wahrscheinlichkeit keinen defekten CD-Player aufweist.

Lösung:

Der Erwartungswert und die Standardabweichung werden aus dem Stichprobenumfang $n = 200$ (vollständige Kontrolle), aus der Wahrscheinlichkeit für ein defektes Gerät ($p = 0,04$) und der Wahrscheinlichkeit für kein defektes Gerät ($q = 1 - p = 0,96$) berechnet (vgl. auch Abschnitte H 4.2 und H4.3).

Erwartungswert: $\mu = EX = n \cdot p = 8$; Standardabweichung: $\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot q} \approx 2,77$

Zur Prüfung, ob die festgestellte Anzahl defekter Geräte signifikant vom Erwartungswert abweicht, wird ein (einseitiger) **rechtsseitiger Signifikanztest** konstruiert. (Sehr) große Werte der Zufallsgröße X sprechen gegen die Nullhypothese.

Testkonstruktion unter Annahme der Binomialverteilung:

(Die Annahme der Binomialverteilung ist gerechtfertigt, da genau zwei mögliche Prüfergebnisse „defekt“ und „nicht defekt“ vorliegen. Die zugehörigen Wahrscheinlichkeiten ändern sich nicht und die Prüfungen sind unabhängig voneinander.)

Nullhypothese $H_0: p_0 \leq 0,04$; $[H_1: p_1 > 0,04]$;

$n = 200$;

gewähltes Signifikanzniveau: $\alpha = 0,05$

$X \sim B_{200; 0,04}$ (bei wahrer H_0)

Ermitteln des Ablehnungsbereiches \bar{A} : $\bar{A} = \{k; k + 1; \dots; 200\}$

Gesucht ist derjenige Wert $k - 1$, für den die Ungleichung $B_{200; 0,04}(\{0; 1; \dots; k - 1\}) \geq 0,95$ erstmals erfüllt ist. Da die summierte Binomialverteilung für diese Werte i. Allg. nicht tabelliert vorliegt, verwenden wir – wie schon im Abschnitt J 1.2 – einen geeigneten Rechner.

Die Ungleichung ist erstmals für den Wert $k - 1 = 13$, also $k = 14$ erfüllt (s. Fig. J 10). Als Ablehnungsbereich \bar{A} folgt dann $\bar{A} = \{14; 15; \dots; 200\}$.

Interpretation / Schlussfolgerung:

Da $X = 12$ und $12 \notin \bar{A}$, kann die Nullhypothese nicht abgelehnt werden. Die Abweichung vom Erwartungswert darf also (auf dem Signifikanzniveau $\alpha = 0,05$) als zufällig angesehen werden.

Ohne Verwendung eines geeigneten Rechners wären bei dieser Testkonstruktion die Berechnungen (wegen fehlender Tabellierung müsste die BERNOULLI-Formel wiederholt angewendet

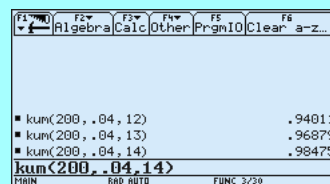


Fig. J 10

werden) sehr aufwändig. Es bietet sich daher an, zu prüfen, ob die Binomialverteilung der Zufallsgröße X durch eine (standardisierte) Normalverteilung (s. Abschnitt H 5.8) approximiert werden kann. Dies ist mit hinreichender Genauigkeit möglich, wenn $\sigma^2 = V(X) > 9$ bzw. $\sigma^2 = V(X) \geq \frac{1}{2} \cdot \sqrt{n}$ gilt (vgl. auch Beispiel J 7). Mit $\sigma^2 = 7,68$ erhalten wir $7,68 < 9$, aber $7,68 \geq \frac{1}{2} \cdot \sqrt{200} \approx 7,07$. Da das „Wurzelkriterium“ erfüllt ist, versuchen wir eine entsprechende Testkonstruktion.

Nullhypothese $H_0: \mu = EX = 8$

(Das explizite Formulieren einer Gegenhypothese ist nicht erforderlich. Es soll letztendlich lediglich entschieden werden, ob bei angenommenem Ausschussanteil $p = 0,04$ der Wert $X = 12$ noch zufällig vom Erwartungswert abweicht. Dies ist dann der Fall, wenn die Nullhypothese nicht abgelehnt werden kann.)
Stichprobenumfang: $n = 200$; gewähltes Signifikanzniveau: $\alpha = 0,05$

Ermitteln des Ablehnungsbereiches \bar{A} :

Unter Verwendung der globalen Näherungsformel von DE MOIVRE-LAPLACE (s. Beispiel J 7) folgt $\Phi\left(\frac{k-1+0,5-\mu}{\sigma}\right) \geq 1-\alpha$. Mit $\mu = 8$, $\sigma = 2,77$ und $\alpha = 0,05$ ergibt sich $\Phi\left(\frac{k-8,5}{2,77}\right) \geq 0,95$. Der Tabelle der Funktionswerte $\Phi(X)$ der Normalverteilung entnimmt man $\Phi(1,64) = 0,9495$ und setzt daher $\frac{k-8,5}{2,77} \geq 1,64$. Äquivalentes Umformen ergibt $k \geq 1,64 \cdot 2,77 + 8,5 \approx 13,04$, also $k = 14$. Der Ablehnungsbereich ist somit $\bar{A} = \{14; 15; \dots; 200\}$.

Interpretation / Schlussfolgerung:

Da $X = 12$ und $12 \notin \bar{A}$, kann die Nullhypothese nicht abgelehnt werden. Auch bei dieser Testkonstruktion unter der Annahme Normalverteilung – der Ablehnungsbereich stimmt mit dem bei der Binomialverteilung (auf dem Signifikanzniveau $\alpha = 0,05$) überein – wird die festgestellte Abweichung vom Erwartungswert als zufällig nachgewiesen.

Die gesuchte Mindestanzahl n der defekten CD-Player in der Stichprobe ist durch Berechnen der Mindestlänge einer BERNOULLI-Kette zu ermitteln. In diesem Fall beschreibe die Zufallsgröße Y die Anzahl der defekten Geräte in einer Stichprobe mit unbekanntem Stichprobenumfang n ; $Y \sim B_{n; 0,04}$.

Wegen $P(Y = 0) \leq 0,10$ folgt $P(Y = 0) = \binom{n}{0} \cdot 0,04^0 \cdot 0,96^n = 0,96^n$.

Aus $0,96^n \leq 0,10$, also $n \cdot \lg 0,96 \leq \lg 0,10$ bzw. $n \geq \frac{\lg 0,10}{\lg 0,96}$ erhält man $n \geq 56,4$ und somit die Mindestanzahl $n = 57$ für den gesuchten Stichprobenumfang.

Soll bei diesem Test auch das Kriterium $\sigma^2 = V(X) > 9$ erfüllt sein, müsste der Stichprobenumfang n vergrößert werden. Die vollständige Kontrolle könnte z. B. auf zwei zufällig ausgewählte Einkäufe der Kaufhauskette ($n = 2 \cdot 200 = 400$; Klumpenstichprobe) ausgedehnt werden. Mit $\sigma^2 = 400 \cdot 0,04 \cdot 0,96 \approx 15,36$ sind dann beide Kriterien erfüllt.

Übersicht zu den Berechnungen:

Normalverteilung von X	Binomialverteilung von X
$H_0: \mu = EX = 16$	$H_0: p_0 = 0,04$
$\Phi\left(\frac{k-16,5}{3,92}\right) \geq 0,95$; $\Phi(1,64) = 0,9495$ [s. Tab. $\Phi(X)$]	$B_{400; 0,04}(\{0; 1; \dots; k-1\}) \geq 0,95$ ist erstmals erfüllt
$k \geq 1,64 \cdot 3,92 + 16,5 \approx 22,93$, also $k = 23$	für $k-1 = 23$ [geeigneter Rechner; 0,966337]
$\bar{A} = \{23; 24; \dots; 400\}$	$\bar{A} = \{24; 25; \dots; 400\}$

H_0 kann nicht abgelehnt werden, wenn der festgestellte Wert (z. B. $X = 20$) nicht Element des Ablehnungsbereiches ist. Die Berechnungen bestätigen die hinreichend genaue Approximation (Abweichung der Ablehnungsbereiche um einen Wert).

• Noch einmal zur Geburtenstatistik

Eingangs des Kapitels J wurden zwei Fragen formuliert:

- * Wie viele Jungen (Mädchen) müssen anteilig (mindestens) unter einer bestimmten Anzahl von Neugeborenen sein, um die Geburtenstatistik „zu bestätigen“?
- * Mit welcher Sicherheit dürfen wir uns für eine Bestätigung entscheiden?

Diese beiden Fragen sollen nun noch einmal aufgegriffen werden.

Beispiel J 13:

Im Beispiel J 7 waren unter 200 Neugeborenen ($n = 200$) 102 Jungen (und folglich 98 Mädchen) registriert worden. Die festgestellten Werte sprachen nicht gegen die Geburtenstatistik (Signifikanzniveau $\alpha = 0,05$; also Mindestsicherheit von 95 %). Es ist nun – jeweils „aus Sicht der Jungengeburten“ und „aus Sicht der Mädchengeburten“ – zu prüfen, welche Aussagen bei Vergrößerung der Stichprobenumfänge ($n = 1000$; $n = 10^6$ und $n = 12 \cdot 10^6$) möglich sind.

Lösung:

Wird der Stichprobenumfang n vergrößert, so kann der bisher (auch im Beispiel J 7) verwendete Rechner bei einigen notwendigen Berechnungen auf der Basis der Binomialverteilung mehr als 30 Minuten benötigen. Wir arbeiten daher mit der hinreichend genauen **Normalverteilung**. (Sie darf genutzt werden, wie eine Überprüfung, vgl. z.B. Beispiel J 7, zeigen würde.)

Unter n Neugeborenen ist mit der Anzahl der Jungengeburten (Mädchengeburten) stets auch die Anzahl der Mädchengeburten (Jungengeburten) bekannt. Die Geburtenstatistik wird praktisch nicht „bestätigt“, wenn in der Stichprobe zu viele Jungen (zu wenig Mädchen) oder zu viele Mädchen (zu wenig Jungen) gezählt werden. Man kann sich daher für die Konstruktion eines einseitigen Signifikanztests (entweder rechts- oder linksseitig) entscheiden. Ausgehend von dem registrierten höheren Anteil (i. Allg. ist dies die Anzahl der Jungengeburten, hier beschrieben durch die Zufallsgröße X), wählen wir einen **rechtsseitigen Signifikanztest** (große Werte von X sprechen gegen die Nullhypothese). Als Nullhypothese wird „vorsichtig“ – mit dem Bestreben, sie abzulehnen – formuliert: „Der registrierte Anteil der Jungengeburten erreicht (höchstens) den langjährigen Erfahrungswert ($H_0: p_0 \leq 0,514$).“

(1a) *Untersuchung aus der Sicht der Jungengeburten* für $n = 1000$; $p = 0,514$

Mit $\alpha = 0,05$, $m = EX = 514$ und $\sigma = \sqrt{V(x)} \approx 15,81$ folgt aus $\Phi\left(\frac{k-1+0,5-\mu}{\sigma}\right) \geq 1 - \alpha$

unter Verwendung des Tabellenwertes $\Phi(1,64) = 0,9495$ der Ausdruck $\frac{k-514,5}{15,81} \geq 1,64$ und nach äquivalentem Umformen $k \geq 1,64 \cdot 15,81 + 514,5 \approx 540,43$; $k = 541$.

Allgemein bei $\alpha = 0,05$: $k \geq 1,64 \cdot \sigma + \mu + 0,5$.

(Zum Rechenweg vergleiche man mit den ausführlicheren Berechnungen in den Beispielen J 7 und J 12). Befinden sich unter 1000 Neugeborenen sehr viele Jungen (mehr als 540 Jungen), so wird die Geburtenstatistik nicht bestätigt.

(1b) *Untersuchung aus der Sicht der Mädchengeburten* für $n = 1000$; $p = 0,486$

Mit $\alpha = 0,05$, $\mu = EX = 486$ und $\sigma = \sqrt{V(x)} \approx 15,81$ folgt in Analogie zu den obigen Berechnungen $k \geq 1,64 \cdot 15,81 + 486,5 \approx 512,43$; $k = 513$.

Befinden sich unter 1000 Neugeborenen sehr viele Mädchen (mehr als 513 Mädchen), wird die Geburtenstatistik nicht bestätigt.

(2a) *Untersuchung aus der Sicht der Jungengeburten* für $n = 10^6$; $p = 0,514$

Die Anzahl der Neugeborenen lag in Deutschland 1999 bei etwa 770 000. Der Stichprobenumfang würde also einen Beobachtungszeitraum (gemessen am Jahrgang 1999) von ca. 1,5 Jahren umfassen.

Mit $\alpha = 0,05$, $\mu = EX = 514\,000$ und $\sigma = \sqrt{V(x)} \approx 499,80$ erhält man

$k \geq 1,64 \cdot 499,80 + 514\,000,5 \approx 14820,17$; $k = 514821$.

Befinden sich unter 1 Million Neugeborenen sehr viele Jungen (mehr als 514 820 Jungen), wird die Geburtenstatistik nicht bestätigt.

(2b) *Untersuchung aus der Sicht der Mädchengeburten* für $n = 10^6$; $p = 0,486$

Mit $\alpha = 0,05$, $\mu = EX = 486\,000$ und $\sigma = \sqrt{V(x)} \approx 499,80$ erhält man

$k \geq 1,64 \cdot 499,80 + 486\,000,5 \approx 486\,820,17$; $k = 486\,821$.

Befinden sich unter 1 Million Neugeborenen sehr viele Mädchen (mehr als 486 820 Mädchen), wird die Geburtenstatistik nicht bestätigt.

(3a) *Untersuchung aus der Sicht der Jungengeburten* für $n = 12 \cdot 10^6$; $p = 0,514$

Der Stichprobenumfang würde einen Beobachtungszeitraum (gemessen am Jahrgang 1999) von ca. 18 Jahren umfassen.

Mit $\alpha = 0,05$, $\mu = EX = 6\,168\,000$ und $\sigma = \sqrt{V(x)} \approx 1731,37$ erhält man

$k \geq 1,64 \cdot 1731,37 + 6\,168\,000,5 \approx 6\,170\,839,45$; $k = 6\,170\,840$.

Befinden sich unter 12 Million Neugeborenen sehr viele Jungen (mehr als 6 170 839 Jungen), wird die (bisherige) Geburtenstatistik nicht bestätigt.

(3b) *Untersuchung aus der Sicht der Mädchengeburten* für $n = 12 \cdot 10^6$; $p = 0,486$

Mit $\alpha = 0,05$, $\mu = EX = 583\,2000$ und $\sigma = \sqrt{V(x)} \approx 1731,37$ erhält man

$k \geq 1,64 \cdot 1731,37 + 5\,832\,000,5 \approx 5\,834\,839,45$; $k = 5\,834\,840$.

Befinden sich unter 12 Million Neugeborenen sehr viele Mädchen (mehr als 5 834 839 Mädchen), wird die (bisherige) Geburtenstatistik nicht bestätigt.

Interpretation/Schlussfolgerung:

Die Berechnungen in (1) bis (3) lassen gewisse Gesetzmäßigkeiten erkennen. Offensichtlich wird die Testqualität mit Vergrößerung des Stichprobenumfangs n immer besser.

Das gewählte Signifikanzniveau legt über den berechneten k -Wert (mit dem Annahme- bzw. Ablehnungsbereich) den Entscheidungsspielraum im Test fest. Dessen „Größe“ kann im vorliegenden Beispiel sehr anschaulich durch die Addition der einander entsprechenden k -Werte (k_J – Jungen; k_M – Mädchen) dargestellt werden. Interpretiert man dabei den über den eigentlichen Stichprobenumfang hinausgehenden Anteil als ein Maß für die Größe des Entscheidungsspielraumes, so wird diese mit zunehmendem n (bei gleich bleibendem Signifikanzniveau α) prozentual immer kleiner. Das heißt, die Güte des Tests (s. auch Abschnitt J 1.4) lässt sich mit wachsendem Stichprobenumfang n deutlich verbessern (s. S. 451).

Addiert man in (1) bis (3) jeweils die k -Werte, so erhält man

(1) $k_J + k_M = 1054$; (2) $k_J + k_M = 1\,001\,642$; (3) $k_J + k_M = 12\,005\,680$.

Anteilig wird der sich ergebende „Überhang“ mit wachsendem n immer geringer.

In (1) beträgt er noch $\frac{54}{1000} = 0,054$ (5,40 %), in (2) $\frac{1642}{10^6} = 0,001642$ ($\approx 0,16$ %);

in (3) nur noch $\frac{5680}{12 \cdot 10^6} \approx 4,73 \cdot 10^{-4}$ ($\approx 0,05$ %).

Dies darf als deutliches Zeichen für die verbesserte Testqualität (bei gleichem Signifikanzniveau α) angesehen werden. Ein Beobachtungszeitraum von ca. 18 Jahren mit ca. $12 \cdot 10^6$ Neugeborenen könnte bereits in der beschreibenden Statistik (nach dem empirischen Gesetz der großen Zahlen; s. Abschnitt H 1.3) zur („recht zuverlässigen“) Angabe von relativen Häufigkeiten genutzt werden. „Auffällige“ Abweichungen von $p = 0,514$ bzw. $p = 0,486$ müssten dann prüfstatistisch auf Signifikanz untersucht werden.

• Untersuchungen zur Mindesthaltbarkeitsdauer (MHD)

In der Lebensmittelindustrie werden Erzeugnisse durch moderne Konservierungsverfahren haltbar gemacht. Die Hersteller sind verpflichtet, eine so genannte Mindesthaltbarkeitsdauer ihrer Erzeugnisse zu garantieren.

Beispiel J 14:

Um die Konkurrenzfähigkeit des Erzeugnisses M zu verbessern, soll geprüft werden, ob durch ein verändertes Konservierungsverfahren eine längere MHD garantiert werden kann. Nach dem alten Verfahren erreichen 50 % aller Erzeugnisse M die angestrebte längere MHD.

Es werden 200 nach dem veränderten Verfahren konservierte Erzeugnisse auf Haltbarkeit untersucht. Die Zufallsgröße X beschreibe die Anzahl der Erzeugnisse M, die die längere MHD erreichen. Als Signifikanzniveau wird $\alpha = 0,05$ gewählt.

(1) Es sei noch nicht bekannt, ob das veränderte Verfahren tatsächlich zu einer längeren oder evtl. sogar kürzeren MHD führt; **zweiseitiger Signifikanztest:**

$$H_0: p_0 = 0,5 \quad [H_1: p_1 \neq 0,5] \quad X \sim B_{200; 0,5} \text{ (bei wahrer } H_0)$$

Aus $B_{200; 0,5}(\{0; 1; \dots; k_L\}) \leq 0,025$ und $B_{200; 0,5}(\{0; 1; \dots; k_R - 1\}) \geq 0,975$ erhält man $k_L = 85$, $k_R - 1 = 114$ bzw. $k_R = 115$ (Tabellenwerte) und somit den Ablehnungsbereich $\bar{A} = \{0; 1; \dots; 85\} \cup \{115; 116; \dots; 200\}$. Ist die Anzahl der Erzeugnisse M, die die längere MHD erreichen, kleiner als 86 oder größer als 114, so hat das veränderte Verfahren signifikante Auswirkungen.

(2) Längerfristige Untersuchungen lassen die Vermutung zu, dass das veränderte Verfahren zu einer längeren MHD führt; (einseitiger) **linksseitiger Signifikanztest:**

$$H_0: p_0 \geq 0,5 \quad [H_1: p_1 < 0,5] \quad X \sim B_{200; 0,5} \text{ (bei wahrer } H_0)$$

Aus $B_{200; 0,5}(\{0; 1; \dots; k\}) \leq 0,05$ erhält man $k = 87$ (Tabellenwert) und somit den Ablehnungsbereich $\bar{A} = \{0; 1; \dots; 87\}$. Erreichen weniger als 88 Erzeugnisse M die längere MHD, kann die Vermutung nicht bestätigt werden (Ablehnung von H_0).

(3) Längerfristige Untersuchungen lassen die Vermutung zu, dass das veränderte Verfahren zu einer kürzeren MHD führt; (einseitiger) **rechtsseitiger Signifikanztest:**

$$H_0: p_0 \leq 0,5 \quad [H_1: p_1 > 0,5] \quad X \sim B_{200; 0,5} \text{ (bei wahrer } H_0)$$

Aus $B_{200; 0,5}(\{0; 1; \dots; k - 1\}) \geq 0,95$ erhält man $k - 1 = 112$ (Tabellenwert) bzw. $k = 113$ und somit den Ablehnungsbereich $\bar{A} = \{113; 114; \dots; 200\}$. Erreichen mehr als 112 Erzeugnisse M die längere MHD, kann die Vermutung nicht bestätigt werden (Ablehnung von H_0).

(4) Längerfristige Untersuchungen lassen die Vermutung zu, dass 80 % aller Erzeugnisse M die längere MHD erreichen; (einseitiger) **linksseitiger Alternativtest:**

$$H_0: p_0 = 0,8 \quad H_1: p_1 = 0,5 \quad X \sim B_{200; 0,8} \text{ (bei wahrer } H_0)$$

Aus $B_{200; 0,8}(\{0; 1; \dots; k\}) \leq 0,05$ erhält man $k = 72$ (Tabellenwert) und somit den Ablehnungsbereich $\bar{A} = \{0; 1; \dots; 72\}$. Erreichen weniger als 73 Erzeugnisse M die längere MHD, kann die Vermutung nicht bestätigt werden (Ablehnung von H_0) und es ist weiterhin von der 50 %-Angabe (H_1) auszugehen.