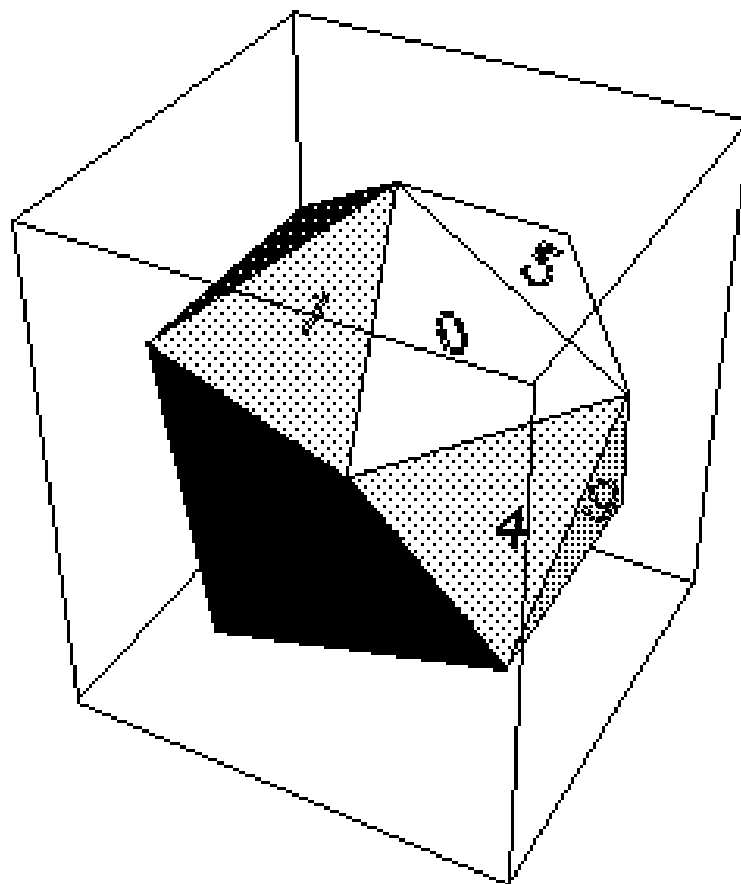


# ***Testen von Hypothesen***



## ***Ein Leitprogramm in Mathematik***

von  
**Otto M. Keiser**

Herausgegeben von  
**Urs Kirchgraber**

---

## 0

**ETH-Leitprogramm "Testen von Hypothesen"**

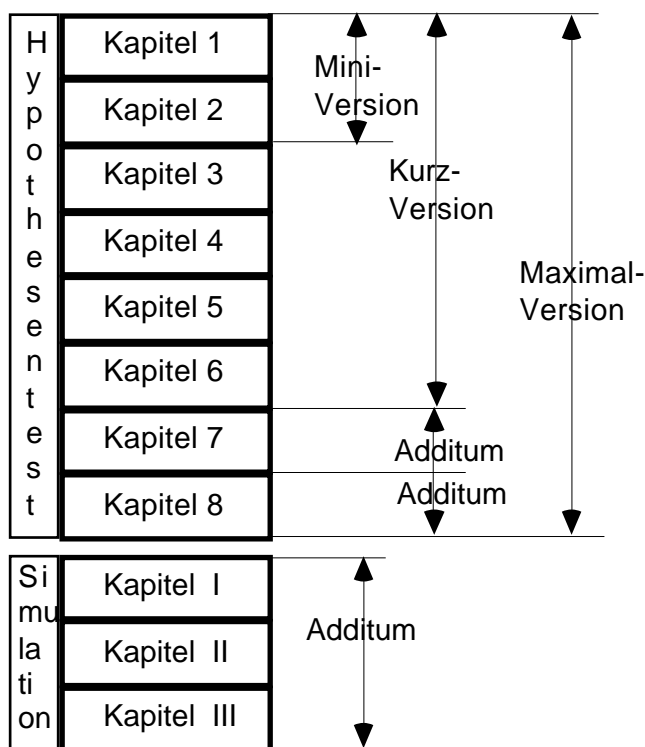
<b>Version</b>	Juni 1998
<b>Fach</b>	Mathematik
<b>Schulbereich</b>	Gymnasium
<b>Vorkenntnisse:</b>	Grundkenntnisse in der Kombinatorik Grundkenntnisse in der Wahrscheinlichkeitsrechnung Grundfertigkeiten auf einem Taschencomputer

**Bearbeitungsmodelle**

Das vorliegende Leitprogramm ist ziemlich umfangreich. Es ist jedoch ohne weiteres möglich, die Arbeit auf die ersten 6, 7 oder 8 Kapitel zu beschränken. Entsprechend kann die an die Schüler abgegebene Zahl der Kopien reduziert werden.

Als *Minimalversion* sind sogar nur die ersten zwei Kapitel denkbar.

Die nicht behandelten Kapitel eignen sich auch als Additum. Der Anhang S über die Simulation kann parallel zu den Kap. 1 bis 8 oder als Additum bearbeitet oder auch weggelassen werden.



<b>Zeitbedarf</b>	<p>Die Arbeit am Leitprogramm kann, wie gesagt, nach dem 6., 7. oder 8. Kapitel ohne weiteres abgebrochen werden.</p> <ul style="list-style-type: none"><li>• Kap. 1 bis 6: 10 Lektionen</li><li>• Kap. 7: 3 Lektionen</li><li>• Kap. 8: 2 Lektionen</li></ul> <p>immer inkl. Hausaufgabenzeit.</p>
<b>Bezugsquelle</b>	<p>Urs Kirchgraber ETH Zürich, Departement Mathematik 8092 Zürich</p>
<b>Autor</b>	<p>Otto M. Keiser Freies Gymnasium in Zürich Arbenzstrasse 19 Postfach 8034 Zürich</p>
<b>Kopierrecht</b>	<p>Für den Schulgebrauch unbeschränkt.</p>
<b>Kommerzielle Nutzung</b>	<p>Nicht erlaubt</p>

### **Dank**

Das Leitprogramm "Testen von Hypothesen" ist während meines Weiterbildungsurlaubes im Rahmen von "ETH für die Schule" entstanden. Ich danke Prof. Urs Kirchgraber, dem Initianten und Leiter dieses Programms, für die Einladung, für seine Ermutigung und Unterstützung und seine Offenheit gegenüber neuen Ideen.

Wichtige Hinweise und Anregungen stammen von Prof. H. R. Künsch. Ihm gebührt ein grosser Dank für seine Bereitschaft, das umfangreiche Manuskript zu lesen und als "statistisches Gewissen" zu wirken. Die Verantwortung für allfällige Fehler liegt jedoch ausschliesslich bei mir.

Dankbar bin ich auch für Anregungen von Moritz Adelmeyer und Niklaus Sigrist. Edgar Stahel und Hansjörg Stocker haben die Rohfassung des Manuskriptes kritisch durchgesehen. Daniel Reichmuth hat das Leitprogramm als erster in einer Klasse des MNG Rämibühl ausprobiert. Marco Bettinaglio hat mich bei der Gestaltung beraten und die Schlussfassung korrigiert.

## **Einführung: Wie viele Knallkörper explodieren?**

Ein Grossverteiler hat im Hinblick auf den bevorstehenden Silvester 100'000 Knallkörper bei einem neuen Lieferanten eingekauft. Er kann und will es sich nicht leisten, schlechte Qualität anzubieten.

Im Kaufvertrag wurde deshalb vereinbart, dass die Sendung höchstens 4% Ausschuss, d.h. höchstens 4000 defekte Exemplare enthalten darf.

Gleichzeitig wurde festgehalten, wie die Qualität der Sendung zu testen ist: Der Einkäufer wählt völlig zufällig 200 Exemplare aus und zündet sie. Wenn mehr als 13 Knallkörper nicht explodieren, so darf der Einkäufer die Sendung zurückweisen und braucht nichts zu bezahlen. Andernfalls muss er die Sendung akzeptieren.

Es wäre nicht verwunderlich, wenn Sie von diesem Prüfverfahren überrascht wären und wenn Sie sich, je nach Standpunkt, die eine oder andere der folgenden Fragen gestellt hätten:

- Ist der Einkäufer nicht viel zu grosszügig, wenn er bis zu 13 defekte Knallkörper akzeptiert? Sollte er die Sendung nicht schon ablehnen, wenn die Stichprobe mehr als 4%, d. h. mehr als 8 defekte Knallkörper enthält?
- Ist dieser Test nicht etwas hart für den Lieferanten? Man sagt sich doch: Wenn die Sendung gerade noch in Ordnung ist, müsste der Einkäufer sowieso mit 8 defekten Knallkörpern rechnen. Aber wenn's „der Zufall will“, geraten „leicht“ einmal auch 10 oder eben 14 defekte Exemplare in die Auswahl. Und im letzten Fall würde die Sendung zu Unrecht abgelehnt.
- Ist dieser Test auch genügend scharf für den Einkäufer? Würde er wirklich entdecken, dass die Sendung beispielsweise 8% Ausschuss enthält statt der erlaubten 4%?

In diesem Leitprogramm geht es darum, solche Fragen zu beantworten. In einem ersten Schritt werden wir Antworten erhalten, indem wir das Prüfverfahren auf dem Computer imitieren, nachspielen oder wie man auch sagt, simulieren. Danach bemühen wir uns um rechnerische Antworten. Hier sind Grundkenntnisse in der Kombinatorik und in der Wahrscheinlichkeitsrechnung gefragt und Sie müssen mit Ihrem Taschenrechner umgehen können.

Wir werden nicht bei diesem einen Qualitätstest stehen bleiben, sondern weitere interessante Fragestellungen aus der Medizin, der Meinungsforschung und mehreren anderen Gebieten anschneiden.

Nebenbei: Es ist Ihnen sicher klar, dass man die Qualität der Knallkörperlieferung grundsätzlich nur aufgrund ausgewählter Einzelexemplare untersuchen kann; denn letzte Klarheit über die Qualität müsste man sich durch vollständige Vernichtung der Sendung erkaufen (im doppelten Sinne des Wortes) ....

## **Inhaltsverzeichnis**

<b>Einführung: Wie viele Knallkörper explodieren?</b>	0.3
<b>Arbeitsanleitung</b>	0.7
<b>1 Ein Modell für den Qualitätstest- Welche Fehler sind möglich?</b>	1.1
Lernkontrolle: Die theoretische Fahrprüfung	1.5
Lösungen der Aufgaben zu Kapitel 1	1.6
<b>2 Computersimulation des Qualitätstests</b>	2.1
2.1 Simulation mit Ikosaedern	2.2
2.2 Die Fehler 1. und 2. Art simuliert	2.5
Lernkontrolle: Der Fehler 2. Art	2.7
Lösungen der Aufgaben zu Kapitel 2	2.8
<b>3 Grundlagen zur Wahrscheinlichkeitsrechnung</b>	3.1
Lernkontrolle: Wie viele Luchse gibt es?	3.6
Lösungen der Aufgaben zu Kapitel 3	3.7
<b>4 Der Qualitätstest unter der rechnerischen Lupe</b>	4.1
4.1 Der Fehler 1. Art - das Signifikanzniveau	4.1
4.2 Der Fehler 2. Art - die Macht des Tests	4.3
Lernkontrolle: Produktionskontrolle	4.4
Lösungen der Aufgaben zu Kapitel 4	4.5
<b>5 Fachbegriffe bei statistischen Tests</b>	
5.1	
5.1 Die Nullhypothese	5.2
5.2 Die Alternativhypothese	5.3
5.3 Die Testgrösse	5.4
5.4 Der Verwerfungsbereich - der Annahmebereich	5.4
Lernkontrolle: Die Begriffe in der Übersicht	5.6
Lösungen der Aufgaben zu Kapitel 5	5.7
<b>6 Exakter Test von Fisher: Wie einen Herzinfarkt behandeln?</b>	6.1
6.1 Die Vierfeldertafel im Licht der Statistikbegriffe	6.2
6.2 Die Ergebnisse einer Computersimulation	6.4
6.3 Wie gross ist     wirklich?	6.5
Lernkontrolle: Noch mehr Herzinfarkte	6.7
Lösungen der Aufgaben zu Kapitel 6	6.8

---

<b>7 Test mit Binomialverteilung: Meinungsumfragen</b>	7.1
7.1 Ein neues Urnenmodell	7.2
7.2 Statistische Begriffe bei der Meinungsumfrage	7.3
7.3 Bestimmung von $\mu$ und $\sigma$ durch Simulation	7.5
7.4 Die Binomialverteilung	7.6
7.5 Die Berechnung von $\mu$ und $\sigma$	7.10
Lernkontrolle: Voraussichtliche Stimmbeteiligung	7.11
Lösungen der Aufgaben zu Kapitel 7	7.12
<b>8 Ein zweiseitiger Vorzeichentest: Welche Sorte pflanzen?</b>	
8.1	
8.1 Die Nullhypothese beim Sortenvergleich	8.2
8.2 Die Alternativhypothese beim Sortenvergleich	8.2
8.3 Die Testgrösse beim Sortenvergleich	8.3
8.4 Der Verwerfungsbereich beim Sortenvergleich	8.4
8.5 Berechnung von $\mu$	8.5
8.6 Berechnung von $\sigma$	8.6
Lernkontrolle: Zwei Messmethoden vergleichen	8.7
Lösungen der Aufgaben zu Kapitel 8	8.8

### **Anhang S: Simulationen**

<b>I Computersimulationen zu Kapitel 2</b>	I.1
I.1 Simulation als Tabellenkalkulation	I.1
I.1.1 Simulation von 10-er Stichproben	I.1
I.1.2 Hundert Simulationen einer 200-er Stichprobe	I.8
I.2 Simulation auf dem Taschencomputer	I.10
Lösungen der Aufgaben zu Kapitel I	I.12
<b>II Simulation zu Kapitel 6</b>	II.1
II.1 Simulation auf dem Taschencomputer	II.1
II.2 Simulation mit einem EXCEL-Makro	II.3
<b>III Simulation zu Kapitel 7</b>	III.1
III.1 Simulation mit EXCEL	III.1
III.2 Simulation auf dem Taschencomputer	III.2

**Anhang K für die Lehrperson**  
**Kapiteltests**

Kapiteltest 1: Pilzkontrolle	K.1
Kapiteltest 2: Gewicht von Tomatenkonserven	K.2
Kapiteltest 3: Therapien nach Herzinfarkt	K.3
Kapiteltest 4: Produktionskontrolle bei Kugelschreiberminen	K.4
Kapiteltest 5: Max & Moritz werfen eine Münze	K.5
Kapiteltest 6: Kriminelle Zwillinge	K.6
Kapiteltest 7: ESP = Extrasensory Perception	K.7
Kapiteltest 8: Der Einfluss des rooming-in	K.8
Quellen	K.9

## Arbeitsanleitung

Die Kapitel dieses Leitprogrammes sind immer etwa gleich aufgebaut. Sie beginnen mit einer kurzen Übersicht, die Sie über das Kapitelthema orientiert. Dann werden die Lernziele beschrieben; sie geben an, was im betreffenden Kapitel gelernt werden soll.

Danach beginnt der eigentliche Text, den Sie selbständig durcharbeiten sollen. Sie können dabei Ihr eigenes Arbeitstempo anschlagen. Und Sie haben anhand von eingestreuten Aufgaben laufend Gelegenheit, zu kontrollieren, ob Sie den Stoff verstanden haben. Die ausführlichen Lösungen finden Sie nämlich am Schluss des jeweiligen Kapitels.

Der Lehrtext endet immer mit einer grösseren Aufgabe als Lernkontrolle. Sie können anhand der ebenfalls abgedruckten Lösung selbständig prüfen, ob Sie die Lernziele des Kapitels erreicht haben.

Danach können Sie sich bei Ihrer Lehrperson zum Kapiteltest anmelden. Das Ergebnis entscheidet, ob Sie im nächsten Kapitel weiterfahren können, oder ob Sie sich nochmals in das alte Kapitel vertiefen müssen.

Bevor Sie nun an die Lektüre des ersten Kapitels gehen, müssen Sie sich noch über die Bedeutung der Symbole orientieren, die den Text am Rand begleiten.



Beim Wegweiser steht, wohin die "Reise" geht und welche Lernziele Sie erreichen sollten.



Wichtige Informationen oder Zusammenfassungen sind von einem Ausrufezeichen begleitet. Oft sind jene zusätzlich eingerahmt.



Immer wenn Sie auf den schaufelnden Arbeiter stossen, müssen Sie auf einem separaten Blatt die Aufgaben lösen und danach Ihre Lösung mit der gedruckten vergleichen.



Am Ende des Kapitels stossen Sie auf die Zolltafel: Sie müssen die Lernkontrolle absolvieren. "Selbstdeklaration" ist angesagt! Sie entscheiden selber, ob Sie reif sind für den Kapiteltest.



## 1

## Ein Modell für den Qualitätstest - Welche Fehler sind möglich?



Das einführende Beispiel von den Knallkörpern schildert einen statistischen Qualitätstest. Anhand dieses praktischen Beispiels erfahren Sie im ersten Kapitel, wie man Tests in der Statistik modelliert und dass Fehler bei solchen Tests unvermeidlich sind. Im Einzelnen werden Sie die Begriffe *Grundgesamtheit*, *Stichprobe* und *Fehler 1. und 2. Art* kennen lernen.

Der Abschnitt informiert abschliessend ganz kurz über das Wesen der Statistik und die Rolle, welche die Wahrscheinlichkeitsrechnung dabei spielt.

### Urnenmodell

Statistiker und Statistikerinnen ziehen es vor, anstelle von Knallkörpern neutrale (gleich grosse) Kugeln zu betrachten. Die Kugeln liegen in ihrer Vorstellung in einer Kiste, der sog. *Urne* (Fig. 1.1). In unserem Fall enthält die Urne  $N=100'000$  Kugeln. Den defekten Knallkörpern entsprechen schwarze Kugeln; ihre (unbekannte) Anzahl nennen wir  $S$ .

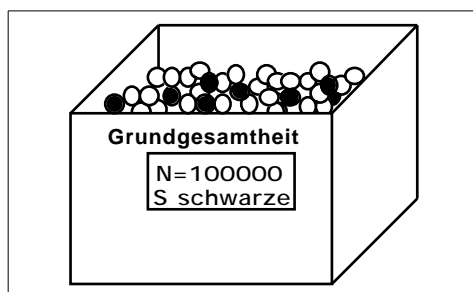


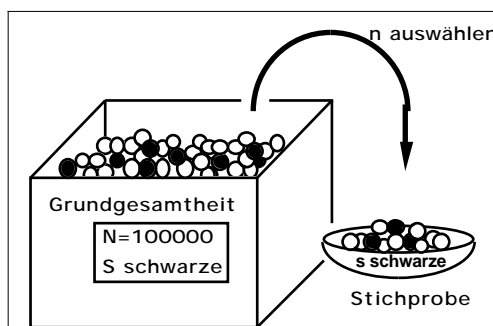
Fig. 1.1: Die Grundgesamtheit im Urnenmodell

Alle anderen Kugeln, ihre Anzahl beträgt  $N-S$ , sind weiss; sie repräsentieren die funktionierenden Knallkörper. Die weissen und schwarzen Kugeln insgesamt bilden die sog. *Grundgesamtheit*.

Dieses Beispiel zeigt Ihnen, wie die Statistik die Realität in ein universales, abstraktes und übersichtliches Gedankenmodell, eben das *Urnenmodell*, übersetzt.

### Stichprobe

Dem zufälligen Auswählen von 200 Knallkörpern in der Realität entspricht im Urnenmodell etwa der folgende Vorgang: Wir mischen die Kugeln kräftig, ziehen mit verbundenen Augen eine Kugel und legen sie beiseite. Dieses Prozedere wiederholen wir 199-mal.



Wir haben so eine sog. *Stichprobe* erhalten (Fig. 1.2). Die Zahl  $n=200$  nennen wir den *Umfang* der Stichprobe.

Fig. 1.2: Eine Stichprobe aus der Grundgesamtheit im Urnenmodell.

Die nachstehende Figur 1.3 zeigt drei mögliche Stichproben vom Umfang  $n=200$  und  $s=1$ ,  $s=5$  und  $s=10$  schwarzen Kugeln.

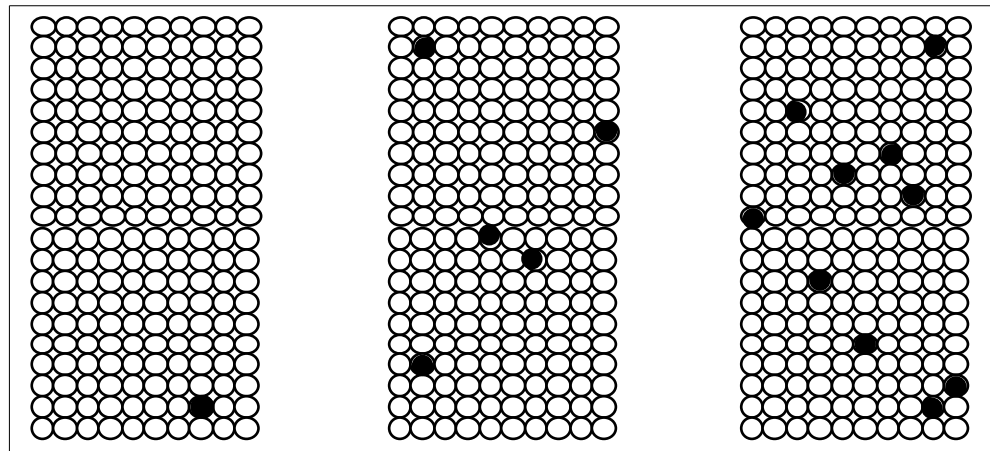


Fig. 1.3: Drei mögliche Stichproben mit dem Umfang 200.



### Aufgabe 1.1

Mit seinem Qualitätstest möchte unser Einkäufer – in der Sprache des Urnenmodells formuliert – herausfinden, ob die Urne (die Grundgesamtheit) nicht mehr als 4000 schwarze Kugeln enthält. Weshalb zählt er nicht einfach in der Urne nach?

### Die Grundidee des Tests



Der Einkäufer hofft, dass er sich mit seiner Stichprobe ein kleines Spiegelbild der Grundgesamtheit verschafft hat. Wenn die Grundgesamtheit gerade noch den Qualitätsanforderungen entspricht, so enthält sie, wie gesagt, genau  $S=4000$  schwarze Kugeln. Der Einkäufer müsste dann mit  $s=8$  schwarzen Kugeln in der Stichprobe rechnen, aber selbstverständlich können es mehr oder weniger sein, in den Extremfällen sogar 0 oder 200.

Gefühlsmässig ist uns jedoch klar: Grosse Abweichungen von 8 sind *unwahrscheinlich*. Und unser Einkäufer will in *diesem* Sinne entscheiden! Wenn die Stichprobe mehr als 13 schwarze Kugeln enthält, glaubt er nicht mehr daran, dass die Grundgesamtheit nur 4000 schwarze Kugeln enthält. Deshalb lehnt er die Sendung ab.

Hand aufs Herz! Sehen Sie eine grundsätzlich andere, realistische Möglichkeit, die Qualität von Knallkörpern zu überprüfen? Aber natürlich ist dieses bestechend einfache Verfahren mit gewissen Risiken behaftet. Diese müssen wir untersuchen und irgendwie messen.

### Der Fehler 1. Art

Beim geschilderten Entscheidungsschema nimmt der Einkäufer (und vor allem der Lieferant!) in Kauf, dass die Sendung *irrtümlicherweise* abgelehnt wird, weil 14 oder mehr schwarze Kugeln in die Stichprobe geraten, *obwohl*

die *Sendung in Ordnung* ist (nicht mehr als 4000 schwarze Kugeln enthält). Man bezeichnet diesen Fehlentscheid als Fehler 1. Art.

Es hängt vom Zufall ab, ob es zu einem solchen Fehler kommt. Deshalb kann man einem Fehler 1. Art eine Wahrscheinlichkeit zuordnen. Diese hängt natürlich von der *tatsächlichen* Anzahl schwarzer Kugeln in der Urne ab. Sie ist am grössten, wenn die maximale Anzahl von 4000 schwarzen Kugeln vorhanden ist. Der Maximalwert dieser Wahrscheinlichkeit heisst traditionsgemäss .



Zusammenfassend können wir sagen: Wenn die *Sendung in Ordnung* ist, sind höchstens 4000 schwarze Kugeln in der Urne und es ist

$$P(\text{Fehler 1. Art}) = \dots$$

In den nächsten Kapiteln werden wir uns näher mit der Bestimmung von befassen.

### Der Fehler 2. Art

Der Fehler 1. Art trifft den Lieferanten. Bei unserem Qualitätstest geht aber auch der Einkäufer ein (beachtliches) Risiko ein: Nehmen wir an, die Urne enthalte weit mehr als 4000, z. B. 8000 schwarze Kugeln. Trotzdem könnte es in der Stichprobe nicht mehr als 13 schwarze Kugeln haben. Der Einkäufer würde in diesem Fall also nicht realisieren, dass die *Sendung nicht* in Ordnung ist und diese akzeptieren. Man spricht hier vom Fehler 2. Art. Wieder hängt es vom Zufall ab, ob dieser Fehler gemacht wird, und man kann diesem eine Wahrscheinlichkeit zuordnen. Diese wird üblicherweise mit  $\beta$  bezeichnet.

Die Wahrscheinlichkeit

$$\beta = P(\text{Fehler 2. Art})$$

ist aber erst festgelegt, wenn wir über die Anzahl schwarzer Kugeln eine Annahme machen. Diese Annahme muss nicht notwendigerweise "4000 schwarze Kugeln", sondern kann z. B. "8000 schwarze Kugeln" sein.



### Fehler 1. und 2. Art (Zusammenfassung)

Bei unserem Qualitätstest sind zwei mögliche *Fehlentscheidungen* grundsätzlich unvermeidlich:

- Wir finden mehr als 13 defekte Knallkörper in der Stichprobe, obwohl die *Sendung in Ordnung* ist (Ausschussanteil 4%). Dann *lehnen wir sie zu Unrecht ab (Fehler 1. Art) und zwar mit der Wahrscheinlichkeit  $\alpha$ .*
- Wir finden höchstens 13 defekte Knallkörper in der Stichprobe, obwohl die *Sendung nicht* in Ordnung ist (mehr als 4% Ausschuss). Wir *akzeptieren sie zu Unrecht (Fehler 2. Art) und zwar mit der Wahrscheinlichkeit  $\beta$ .*

Die Darstellung der Situation im Baumdiagramm ist instruktiv (Fig. 1.4):

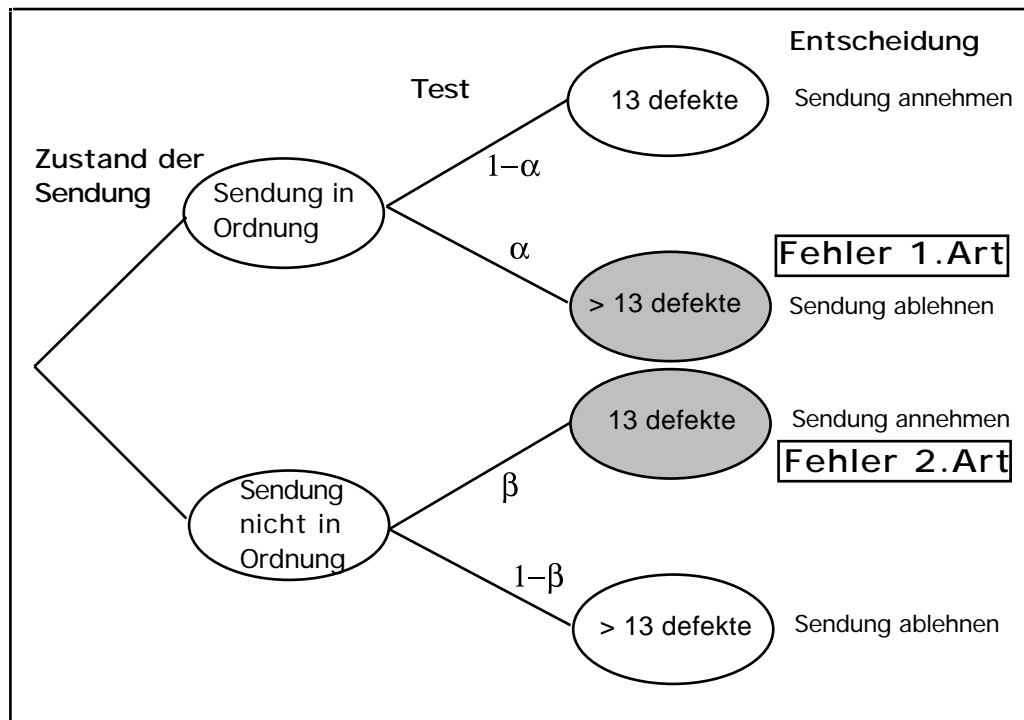


Fig. 1.4: Die Fehler 1. Art und 2. Art und ihre Wahrscheinlichkeiten  $\alpha$  und  $\beta$  im Baumdiagramm

### Beurteilende Statistik

Der Qualitätstest ist ein typisches Beispiel aus der sog. *beurteilenden* Statistik: Eine Grundgesamtheit soll aufgrund einer Stichprobe, d. h. aufgrund einer *zufälligen* Auswahl, beurteilt werden. In unserem Fall soll von der (relativen) Häufigkeit der schwarzen Kugeln in der Stichprobe auf diejenige in der Grundgesamtheit geschlossen werden (Fig. 1.5).

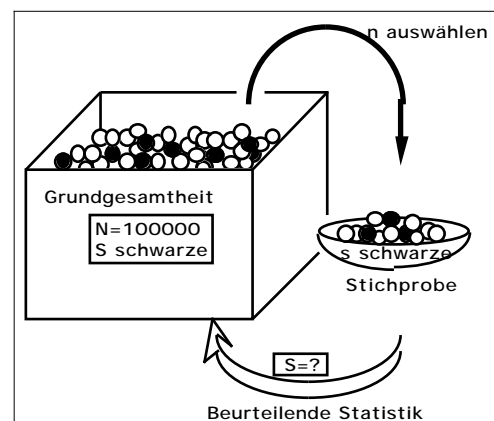


Fig. 1.5: Die beurteilende Statistik im Urnenmodell

Es ist unvermeidlich, dass der Schluss mit einer gewissen Wahrscheinlichkeit falsch ist. Wir werden diese Wahrscheinlichkeit in den nächsten Kapiteln auf zwei ganz verschiedene Arten ermitteln: 1. Durch Simulation des Qualitätstests auf dem Computer und 2. mit Hilfe der Wahrscheinlichkeitsrechnung.

### Zusatzinformation:

Intuitiv ist Ihnen sicher klar, dass ein Qualitätstest um so zuverlässiger ist, je grösser der Umfang  $n$  der Stichprobe ist. Aber je grösser  $n$  ist, um so grösser sind die Testkosten. Eine weitere Aufgabe der Statistik ist es,  $n$  so klein als möglich und so gross wie nötig festzulegen. Dieses wichtige und interessante Problem steht aber nicht im Zentrum dieses Leitprogramms.

**Haben Sie die Lernziele erreicht?**

Sie sind bereits am Ende des ersten Kapitels angelangt. In der folgenden *Lernkontrolle* geht es um eine neue statistische Fragestellung. Sie sollen versuchen, hier die Bedeutung der eben eingeführten Begriffe und Variablen zu erkennen.

Wenn Ihnen die Aufgabe Schwierigkeiten bereitet, ist es angezeigt, den vorangehenden Abschnitt nochmals durcharbeiten. Schlagen Sie nicht zu früh im Lösungsteil auf der nächsten Seite nach. Wenn Sie aber das Gefühl haben, dass Sie die Sache beherrschen, so melden Sie sich gleich anschliessend bei Ihrer Lehrperson zum Kapiteltest.

**Lernkontrolle: Die theoretische Fahrprüfung**

KandidatInnen für die Autofahrprüfung müssen sich u. a. auch über ihre theoretischen Kenntnisse der Verkehrsregeln ausweisen. Das Strassenverkehrsamt möchte wissen, ob sie sich in wenigstens 475 von 500 Standardverkehrssituationen (theoretisch) richtig verhalten. Dazu werden ihnen im Rahmen eines Tests 50 zufällig ausgewählte Verkehrssituationen präsentiert und sie müssen aus verschiedenen angebotenen Reaktionen die richtige auswählen (multiple choice).

Der Test gilt als bestanden, wenn wenigstens 45 Fragen richtig beantwortet werden.

☐ a

Weshalb wohl legt das Strassenverkehrsamt den KandidatInnen nur 50 und nicht alle 500 Verkehrssituationen vor?

☐ b

Man kann das Vorgehen des Strassenverkehrsamtes als einen Qualitätstest ansehen. Beschreiben Sie für dieses Beispiel die Grundgesamtheit und die Stichprobe.

☐ c

Zeichnen Sie ein Urnenmodell. Welche Rolle spielen die schwarzen Kugeln? Welche Bedeutung und/oder Werte haben die Variablen  $N$ ,  $S$ ,  $n$ ,  $s$ ?

☐ d

Nach der Prüfung wird, wie gesagt, entschieden, ob der Test bestanden wurde. Dabei kann es zu einem Fehler 1. Art kommen. Worin besteht er?

☐ e

Worin besteht der mögliche Fehler 2. Art?

☐ f

Welchen der beiden Fehler erachten Sie als problematischer? Weshalb?

## Lösungen der Aufgaben zu Kapitel 1

### Lösung Aufgabe 1.1

Das Abzählen der schwarzen Kugeln ist nicht möglich, weil man ja in der Realität erst durch das Abbrennen des der Kugel entsprechenden Knallkörpers weiss, ob die Kugel schwarz ist.

### Lösung der Lernkontrolle

a

U. a. aus Zeitgründen ist es nicht möglich, den KandidatInnen alle 500 Fragen vorzulegen.

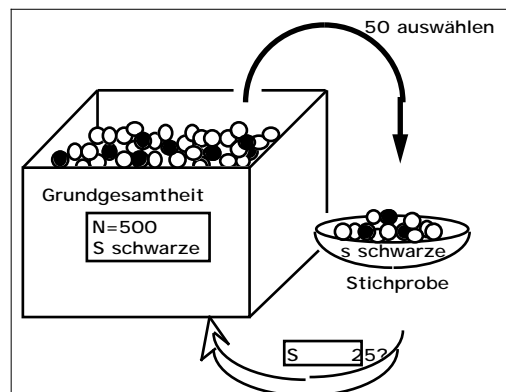
b

Die 500 Verkehrssituationen bilden die Grundgesamtheit. Die 50 ausgewählten Testsituationen bilden die Stichprobe.

c

Jeder der 500 Verkehrssituationen ordnen wir eine Kugel zu:  $N=500$ .

Denjenigen Verkehrssituationen, in denen sich der (die) Kandidat(in) in der momentanen Testsituation nicht auskennt, ordnen wir schwarze Kugeln zu. An ihrer Anzahl  $S$  ist das Strassenverkehrsamt interessiert; es sollte  $S = 25$  sein.



Die 50 für den Test ausgewählten Verkehrssituationen bilden, wie gesagt, die Stichprobe; der Stichprobenumfang ist also  $n=50$ .

Den falsch gelösten Fragen entsprechen die schwarzen Kugeln in der Stichprobe; ihre Anzahl  $s$  sollte kleiner oder gleich 5 sein. (Andernfalls ist der Test nicht bestanden.)

d

Der Fehler 1. Art im *Urnenmodell*: Auch wenn die Urne unter 500 höchstens 25 schwarze Kugeln enthält, können mehr als 5 schwarze in die Stichprobe gelangen.

Der Fehler 1. Art in der *Realität*: Wenn einem Kandidaten im Test beispielsweise ausgerechnet jene 10 Verkehrssituationen, die er nicht beherrscht, präsentiert werden, wird er den Test nicht bestehen, obwohl er eigentlich die Anforderungen erfüllt.

e

Der Fehler 2. Art im *Urnenmodell*: Auch wenn die Urne mehr als 25 schwarze Kugeln enthält, können weniger als 5 schwarze in die Stichprobe gelangen.

Der Fehler 2. Art in der *Realität*: Auch KandidatInnen, die insgesamt bei mehr als 25 Aufgaben eine falsche Antwort gäben, können die Prüfung bestehen, weil durch Zufall nur wenige davon für die Prüfung ausgewählt wurden. Im Extremfall können KandidatInnen, welche die Verkehrsregeln praktisch nicht kennen, die Prüfung bestehen.

Bei den obigen Überlegungen haben wir bewusst ausser acht gelassen, dass es für jeden Kandidaten mehr als zwei Kategorien von Fragen gibt, nämlich:

- Solche, bei denen der Kandidat die richtige Antwort weiss.
- Solche, bei denen der Kandidat glaubt, die richtige Antwort zu kennen, aber sie ist falsch.
- Solche, bei denen einfach geraten wird.

f

Der Fehler 2. Art kann sich hier viel verhängnisvoller auswirken, weil VerkehrsteilnehmerInnen ans Steuer gelassen werden, obwohl sie sich in den Verkehrsregeln nicht auskennen. Der Fehler 1. Art ist für die Betroffenen zwar schmerzlich, aber sie werden ihren Test bestimmt das nächste Mal bestehen....

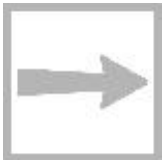
## 2

**Computersimulation des Qualitätstests**

In diesem Kapitel erreichen wir ein erstes entscheidendes Ziel: Wir ermitteln *näherungsweise* die *Wahrscheinlichkeiten*  $\alpha$  und  $\beta$  für die Fehler 1. und 2. Art – und zwar ohne eigentliche Rechnung!

Wie ist das möglich? Sie erinnern sich, dass die Wahrscheinlichkeit eines zufälligen Ereignisses die *idealisierte relative Häufigkeit* ist, mit welcher dieses vorkommt, wenn man das zugehörige Zufallsexperiment sehr oft wiederholt. Wenn wir beispielsweise zwei ideale Würfel sehr oft gleichzeitig werfen, so werden wir durchschnittlich unter je 18 Würfeln einmal die Summe 3 erhalten. Deshalb legen wir fest:

$$P(\text{„Summe 3“}) = \frac{1}{18} = 0.0555 = 5.55\%.$$



Mit dieser Idee können wir näherungsweise auch die Wahrscheinlichkeit für den Fehler 1. Art bestimmen. Wir betrachten eine Urne mit 100'000 Kugeln, von denen exakt 4000 schwarz sind. Nun simulieren wir den Qualitätstest hundertmal, indem wir hundertmal je 200 Kugeln aus der vollen Urne auswählen. Jedesmal prüfen wir nach, ob die Stichprobe mehr als 13 schwarze enthält, was ja einem Fehler 1. Art entsprechen würde. Wenn dies in z. B. vier der hundert Ziehungen vorkommt, hat der Fehler 1. Art die relative Häufigkeit 4% und wir würden 4% schätzen.

Bitte erschrecken Sie nicht vor dieser Aufgabe! Selbstverständlich werden wir nicht 100 mal unter 100'000 *realen* Kugeln 200 auswählen und die schwarzen zählen, sondern wir simulieren diesen Vorgang praktisch gleichwertig auf dem Computer.

**Der Stellenwert der Simulation**

Denken Sie nicht, dass es sich bei dieser spielerischen Methode um ein unseriöses oder unwissenschaftliches Verfahren handelt. Im Gegenteil: Immer mehr Statistikanwender entschliessen sich, ihr statistisches Problem nicht mit Hilfe einer abstrakten (Wahrscheinlichkeits-) Rechnung zu durchleuchten, sondern mittels der viel anschaulicheren Simulation. Ermöglicht wird dies durch die immensen Speicher- und Rechenkapazitäten der allgegenwärtigen Computer. Und man kann theoretisch zeigen (sog. Gesetz der grossen Zahl), dass die simulierten Resultate verlässlich sind, ja, man kann sogar die Grösse des Fehlers abschätzen (zentraler Grenzwertsatz). Bei komplizierten Problemen ist die Simulationsmethode manchmal sogar die einzige Möglichkeit, ein Resultat zu erhalten.

**Aufgabe 2.0.1**

*Eine Repetitionsaufgabe, die Sie wahrscheinlich früher schon einmal gelöst haben: Mit welcher Wahrscheinlichkeit würfelt man mit zwei idealen Würfeln die Summe 7?*





### Aufgabe 2.0.2

Wie gross schätzen Sie bei unserem Knallkörper–Qualitätstest  $\alpha$ ?

☐ 1‰

☐ 1%

☐ 2%

☐ 5%

## 2.1

### Simulation mit Ikosaedern

Wir wollen, wie gesagt,  $\alpha$  durch Simulation ermitteln. Damit wir die Simulation auf dem Computer besser verstehen, führen wir zuvor eine Simulation „von Hand“ durch. Um die Übersicht zu behalten, beschränken wir uns dabei auf 10-er Stichproben.

Als erstes wollen wir erreichen, dass jede einzelne Kugel identifizierbar und unterscheidbar von den anderen ist. Dazu denken wir uns die 100'000 Kugeln numeriert mit den Nummern

0, 1, 2, ..., 99998, 99999.

Am übersichtlichsten wird es, wenn wir den 4000 schwarzen Kugeln gerade die Nummern 0, 1, 2, ..., 3998, 3999 zuordnen.

#### Zufällig eine Kugel ziehen

Zufällig eine Kugel wählen ist nun gleichbedeutend mit zufällig eine der obgenannten Nummern würfeln. Am einfachsten geht das mit dem Ikosaeder, dem regulären Zwanzigflächner (Fig. 2.1), aus welchem wir wie folgt einen „Zehnerwürfel“ konstruieren: Wir malen auf ein erstes beliebig gewähltes gleichseitiges Dreieck eine 0 und gleich anschliessend eine 0 auch auf das diametral gegenüberliegende Dreieck. In analoger Weise beschriften wir je Paare gegenüberliegender Seitenflächen mit den Ziffern 1, 2, ..., 9 (man kann solche Würfel sogar fixfertig in Spielwarengeschäften kaufen).

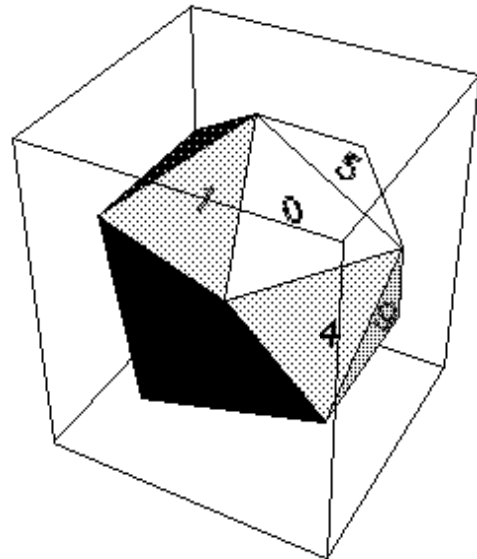


Fig. 2.1: Das mit den Ziffern 0, 1, ..., 9 beschriftete Ikosaeder.

Bei jedem Wurf eines solchen idealen Ikosaeders erscheint mit gleicher Wahrscheinlichkeit eine der zehn Ziffern auf der zuoberst liegenden Seitenfläche. Werfen wir den Ikosaeder fünfmal nacheinander und notieren die gewürfelten Ziffern hintereinander, so erhalten wir gerade eine der obgenannten 100'000 Zahlen, z. B. 01033.

Die Kugel mit der Nummer 1033 wurde gewürfelt; sie ist zufällig gerade schwarz.

### **Zufällig eine Stichprobe ziehen**

Diese umständliche Prozedur wurde 10-mal durchgeführt und so ergab sich die folgende 10-er Stichprobe:

1033 - 45726 - 93952 - 14152 - 3070 - 64939 - 81283 - 92019 - 30218 - 90733

Bei jeder Nummer wurde überprüft, ob sie nicht vorher schon einmal vorgekommen ist. Man hätte sie allenfalls einfach weggelassen.

Nur nebenbei: Die zehn gewürfelten Zahlen sind mit der Wahrscheinlichkeit 0.99945 verschieden, d. h. die gleiche Nummer wird sehr selten mehr als einmal gewürfelt.

### **Das 0-1-Protokoll**

Unsere Stichprobe enthält zwei schwarze Kugeln, nämlich die Nummern 1033 und 3070. Da uns eigentlich jeweils nur interessiert, ob die gezogene Kugel schwarz oder weiss ist, genügt es, das sog. 0-1-Protokoll der Stichprobe zu notieren:

1	0	0	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---

Dabei haben wir schwarze Kugeln durch eine 1, weisse durch eine 0 gekennzeichnet. Diese Darstellung hat den Vorteil, dass wir die Anzahl der schwarzen Kugeln einfach als Summe aller 10 gewürfelten 0-er und 1-er erhalten.

### **25 Simulationen einer 10-er Stichprobe**

In der folgenden Tabelle sind 25 simulierte 10-er Stichproben zusammengestellt und in der letzten Spalte wurde je die Anzahl schwarzer Kugeln notiert.

	1	2	3	4	5	6	7	8	9	10	Anzahl schwarze
Simulation 1	0	0	0	0	0	0	0	1	0	0	1
Simulation 2	0	0	0	0	0	0	1	0	0	0	1
Simulation 3	0	0	0	0	1	0	0	0	0	0	1
Simulation 4	0	0	0	0	0	0	0	0	0	0	0
Simulation 5	0	0	0	0	0	0	0	0	0	0	0
Simulation 6	0	0	0	0	0	0	0	0	0	0	0
Simulation 7	0	0	0	0	0	0	0	0	0	0	0
Simulation 8	0	0	0	0	0	0	0	0	0	0	0
Simulation 9	0	0	0	0	0	0	0	1	0	0	1
Simulation 10	0	0	0	0	0	0	0	0	0	0	0
Simulation 11	0	0	1	0	0	0	0	0	0	0	1
Simulation 12	1	0	0	0	0	0	0	0	0	0	1
Simulation 13	0	0	0	0	0	0	0	0	0	0	0

Simulation 14	0	1	0	0	0	0	0	0	0	0	1
Simulation 15	0	0	0	0	0	0	0	0	0	0	0
Simulation 16	0	0	0	0	0	0	1	1	0	0	2
Simulation 17	0	0	0	0	0	0	0	0	0	0	0
Simulation 18	0	0	0	0	0	0	0	0	0	0	0
Simulation 19	0	0	0	0	0	0	0	0	0	0	0
Simulation 20	1	0	1	0	0	0	0	0	0	0	2
Simulation 21	0	0	0	0	0	0	0	0	0	0	0
Simulation 22	0	0	0	0	0	0	0	0	0	0	0
Simulation 23	0	0	0	1	0	0	0	0	0	0	1
Simulation 24	0	0	0	0	0	0	0	0	0	1	1
Simulation 25	0	0	0	0	0	0	0	0	0	0	0

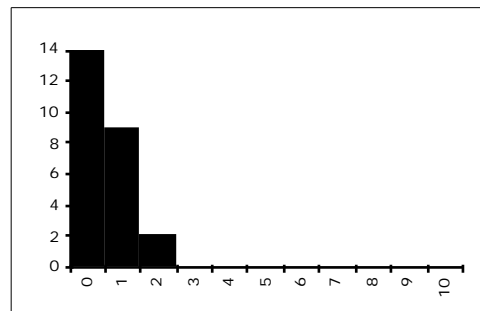
### Absolute Häufigkeiten

Theoretisch kann eine 10-er Stichprobe 0 bis 10 schwarze Kugeln enthalten. In den obigen 25 Simulationen sind aber nur die Werte 0, 1 und 2 vorgekommen und zwar mit den in der nachstehenden Tabelle zusammengestellten absoluten Häufigkeiten:

Anzahl schwarze Kugeln	0	1	2	3	4	5	6	7	8	9	10
Anzahl Stichproben	14	9	2	-	-	-	-	-	-	-	-

### Das Histogramm

Die nebenstehende Figur 2.2 zeigt das Histogramm (Säulendiagramm) dieser Häufigkeitstabelle. Auf der waagrechten Achse ist die mögliche Anzahl der schwarzen Kugeln aufgetragen und auf der senkrechten die zugehörigen absoluten Häufigkeiten.



Das Histogramm ist anschaulicher als die Häufigkeitstabelle. Wir sehen z. B. auf einen Blick, dass in den meisten Fällen keine oder genau eine schwarze Kugel gezogen wurde.

Fig. 2.2: Histogramm der absoluten Häufigkeiten von Stichproben mit 0, 1, ..., 10 schwarzen Kugeln.

### Relative Häufigkeiten

Aus den absoluten können wir die *relativen* Häufigkeiten berechnen. Das Ereignis „Keine schwarze Kugel“ hat z. B. die relative Häufigkeit:

$$\frac{\text{absolute Häufigkeit}}{\text{Anzahl Simulationen}} = \frac{14}{25} = 0.56 = 56\%.$$

Analog ergeben sich die relativen Häufigkeiten der anderen Ereignisse.



### Aufgabe 2.1.1

Berechnen Sie die restlichen relativen Häufigkeiten und ergänzen Sie die nachfolgende Tabelle.

Anzahl defekte	0	1	2	3	4	5	6	7	8	9	10
Relative Häufigkeit (%)	56										

### Wahrscheinlichkeiten

Gute und zuverlässige Schätzwerte für die Wahrscheinlichkeiten erhalten wir, wenn wir das zugehörige Zufallsexperiment sehr oft wiederholen. Obwohl 25 in der Statistik eine eher kleine Zahl ist, wollen wir im vorliegenden Fall doch annehmen, dass das Ereignis

„Keine schwarze Kugel in einer Stichprobe von 10 Kugeln“

ungefähr die Wahrscheinlichkeit 56% hat:

$$P(\text{„Keine schwarze Kugel“}) = 56\%$$

Analoges gilt für die Ereignisse „Genau 1 schwarze“, „Genau 2 schwarze“ etc.



### Aufgabe 2.1.2

Stellen Sie sich vor, dass sich Einkäufer und Lieferant den folgenden Testplan vereinbart hätten:

Es werden zufällig 10 Knallkörper ausgewählt und gezündet. Wenn diese Stichprobe mehr als einen defekten Knallkörper enthält, wird die Sendung nicht akzeptiert.

- ☐ a) Worin besteht der Fehler 1. Art bei diesem Test?
- ☐ b) Wie oft wird der Fehler 1. Art in den obigen 25 Simulationen gemacht?
- ☐ c) Wie gross ist etwa die Wahrscheinlichkeit  $\alpha$  für den Fehler 1. Art bei diesem einfachen Qualitätstest?

## 2.2

### Die Fehler 1. und 2. Art simuliert

Sie haben im letzten Abschnitt erfahren, wie *im Prinzip* die Wahrscheinlichkeit für die Fehler 1. Art über Simulation ermittelt werden kann. Bei einer Computersimulation verwendet man anstelle des Ikosaeders den Zufallsgenerator eines Computers. Das ist ein Programm, das Zahlen erzeugt, die man von "echten" Zufallszahlen praktisch nicht unterscheiden kann. In diesem Abschnitt präsentieren wir Ihnen nun in Form von EXCEL-Tabellen und Histogrammen die Auswertung von hundert Computersimulationen einer 200-er Stichprobe. Es ist Ihnen wohl klar, dass jede neue Serie von Simulationen leicht verschiedene Resultate liefert.

### Häufigkeitstabelle für den Fehler 1. Art

Hier rechnet die Simulation mit genau 4000 schwarzen unter den 100'000 Kugeln in der Urne. Die nachstehende Tabelle ist wie diejenige auf Seite 2.4 aufgebaut: In der 2. Zeile stehen die absoluten Häufigkeiten von Stichproben mit  $s$  schwarzen Kugeln ( $s=0, 1, \dots, 20$ ; diese Werte stehen in der 1. Zeile). Beispielsweise enthielten 18 der 100 simulierten Stichproben 8 schwarze Kugeln.

Tabelle3																						
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	Y
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2	0	0	1	0	6	8	18	14	18	6	8	10	2	6	2	1	0	0	0	0	0	
3																						
4																						
5																						

Fig. 2.2: Häufigkeitstabelle für die Anzahl schwarzer Kugeln in einer 200-er Stichprobe.



### Aufgabe 2.2.1

Wie viele der 100 Stichproben enthielten mehr als 10 schwarze Kugeln?



### Wahrscheinlichkeit des Fehlers 1. Art

Der Fehler 1. Art ergibt sich jedesmal, wenn eine Stichprobe mehr als 13 schwarze Kugeln enthält. Im ganzen gibt es gemäss Häufigkeitstabelle nur 3 Stichproben mit mehr als 13 schwarzen Kugeln, nämlich zwei mit 14 und eine mit 15. Der Fehler 1. Art hat bei dieser Simulation also die relative Häufigkeit von 3%. Wir dürfen diesen Wert als Schätzwert für die Wahrscheinlichkeit des Fehlers 1. Art betrachten:

3%

### Was bedeutet $\alpha = 3\%$ für die Praxis?

Für den Qualitätstest bei den Knallkörpern bedeutet 3%: Ein Einkäufer, der diesen oder einen ähnlichen Einkauf häufig macht, wird in durchschnittlich 3 von 100 Einkäufen die Knallkörpersendung irrtümlicherweise ablehnen!

### Der Fehler 1. Art im Histogramm

Im nebenstehenden Histogramm der Häufigkeitstabelle (Fig. 2.3) sind die Stichproben, die zu einem Fehler 1. Art führen, durch punktierte Säulen unterschieden.

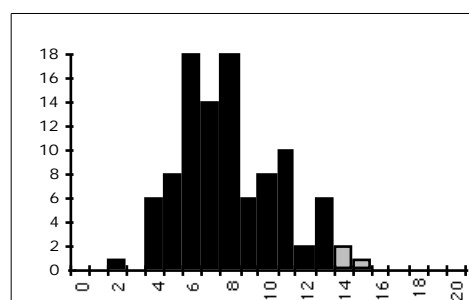


Fig. 2.3: Histogramm der Häufigkeit schwarzer Kugeln in Hundert 200-er Stichproben.

Damit haben wir auch schon das Ende dieses Kapitels erreicht. Wenn Sie die folgende Lernkontrolle zufriedenstellend gelöst haben, sind Sie reif für den Kapiteltest. Zunächst jedoch noch ein Hinweis:

### Die Technik der Computersimulation

Wenn Sie erfahren möchten, wie eine Computersimulation funktioniert und wie man sie programmiert, müssen Sie im Anhang S "Simulationen" das Kap. I studieren. Ihre Lehrperson hat Sie orientiert, ob und wie weit Sie sich in diese technische Seite des Problems einarbeiten sollen.



### Lernkontrolle: Der Fehler 2. Art

Wir stellen uns die Frage: Wie gross ist die Wahrscheinlichkeit für den Fehler 2. Art, wenn die Urne statt 4000 genau doppelt so viele schwarze Kugeln enthält?

Hier zunächst das Ergebnis von 100 Simulationen in Form einer EXCEL-Tabelle:

	Tabelle3																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	0	0	0	0	0	0	0	1	1	2	6	4	5	6	9	12	16	7	10	8
3																				
4																				

a

Die obige Tabelle ist auf der rechten Seite unvollständig! Ergänzen Sie diese im folgenden Sinne: Wie viele der 100 Stichproben enthielten mehr als 20 schwarze Kugeln?

b

Bei welchen, bzw. bei wie vielen Stichproben macht man den Fehler 2. Art?

c

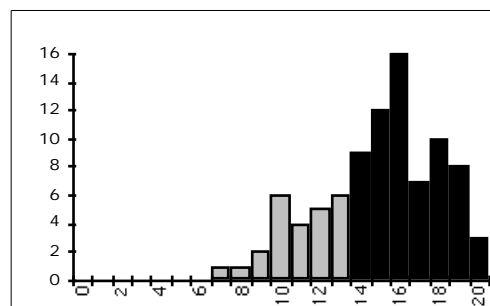
Wie gross ist demnach ungefähr die Wahrscheinlichkeit  $\beta$  für den Fehler 2. Art?

d

Was sagen Sie zum erhaltenen Wert für  $\beta$ ? Machen Sie Vorschläge, um eine Verbesserung zu erreichen.

e

Die nebenstehende Figur zeigt das (ebenfalls rechts angeschnittene) Histogramm der obigen Häufigkeitstabelle. Interpretieren Sie die Darstellung.



## Lösungen der Aufgaben zu Kapitel 2

### Lösung Aufgabe 2.0.1

Es gilt:

$$7=1+6=2+5=3+4=4+3=5+2=6+1.$$

Unter den 36 verschiedenen Wurfresultaten (die beiden Würfel müssen durch irgend ein Merkmal, z. B. durch ihre Farbe, von einander unterscheidbar sein), ergeben also 6 die Summe 7. Weil alle Wurfresultate gleich wahrscheinlich sind, gilt:

$$P(\text{„Summe 7“}) = \frac{6}{36} = \frac{1}{6} = 16.7\%.$$

### Lösung Aufgabe 2.0.2

Lassen Sie sich von den Simulationen im folgenden Abschnitt überraschen.

### Lösung Aufgabe 2.1.1

Anzahl schwarze Kugeln	0	1	2	3	4	5	6	7	8	9	10
Relative Häufigkeit (%)	56	36	8	0	0	0	0	0	0	0	0

### Lösung Aufgabe 2.1.2

a

Bei diesem Test besteht der Fehler 1. Art darin, dass die Sendung zurückgewiesen wird, obwohl sie in Ordnung ist. Gleichbedeutend damit ist: Es hat mehr als einen defekten Knallkörper in der 10-er Stichprobe, obwohl die ganze Sendung höchstens 4000 defekte Stücke enthält.

b

Wir haben 25 10-er Stichproben simuliert und zwar bei einer Urne, die unter 100'000 genau 4000 schwarze Kugeln enthält. Zwei dieser 25 Stichproben enthalten mehr als eine schwarze Kugel und führen deshalb auf einen Fehler 1. Art.

c

Das Ereignis "Fehler 1. Art" hat also die *relative Häufigkeit*:

$$\frac{2}{25} = 0.08 = 8\%.$$

Als Schätzwert der Wahrscheinlichkeit für den Fehler 1. Art erhalten wir somit 8%.

**Lösung Aufgabe 2.2.1**

Aus der Tabelle kann entnommen werden, dass

$$1 + 16 + 18 + 14 + 18 + 6 + 8 = 79$$

der 100 Stichproben 10 oder weniger schwarze Kugeln enthielten. Folglich enthalten 21 Stichproben mehr als 10 schwarze. Eine *direkte* Zählung bestätigt dieses Ergebnis:

$$10 + 2 + 6 + 2 + 1 = 21.$$

---

**Lösung der Lernkontrolle**

a

Die Summe aller Häufigkeiten in der Tabelle beträgt 90. Folglich enthielten 10 Stichproben mehr als 20 schwarze Kugeln.

b

Den Fehler 2. Art begeht man jedesmal dann, wenn höchstens 13 schwarze Kugeln in die Stichprobe geraten (obwohl die Urne 8000 schwarze Kugeln enthält). Gemäss der Häufigkeitstabelle enthalten 25 unserer 100 simulierten 200-er Stichproben weniger als 14 schwarze Kugeln.

c

Der Fehler 2. Art hat also bei dieser Simulation die relative Häufigkeit 25%. Für seine Wahrscheinlichkeit ergibt sich folglich der Schätzwert:

25%
-----

d

ist unerwartet gross. Man könnte verkleinern (um 6%), wenn man die Sendung schon ab 13 defekten Knallkörpern zurückweisen würde. Dabei würde allerdings grösser, nämlich aufgrund der Häufigkeitstabelle 9%.

Wenn die Werte von und so nicht akzeptierbar sind, muss der Stichprobenumfang vergrössert werden.

e

Wir haben im Histogramm die Stichproben, die zu einem Fehler 2. Art führen, als punktierte Säulen hervorgehoben.

---



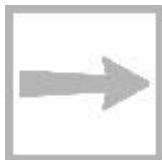
## 3

## Grundlagen zur Wahrscheinlichkeitsrechnung

Durch die Simulation haben wir eine gute Vorstellung von der Bedeutung und der Grösse von  $\hat{p}$  resp.  $\hat{p}_n$  erhalten. In diesem Kapitel wollen wir die Grundlagen erarbeiten, um  $\hat{p}$  und  $\hat{p}_n$  im nächsten Kapitel formelmässig *berechnen* zu können. Das kann sich als sehr hilfreich erweisen: Weiss man beispielsweise, wie  $\hat{p}$  von  $N$ ,  $S$  und  $n$  abhängt, so kann man evtl. ausrechnen, wie gross der Stichprobenumfang  $n$  gewählt werden muss, damit  $\hat{p}$  einen gewissen gewünschten Wert nicht überschreitet.

Wir setzen hier voraus, dass Sie die wichtigsten Grundbegriffe aus der Kombinatorik und der Wahrscheinlichkeitsrechnung mitbringen. Trotzdem ist das Kapitel sehr ausführlich gehalten.

Wir können  $\hat{p}$  und  $\hat{p}_n$  berechnen, sobald wir die folgende Grundaufgabe gelöst haben.



### Die Grundaufgabe

Eine Urne enthält  $N=100'000$  Kugeln, davon sind  $S=4000$  schwarz. Man entnimmt der Urne mit irgendeinem Zufallsmechanismus – denken Sie an die Auswahlmaschinen beim Zahlenlotto –  $n=200$  Kugeln (Fig. 3.1).

Gesucht ist die Wahrscheinlichkeit dafür, dass die Stichprobe  $s$  schwarze Kugeln enthält. (Dabei ist  $s$  eine bestimmte Zahl mit  $0 \leq s \leq n$ .)

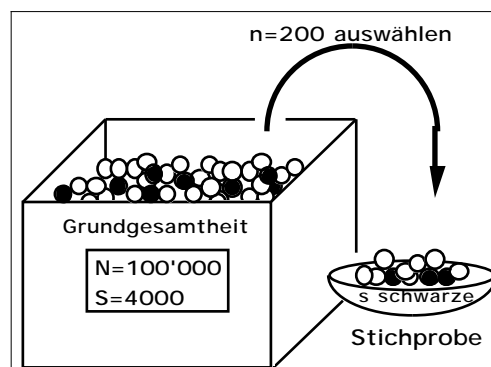


Fig. 3.1: Urnenmodell für die Ziehung von 200 Kugeln

### Die Kugelnummern

Wie schon im letzten Kapitel nehmen wir an, die 100'000 Kugeln seien beginnend mit 0 durchnummeriert, wobei die Nummern 0, 1, ..., 3999 für die schwarzen Kugeln reserviert sind. Mit Hilfe dieses Tricks erreichen wir, dass die Kugeln künstlich unterscheidbar werden, was die nachfolgenden Erklärungen hoffentlich vereinfacht.

### Stichproben und Teilmengen

Wenn wir zur schwarzen Kugel mit der Nr. 0 weitere 199 weisse Kugeln aussuchen, entsteht eine „Stichprobe mit 1 schwarzen Kugel“. Wenn wir die gleichen 199 weissen Kugeln mit der schwarzen Kugel Nr. 1 kombinieren, haben wir wieder eine „Stichprobe mit genau einer schwarzen Kugel“. Das Ereignis „Stichprobe mit 1 schwarzen Kugel“ kann somit allein hinsichtlich der schwarzen Kugel auf 4000 verschiedene Arten zustande kommen.

Etwas mathematischer können wir sagen: Es gibt viele verschiedene *Teilmengen* mit 200 Kugeln, die auf das Ereignis „Stichprobe mit 1 schwarzen Kugel“ führen. Wir bezeichnen solche Teilmengen als *günstig* für das Ereignis. Für die *Anzahl* der *günstigen* Teilmengen verwenden wir das Symbol  $g_1$ .

Alles Gesagte gilt analog für Stichproben mit  $s$  schwarzen Kugeln. Insbesondere verstehen wir unter  $g_s$  die Anzahl günstiger Teilmengen für eine „Stichprobe mit  $s$  schwarzen Kugeln“.

### Unterschied zwischen Stichprobe und Teilmenge



Vielleicht wird der Unterschied zwischen Stichprobe und Teilmenge noch durch die folgende Bemerkung verdeutlicht: Zwei *Teilmengen* mit je 200 Kugeln betrachten wir dann als verschieden, wenn es in der einen Teilmenge wenigstens eine Kugelnummer gibt, die in der anderen Teilmenge nicht vorkommt. Dagegen betrachten wir zwei *Stichproben* als verschieden, wenn sie sich in der *Anzahl der schwarzen Kugeln* unterscheiden; es gibt mit anderen Worten in unserer Betrachtungsweise nur 201 verschiedene Stichproben, nämlich solche mit 0, 1, 2, ..., 200 schwarzen Kugeln.



#### Aufgabe 3.1

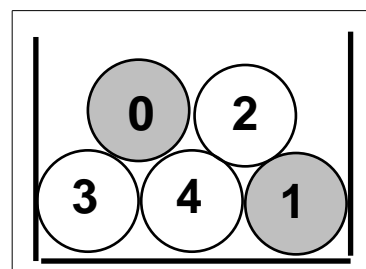
Eine Urne enthält 2 schwarze und 3 weiße (numerierte) Kugeln.

a

Bilden Sie alle möglichen Teilmengen mit 3 Kugeln.

b

Welche bzw. wie viele der Teilmengen führen auf Stichproben mit 0, 1, 2, 3 schwarzen Kugeln?



### Alle Teilmengen mit 200 Elementen sind gleich wahrscheinlich

Für die weiteren Berechnungen ist nun entscheidend, dass wir postulieren: Alle möglichen Teilmengen mit 200 Kugeln haben die gleiche Wahrscheinlichkeit, ausgewählt zu werden.

Beachten Sie, dass dies für *Stichproben* im oben definierten Sinne nicht zutrifft: Eine Stichprobe mit 200 schwarzen Kugeln hat praktisch die Wahrscheinlichkeit 0, Stichproben mit ungefähr 10 schwarzen Kugeln sind jedoch recht häufig.

Die Wahrscheinlichkeit einer Stichprobe ist um so grösser, je grösser die zugehörige Anzahl der günstigen Teilmengen ist. In der Wahrscheinlichkeitsrechnung haben Sie vermutlich die folgende Formel kennen gelernt:



$$P(\text{„s schwarze Kugeln“}) = \frac{\text{Anzahl günstiger Teilmengen}}{\text{Anzahl möglicher Teilmengen}} = \frac{g_s}{m}$$

Hierbei bedeuten:

- $g_s$  = Anzahl der (günstigen) Teilmengen mit  $n=200$  Kugeln, von denen  $s$  schwarz sind ( $s$  ist eine bestimmte Zahl mit  $0 \leq s \leq 200$ ).
- $m$  ist die Anzahl der (möglichen) Teilmengen mit  $n=200$  Kugeln.



### Aufgabe 3.2

Eine Urne enthält 2 schwarze und 3 weisse Kugeln (siehe Aufgabe 3.1). Es wird eine Stichprobe vom Umfang 3 gezogen. Mit welchen Wahrscheinlichkeiten enthält sie 0, 1, 2 oder 3 schwarze Kugeln?

### $m$ = Anzahl der möglichen Teilmengen

Wiederholen wir das Ergebnis der Aufgabe 3.1. Mit 5 Kugeln, die mit 0, 1, 2, 3, 4 numeriert sind, kann man 10 verschiedene Teilmengen mit 3 Kugeln bilden, nämlich:

$$\begin{array}{ccccc} \{0, 1, 2\} & \{0, 1, 3\} & \{0, 1, 4\} & \{0, 2, 3\} & \{0, 2, 4\} \\ \{0, 3, 4\} & \{1, 2, 3\} & \{1, 2, 4\} & \{1, 3, 4\} & \{2, 3, 4\}. \end{array}$$

In der Kombinatorik wird gezeigt, dass diese Anzahl gleich  $\binom{5}{3}$  ist. Wie Sie sich wohl erinnern, berechnet man diesen sog. *Binomialkoeffizienten* mit einem der folgenden beiden Brüche:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 10$$

Konsultieren Sie evtl. Ihre Formelsammlung! (Berechnungsbeispiele folgen!)

Analog gilt: Mit  $N=100'000$  (unterscheidbaren) Kugeln kann man

$$m = \binom{N}{n} = \frac{100'000}{200}$$

verschiedene Teilmengen mit  $n=200$  Kugeln bilden.

### Binomialkoeffizienten auf dem Taschenrechner

Die nachstehende Fig. 3.2 zeigt, wie man diesen Binomialkoeffizienten auf einem modernen Taschenrechner berechnet ( $nCr$  bedeutet hier "number of combinations", deutsch Binomialkoeffizient. Um die Eingabe von Grossbuchstaben zu umgehen, steht  $nn$  anstelle von  $N$ ).

Fig. 3.2 : Die Berechnung eines Binomialkoeffizienten, wie sie auf einem modernen Taschen-computer aussehen mag.

```
100000→nn
200→n
nCr(nn,n)      1.03903E625
```

Das Ergebnis  $m=1.04 \cdot 10^{625}$  ist eine unvorstellbar grosse Zahl. Vergleiche: Die Anzahl der Elektronen im ganzen Weltall ist deutlich kleiner als  $10^{80}$ .



### Aufgabe 3.3

Eine Urne enthält 5 (numerierte) Kugeln. Wie viele verschiedene Teilmengen mit 3 Kugeln kann man aufgrund der eben angegebenen Formel bilden? Vergleichen Sie Ihr Ergebnis mit demjenigen in Aufgabe 3.1.

#### $g_s$ = Anzahl der günstigen Teilmengen

Zu den günstigen Teilmengen gehören solche mit genau  $s$  schwarzen und  $(n-s)$  weissen Kugeln.

Ihre Anzahl ergibt sich aufgrund der folgenden *Idee*: Wir kombinieren alle möglichen Mengen mit  $s$  schwarzen Kugeln mit allen möglichen Mengen mit  $(n-s)$  weissen Kugeln.

Gemäss den vorherigen Überlegungen können wir aus den  $S=4000$  schwarzen Kugeln  $\binom{S}{s}$  verschiedene Teilmengen mit  $s$  Kugeln wählen.

Jede dieser „schwarzen“ Mengen können wir nun mit den  $\binom{N-S}{n-s}$  verschiedenen „weissen“ Mengen zu einer Menge mit  $n=200$  Kugeln ergänzen (Fig. 3. 3).

Es gibt also total

$$g_s = \binom{S}{s} \binom{N-S}{n-s}$$

verschiedene Teilmengen mit  $n=200$  Kugeln, von welchen  $s$  schwarz sind.

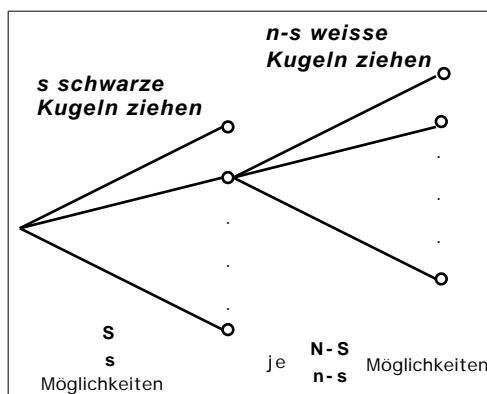


Fig. 3.3: Baumdiagramm für die Erzeugung von Stichproben mit  $s$  schwarzen Kugeln.



### Aufgabe 3.4

Eine Urne enthält 2 schwarze und 3 weisse (numerierte) Kugeln. Wie viele verschiedene Teilmengen mit genau 1 schwarzen Kugel gibt es?

Wenden Sie die eben hergeleitete Formel an und vergleichen Sie das Ergebnis mit demjenigen in Aufgabe 3.1.



#### Die Lösung der Grundaufgabe

$$P(\text{"s schwarze Kugeln"}) = \frac{g_s}{m} = \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}$$

- $N$  = Anzahl Kugeln in der Urne (z. B.  $N=100'000$ )
- $S$  = Anzahl schwarzer Kugeln in der Urne (z. B.  $S = 4000$ )
- $n$  = Umfang der Stichprobe (z. B.  $n = 200$ )
- $s$  = Anzahl schwarzer Kugeln in der Stichprobe (z. B.  $s = 13$ )

Auch diese Formel finden Sie in Ihrer Formelsammlung. Suchen Sie die Formel unter dem Stichwort „Ziehen ohne Zurücklegen“ oder unter „Hypergeometrische Verteilung“.

### Rechenbeispiel

Berechnen wir die Wahrscheinlichkeit, dass die Stichprobe genau die Grundgesamtheit wiedergibt, d.h. genau 8 schwarze Kugeln enthält. Gemäss Fig. 3.4 erhalten wir auf dem Taschencomputer:

$$P(\text{„8 schwarze Kugeln“}) = 14.3\%.$$

$$\begin{array}{l} 100000 \rightarrow nn \\ 4000 \rightarrow ss \\ 200 \rightarrow n \\ 8 \rightarrow s \\ \frac{nCr(ss,s) \cdot nCr(nn-ss,n-s)}{nCr(nn,n)} \quad .142605 \end{array}$$

Fig. 3.4: So oder ähnlich könnte sich die Berechnung der Wahrscheinlichkeit  $P(\text{„8 schwarze“})$  auf einem modernen Taschencomputer präsentieren.



### Aufgabe 3.5

Eine Urne enthält 2 schwarze und 3 weisse Kugeln. Es wird eine Stichprobe vom Umfang 3 gezogen. Mit welchen Wahrscheinlichkeiten enthält sie 0, 1, 2 oder 3 schwarze Kugeln?

Rechnen Sie mit der eben hergeleiteten Formel und zwar mit und ohne Taschenrechner und vergleichen Sie die Ergebnisse mit denjenigen in Aufgabe 3.2.

### Aufgabe 3.6

Einmal tippen beim Schweizerischen Zahlenlotto heisst, die 6 Zahlen zu erraten, die bei der nächsten Ziehung aus den Zahlen 1, 2, ..., 45 ausgelost werden.

Wie gross ist die Wahrscheinlichkeit, dass Sie mit einem Tip 4 richtige Zahlen erraten?

Hinweis für diese nicht ganz einfache Aufgabe: Betrachten Sie eine Urne mit 45 nummerierten Kugeln, wovon einige (welche?) schwarz sind.

In diesem kurzen Kapitel ging es darum, die Wahrscheinlichkeiten für gewisse Stichproben zu berechnen. Prüfen Sie nun anhand der folgenden Lernkontrolle, ob Sie in der Lage sind, die hergeleitete Formel in einem praktischen Beispiel anzuwenden.

Wenn Sie ohne Hilfe des Lösungsteils zum richtigen Ergebnis gekommen sind, werden Sie auch den anschliessenden Kapiteltest ohne Schwierigkeiten bestehen.



### Lernkontrolle: Wie viele Luchse gibt es?

In einem „abgeschlossenen“ Territorium, in welchem der Luchs vorher ausgestorben war, wurden zehn markierte Luchse ausgesetzt. Einige Jahre später wollte man erfahren, wie sich die Luchspopulation entwickelt hat. Zu diesem Zweck stellte man in einer Nacht Fallen auf. Dabei wurden drei markierte und zwei andere Luchse gefangen.

**a**

Weshalb ermittelt man die Populationsstärke so umständlich? Weshalb zählt man nicht einfach die Luchse?

**b**

Beschreiben Sie für dieses Beispiel die Grundgesamtheit und die Stichprobe.

**c**

Zeichnen Sie ein Urnenmodell für die gemachte Untersuchung. Welche Rolle spielen dabei die schwarzen Kugeln? Welche der Grössen  $N$ ,  $S$ ,  $n$ ,  $s$  ist unbekannt und wird gesucht? Welche Werte haben die anderen Grössen?

**d**

Inwiefern unterscheidet sich die neue Fragestellung fundamental von derjenigen des früher betrachteten Qualitätstests?

**e**

Wie viele Luchse gibt es schätzungsweise im untersuchten Territorium? Begründen Sie bitte Ihre Antwort!

**f**

Beim oben beschriebenen Experiment wurden drei markierte und zwei andere Luchse gefangen. Es sind beim geschilderten Vorgehen andere, extreme Ergebnisse denkbar. Welche und welche (evtl. falschen) Schlüsse könnten sich dabei ergeben?

**g**

Nehmen Sie an, es gäbe insgesamt 18 Luchse. Welche Wahrscheinlichkeit hat dann das geschilderte Fangergebnis?

**h**

Bei welcher Luchspopulation mit zwischen 15 und 20 Tieren hat das Fangergebnis die grösste Wahrscheinlichkeit?

## Lösungen der Aufgaben zu Kapitel 3

### Lösung Aufgabe 3.1

Damit wir die 5 Kugeln voneinander unterscheiden können, denken wir sie nummeriert mit 0, 1, 2, 3, 4. Die 2 schwarzen haben die Nummern 0 und 1.

a

Es gibt insgesamt 10 Teilmengen mit 3 Kugeln, nämlich:

$$\begin{array}{ccccc} \{0, 1, 2\} & \{0, 1, 3\} & \{0, 1, 4\} & \{0, 2, 3\} & \{0, 2, 4\} \\ \{0, 3, 4\} & \{1, 2, 3\} & \{1, 2, 4\} & \{1, 3, 4\} & \{2, 3, 4\}. \end{array}$$

b

- Eine Stichprobe mit 0 schwarzen bildet nur die Teilmenge {2, 3, 4}:  $g_0=1$ .
- Eine Stichprobe mit 1 schwarzen bilden die 6 Teilmengen:
 
$$\begin{array}{ccc} \{0, 2, 3\} & \{0, 2, 4\} & \{0, 3, 4\} \\ \{1, 2, 3\} & \{1, 2, 4\} & \{1, 3, 4\}; \end{array}$$
 also ist  $g_1=6$ .
- Eine Stichprobe mit 2 schwarzen bilden die 3 Teilmengen:
 
$$\{0, 1, 2\} \quad \{0, 1, 3\} \quad \{0, 1, 4\};$$
 also ist  $g_2=3$ .
- Stichproben mit 3 oder mehr schwarzen sind natürlich unmöglich:  $g_3=0$ .

### Lösung Aufgabe 3.2

Wir berechnen die Wahrscheinlichkeiten mit der Formel:

$$P(\text{„s schwarze Kugeln“}) = \frac{\text{Anzahl günstiger Teilmengen}}{\text{Anzahl möglicher Teilmengen}} = \frac{g_s}{m}.$$

In der vorhergehenden Aufgabe 3.1 haben wir bereits einige Vorarbeiten geleistet:

- Die Anzahl möglicher Teilmengen mit 3 Kugeln beträgt 10:

$$m=10$$

- Die Anzahl Teilmengen mit 0 schwarzen und 3 weissen Kugeln ist  $g_0=1$ ; folglich gilt:

$$P(\text{„0 schwarze Kugeln“}) = \frac{1}{10} = 0.1$$

- Die Anzahl Teilmengen mit 1 schwarzen und 2 weissen Kugeln ist  $g_1=6$ ; folglich gilt:

$$P(\text{„1 schwarze Kugel“}) = \frac{6}{10} = 0.6$$

- Analog erhalten wir:

$$P(\text{"2 schwarze Kugeln"}) = \frac{3}{10} = 0.3$$

$$P(\text{"3 schwarze Kugeln"}) = \frac{0}{10} = 0$$

### **Lösung Aufgabe 3.3**

Mit den 5 nummerierten Kugeln kann man gemäss Formel

$$m = \frac{5}{3} = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 10$$

verschiedene Teilmengen mit 3 Kugeln bilden. Das Ergebnis von Aufgabe 3.1 wird bestätigt.

### **Lösung Aufgabe 3.4**

Wir setzen  $N=5$ ,  $S=2$ ,  $n=3$ ,  $s=1$  in die Formel

$$g_s = \frac{S}{s} \cdot \frac{N-S}{n-s}$$

ein und erhalten so:

$$g_1 = \frac{2}{1} \cdot \frac{3}{2} = 2 \cdot 3 = 6,$$

wie früher!

### **Lösung Aufgabe 3.5**

Gegeben sind:  $N=5$ ,  $S=2$ ,  $n=3$ ,  $s=0$  (bzw.  $s=1$ ,  $s=2$ ,  $s=3$ ).

Wenn wir diese Werte in die Formel

$$P(\text{"s schwarze Kugeln"}) = \frac{g_s}{m} = \frac{\frac{S}{s} \cdot \frac{N-S}{n-s}}{\frac{N}{n}}$$

einsetzen, erhalten wir:

$$P(\text{"0 schwarze Kugeln"}) = \frac{\frac{2}{0} \cdot \frac{3}{3}}{\frac{5}{3}} = \frac{1 \cdot 1}{10} = \frac{1}{10}$$



$$P(\text{"1 schwarze Kugel"}) = \frac{\frac{\binom{2}{1} \binom{3}{2}}{\binom{5}{3}} = \frac{2 \cdot 3}{10} = \frac{6}{10}}$$

$$P(\text{"2 schwarze Kugeln"}) = \frac{\frac{\binom{2}{2} \binom{3}{1}}{\binom{5}{3}} = \frac{1 \cdot 3}{10} = \frac{3}{10}}$$

Die oben verwendete Formel ist nur für  $0 \leq s \leq n$  gültig. Somit müssen wir z. B. mit dem gesunden Menschenverstand entscheiden, dass:

$$P(\text{"3 schwarze Kugeln"}) = 0.$$

### Lösung Aufgabe 3.6

Zuerst übersetzen wir das Lottoproblem ins Urnenmodell:

- Die Urne enthält  $N=45$  Kugeln, numeriert mit 1, 2, ..., 45.
- Die 6 durch den Tip bezeichneten Kugeln denken wir uns schwarz:  $S=6$ .
- Eine Lottoziehung entspricht einer Stichprobe vom Umfang  $n=6$ .
- Gefragt ist die Wahrscheinlichkeit, dass  $s=4$  schwarze Kugeln in die Stichprobe geraten.

Werten wir die hergeleitete Formel mit diesen Werten von  $N$ ,  $S$ ,  $n$  und  $s$  aus, so erhalten wir (Fig. 3.5):

$$P(\text{"4 schwarze"}) = 0.14\% = 1.4\text{‰}.$$

Unter tausend Tips, können wir also mit 1 bis 2 Vierern rechnen.

$$\frac{\binom{45}{6} \rightarrow \frac{\binom{6}{4} \cdot \binom{39}{2}}{\binom{45}{6}} = \frac{nCr(ss,s) \cdot nCr(nn-ss,n-s)}{nCr(nn,n)} = .001365$$

### Lösung der Lernkontrolle

a

Luchse sind scheue, nachtaktive Tiere und deshalb schwierig zu beobachten. Zudem lassen sich Tiere aus der Ferne schlecht identifizieren, bzw. auseinanderhalten.

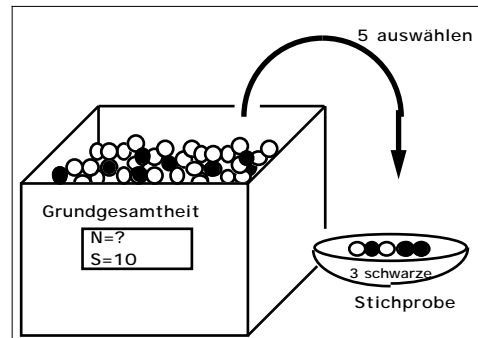
b

Die Menge aller im Territorium lebenden Luchse ist die Grundgesamtheit. Die fünf gefangenen Luchse bilden eine Stichprobe.

c

Urnenmodell (siehe nebenstehende Fig.)

- Die Anzahl der Luchse, d. h. der Umfang  $N$  der Grundgesamtheit ist gesucht.
- Den ausgesetzten Luchsen entsprechen die schwarzen Kugeln:  $S=10$ .



- Die gefangenen Luchse spielen die Rolle der Stichprobe; Stichprobenumfang  $n=5$ .
- Unter den gefangenen Luchsen gibt es 3 markierte, d. h. die Stichprobe enthält  $s=3$  schwarze Kugeln.

d

Im Vergleich zum Qualitätstest, wo der Umfang der Grundgesamtheit (= Anzahl Kugeln in der Urne) gegeben und die Anzahl der schwarzen Kugeln unbekannt sind, ist es hier gerade umgekehrt: Der Umfang der Grundgesamtheit  $N$  ist gesucht, und die Anzahl  $S$  der schwarzen Kugeln ist gegeben.

e

In der folgenden Rechnung setzen wir voraus, dass alle 10 ausgesetzten und markierten Luchse noch im Territorium leben. Ferner nehmen wir an, dass die markierten Luchse in der Stichprobe und in der Grundgesamtheit den gleichen Anteil, d. h. den gleichen Quotienten:

$$\frac{\text{Anzahl markierte}}{\text{Anzahl total}}$$

haben.

Diese Überlegung führt auf die Gleichung:

$$\frac{3}{5} = \frac{10}{N} \quad N = \frac{10 \cdot 5}{3} = 16.7.$$

Man kann unter den gegebenen Voraussetzungen also mit 16 oder 17 Luchsen rechnen.

f

Mögliche *extreme* Fangergebnisse und ihre (evtl. falschen) *Interpretationen* sind:

- Es werden *keine* markierten Luchse gefangen: Alle oder die meisten markierten Luchse sind tot oder es gibt so viele andere Luchse, dass es unwahrscheinlich ist, dass ein markierter in die (wenigen) Fallen geht.
- Es werden *nur markierte* Luchse gefangen: Die Luchspopulation hat sich (praktisch) nicht weiter entwickelt.

g

Wir setzen  $N=18$ ,  $S=10$ ,  $n=5$  und  $s=3$  und erhalten:

$$P(\text{"3 schwarze Kugeln"}) = \frac{\frac{S}{N} \cdot \frac{N-S}{N-1} \cdot \frac{N-S-1}{N-2}}{\frac{n!}{s!(n-s)!}} = 39.2\%$$

h

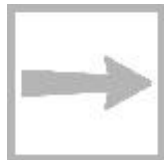
Analog ergeben sich:

N=	15	16	17	18	19	20
P("3 schwarze")=	40.0%	41.2%	40.7%	39.2%	37.2%	34.8%

$N=16$  hat offensichtlich die grösste Wahrscheinlichkeit.

---

## 4

**Der Qualitätstest unter der rechnerischen Lupe**

In diesem Kapitel wollen wir die Wahrscheinlichkeiten und für die Fehler 1. und 2. Art bei unserem Qualitätstest *berechnen*. Nach dem Studium dieses Kapitels werden Sie in der Lage sein, und für einen vergleichbaren neuen Test *selbständig* zu bestimmen. Darüber hinaus lernen Sie, mit zwei neuen Begriffen, nämlich dem *Signifikanzniveau* und der *Macht* des Tests, sachverständig umzugehen.

## 4.1

**Der Fehler 1. Art - das Signifikanzniveau**

Wiederholen wir: Bei unserem Qualitätstest machen wir den Fehler 1. Art, wenn mehr als 13 schwarze Kugeln in die Stichprobe geraten, obwohl die Urne nicht mehr als 4000 schwarze Kugeln enthält. Wie gross ist die Wahrscheinlichkeit für diesen Fehlentscheid?

Die Wahrscheinlichkeit hängt, wie erwähnt, von der tatsächlichen Anzahl schwarzer Kugeln in der Urne ab; sie ist am grössten, wenn gerade 4000 schwarze Kugeln in der Urne sind und dieser Maximalwert heisst *Irrtumswahrscheinlichkeit*.

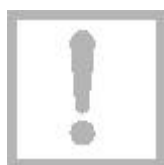
**Berechnung von  $\alpha$** 

Wie Sie gleich feststellen werden, ist es gar nicht so schwierig, zu berechnen. Gehen Sie einfach die folgenden Umformungen aufmerksam durch:



$$\begin{aligned}
 &= P(\text{"mehr als 13 schwarze Kugeln in der Stichprobe"}) \\
 &= P(\text{"14 oder 15 oder ... 199 oder 200 schwarze Kugeln"}) \\
 &= 1 - P(\text{"0 oder 1 oder ... 13 schwarze Kugeln"}) \\
 &= 1 - [P(\text{"0 schwarze"}) + P(\text{"1 schwarze"}) + \dots + P(\text{"13 schwarze"})]
 \end{aligned}$$

Für diese Wahrscheinlichkeiten haben wir im letzten Kapitel eine Formel hergeleitet. Damit der Term überblickbar bleibt, schreiben wir ihn mit dem Summenzeichen:



$$= 1 - \sum_{s=0}^{13} P(\text{"s schwarze Kugeln"}) = 1 - \sum_{s=0}^{13} \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}$$

Wir können die Formel fast eins zu eins in den Taschenrechner eingeben (Fig. 4.1).

Nach einigen Minuten Rechenzeit erhalten wir „grosso modo“ das gleiche Resultat wie mit der Computersimulation:

$$\alpha = 0.0310 = 3.10 \%$$

Nochmals die Interpretation für die

Realität: Wenn der Einkäufer diesen Test *häufig* durchführt, so wird er durchschnittlich 3 von 100 Sendungen *irrtümlicherweise ablehnen*.

$$1 - \frac{nCr(ss,s) \cdot nCr(nn-ss,n-s)}{nCr(nn,n)} = 0.0310$$

s=0

Fig. 4.1: Die Irrtumswahrscheinlichkeit auf einem Taschenrechner

### Das Signifikanzniveau

Wenn der Statistiker mehr als 13 defekte Knallkörper in der Stichprobe findet, sagt er, das Testergebnis sei *signifikant* (bezeichnend) für eine Sendung mit unzulässiger Qualität. Es ist in der Statistik üblich, von *Signifikanz* zu sprechen, wenn die Irrtumswahrscheinlichkeit höchstens 5% ist. Man sagt dann auch, man habe auf dem *5%-Signifikanzniveau* entschieden.

Diese Konvention gilt eventuell nicht, wenn es um sehr wichtige Entscheidungen geht, beispielsweise um Menschenleben. Dann versucht man oft auf dem 1%-Niveau zu entscheiden.

Merken wir uns:



Bei einem statistischen Test sprechen wir (üblicherweise) von *Signifikanz*, wenn die Irrtumswahrscheinlichkeit 5% ist.



### Aufgabe 4.1

Wir ändern den besprochenen Qualitätstest wie folgt ab: Die Sendung wird *abgelehnt*, wenn mehr als ein Dutzend defekter Knallkörper in die Stichprobe geraten. Die neue Irrtumswahrscheinlichkeit sei  $\alpha'$ .

a

Entscheiden Sie, ohne zu rechnen: Ist  $\alpha'$  grösser oder kleiner als  $\alpha$ ?

b

Berechnen Sie  $\alpha'$ .

c

Ermitteln Sie  $\alpha'$  aufgrund der Computersimulation in Kap. 2.2, Seite 2.6.

d

Sind auch beim neuen Entscheidungsschema *signifikante* Entscheidungen möglich?

## 4.2 Der Fehler 2. Art - die Macht des Tests

Einen Fehler 2. Art begehen wir bekanntlich bei unserem Qualitätstest, wenn wir nicht merken, dass die Sendung nicht in Ordnung ist und sie also irrtümlicherweise akzeptieren. Wie gross ist die Wahrscheinlichkeit für diesen Irrtum?

Im Urnenmodell lautet die Frage: Wie gross ist die Wahrscheinlichkeit, dass in die Stichprobe 13 oder weniger schwarze Kugeln geraten, obwohl die Grundgesamtheit mehr als 4000 schwarze Kugeln enthält.

Diese Frage lässt sich *nicht generell* beantworten: Es kommt darauf an, wie viele schwarze Kugeln die Urne tatsächlich enthält. Wir können beispielsweise die Wahrscheinlichkeit berechnen, dass unser Test nicht merkt, dass die Urne 8000, statt nur 4000, schwarze Kugeln enthält. Das wollen wir im folgenden tun.

### Berechnung von $\beta$



$$\begin{aligned} &= P(\text{"höchstens 13 schwarze in der Stichprobe"}) \\ &= P(\text{"0 oder 1 oder 2 oder .... oder 13 schwarze"}) \\ &= P(\text{"0 schwarze"}) + P(\text{"1 schwarze"}) + \dots + P(\text{"13 schwarze"}) \end{aligned}$$

Schreiben wir die Summe mit Hilfe des Summenzeichens, ersetzen wir die Wahrscheinlichkeiten mit der hergeleiteten Formel und werten wir schliesslich den Term auf dem Taschenrechner aus, so erhalten wir:

$$= \sum_{s=0}^{13} \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}} = 0.264 = 26.4\%$$

Dabei haben wir folgende Werte eingegeben:

- $N = 100'000$
- $S = 8000$
- $n = 200$

Das Resultat liegt in der Nähe des Wertes von 25%, den wir mit der Computersimulation erhalten haben.

Die Wahrscheinlichkeit ist, wie schon früher erwähnt, unerwartet gross: In durchschnittlich rund einem Viertel der Tests würden wir *nicht* merken, dass die Qualität doppelt so schlecht ist, wie vereinbart. Wenn der Einkäufer sich dieses Risiko nicht leisten kann oder will, muss er seinen Test verbessern.

### Die Macht des Tests

Wenn die Wahrscheinlichkeit ist, dass wir *nicht* entdecken, dass die Sendung 8000 defekte Knallkörper enthält, so ist eben  $1 -$  die Wahrscheinlichkeit, diesen Tatbestand aufzudecken:

$$1 - = 1 - 26.4\% = 73.6\%.$$

Man bezeichnet deshalb  $1 -$  als die *Macht des Tests* bei einem effektiven Ausschussanteil von 8% (8000 unter 100'000).

Merken wir uns:



Unter der *Macht* eines Tests verstehen wir die Wahrscheinlichkeit  $1 -$ , mit welcher ein Fehler 2. Art verhindert wird. Sie hängt vom tatsächlichen Ausschussanteil ab.



### Aufgabe 4.2

Machen Sie mindestens einen Vorschlag, wie man  $\beta$  beim ursprünglichen Qualitätstest verkleinern, bzw. wie man die Macht des Tests vergrößern könnte.



### Lernkontrolle: Produktionskontrolle

Auf einer Verpackungsanlage werden stündlich 3600 Dosen mit Tomaten gefüllt. Es kann nicht ganz verhindert werden, dass gewisse Dosen ein zu kleines Abtropfgewicht aufweisen. Der Produktionsprozess müsste jedoch gestoppt werden, wenn der Anteil an zu leichten Dosen über 2% steigt. Um diesen Qualitätsstandard zu halten, werden stündlich 100 zufällig ausgewählte Dosen kontrolliert.

**a**

Bei welcher Anzahl zu leichter Dosen sollte der Produktionsprozess unterbrochen werden? Signifikanzniveau 5%.

Lösungshinweis: Lösen Sie diese neue Aufgabe durch sinnvolles Probieren.

**b**

Wie oft pro Monat würde bei Ihrer Entscheidung der Produktionsprozess irrtümlicherweise gestoppt, wenn die Anlage exakt 2% Ausschuss produziert?

**c**

Entscheiden Sie aufgrund Ihres Ergebnisses in Teilaufgabe a: Wie gross ist die Wahrscheinlichkeit für den Fehler 2. Art, falls angenommen wird, dass der Ausschussanteil doppelt so hoch ist wie erlaubt?

Wenn Sie diese Aufgabe richtig gelöst und die Lösung verstanden haben, können Sie sich zum Kapiteltest anmelden.

## Lösungen der Aufgaben zu Kapitel 4

### Lösung Aufgabe 4.1

a)  $\alpha' = P(\text{"mehr als 12 schwarze Kugeln in der Stichprobe"})$   
 $= P(\text{"13 oder 14 oder ... 199 oder 200 schwarze Kugeln"})$   
 $> P(\text{"14 oder 15 oder ... 199 oder 200 schwarze Kugeln"})$   
 $= \dots$

Also:  $\alpha' > \dots$

b)

Hier die rechnerische Bestätigung:

$$\alpha' = 1 - \sum_{s=0}^{12} P(\text{"s schwarze Kugeln"}) = 1 - \sum_{s=0}^{12} \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}} = 0.0597 = 6\%$$

$\alpha'$  ist fast doppelt so gross wie  $\alpha$ .

c)

Bei der Computersimulation in Kap. 2.2 haben wir sogar noch einen erheblich grösseren Wert erhalten:  $\alpha' = 9\%$ .

d)

Nein, bei dieser Variante des Qualitätstests würde man die Sendung nicht auf dem 5%-Niveau ablehnen.

### Lösung Aufgabe 4.2

#### Vorschlag 1

Wenn man *höchstens ein Dutzend* defekte Knallkörper in der Stichprobe erlaubt, ergibt sich eine *kleinere* Wahrscheinlichkeit für den Fehler 2. Art:

$$\beta = \sum_{s=0}^{12} \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}} = 0.182 = 18.2\%$$

Aber in der vorhergehenden Aufgabe haben wir schon gesehen, dass diese Änderung die Irrtumswahrscheinlichkeit *vergrössert*.



**Vorschlag 2**

Man *vergrössert den Umfang*  $n$  der Stichprobe und passt die Testentscheidung an, denn: Je grösser  $n$  ist, um so „zuverlässiger“ ist die Stichprobe als Abbild der Grundgesamtheit und um so kleiner fallen die Fehlerwahrscheinlichkeiten *und* aus.

**Lösung der Lernkontrolle**

Wir übersetzen das Problem ins Urnenmodell:

- Die Urne enthält  $N=3600$  Kugeln; dies entspricht der stündlichen Produktion.
- Einige der Kugeln sind schwarz; sie entsprechen den zu leichten Dosen. Wenn der Produktionsprozess in Ordnung ist, sollte die Urne nicht mehr als  $S=72$  schwarze Kugeln enthalten (2% von 3600).
- In einer zufälligen Stichprobe von  $n=100$  müssten wir dann mit ungefähr 2 schwarzen Kugeln rechnen.
- Wesentlich mehr als 2 schwarze in der Stichprobe sind m. a. W. unwahrscheinlich und wecken unseren Verdacht bezüglich der Qualität der Grundgesamtheit: Wir stoppen und überprüfen den Produktionsprozess.

**a**

Was heisst aber „wesentlich“ mehr als 2? Wir versuchen, diese *gesuchte Grenze*  $g$  so festzulegen, dass wir den Produktionsprozess irrtümlicherweise mit einer Wahrscheinlichkeit 5% stoppen.

- Wie gross ist die Irrtumswahrscheinlichkeit für Grenze  $g=3$ ? Im ungünstigsten Fall, d. h. für  $S=72$ , gilt:

$$\begin{aligned}
 &= P(\text{"mehr als 3 schwarze Kugeln unter 100"}) \\
 &= 1 - \sum_{s=0}^3 P(\text{"s schwarze Kugeln"}) \\
 &= 1 - \sum_{s=0}^3 \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}} = 13.8\%
 \end{aligned}$$

Auf der Basis  $g=3$  würden wir also *nicht* auf dem Signifikanzniveau von 5% entscheiden.

- $g=4$ : Analoge Rechnung ergibt:  $\approx 4.83\%$ .

Man wird also den Produktionsprozess stoppen, wenn die Stichprobe 5 oder mehr zu leichte Dosen enthält.

**b**

Dieser Entscheid ist dann durchschnittlich in rund 5 von 100 Fällen falsch.

In einem Monat mit 400 Stunden Schichtarbeit würde die Produktion also durchschnittlich 20-mal (irrtümlicherweise) gestoppt, wenn die Anlage exakt 2% Ausschuss produzieren würde.

C

Gesucht ist die Wahrscheinlichkeit für den Fehler 2. Art bei 4% Ausschuss.

Wir übersetzen die Frage ins Urnenmodell: Mit welcher Wahrscheinlichkeit enthält eine Stichprobe von  $n=100$  Kugeln weniger als 5 schwarze Kugeln, obwohl sich  $S=144$  schwarze Kugeln in der Urne befinden ( $144=4\%$  von 3600).

Es gilt:

$$\begin{aligned}
 &= P(\text{"weniger als 5 schwarze Kugeln unter 100"}) \\
 &= \sum_{s=0}^4 P(\text{"s schwarze Kugeln"}) \\
 &= \sum_{s=0}^4 \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}} = 62.9\%
 \end{aligned}$$

Die Macht dieses Tests ist überraschend gering. In weniger als 40% der Fälle würde er entdecken, dass der Ausschussanteil auf das Doppelte des Erlaubten geklettert ist. Vermutlich muss man den Umfang der Stichprobe erhöhen.

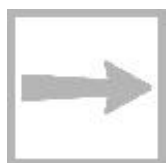
## 5

**Fachbegriffe bei statistischen Tests**

Das Beispiel Knallkörper hat Sie mit der Denk- und Schlussweise bei einem statistischen Test ein bisschen vertraut gemacht. Und Sie haben bereits zwei wichtige Begriffe (Fehler 1. und 2. Art) kennen gelernt. Hier eine Zusammenfassung:



<b>Zweifel:</b>	Ist die Sendung in Ordnung, d. h. ist der Ausschussanteil in unserer Sendung wirklich nicht höher als 4%?
<b>Daten erheben:</b>	Eine Stichprobe von 200 zufällig ausgewählten Knallkörpern prüfen.
<b>Entscheidungsregel:</b>	Sendung ablehnen, wenn die Stichprobe mehr als 13 defekte Exemplare enthält.
<b>Fehler 1. Art:</b>	Wir lehnen die Sendung ab, <i>obwohl</i> sie in Ordnung ist : <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 5px auto;"> <math>= P(\text{„Fehler 1. Art bei 4\% Ausschuss“})</math> </div>
<b>Fehler 2. Art</b>	Wir akzeptieren die Sendung, obwohl sie <i>nicht</i> in Ordnung ist: <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 5px auto;"> <math>= P(\text{„Fehler 2. Art bei z.B. 8\% Ausschuss“})</math> </div>



In den weiteren Tests wird es sich als hilfreich erweisen, wenn wir die folgenden vier neuen Fachbegriffe in unsere Überlegungen einbauen können:

- Nullhypothese
- Alternativhypothese
- Testgrösse
- Verwerfungsbereich

Sie werden diese Begriffe nicht als abstrakte Definitionen, sondern ebenfalls anhand unseres Qualitätstests kennen lernen. Und Sie werden schon bald sehen, dass es sich dabei nicht um (überflüssigen) Ballast handelt. Im Gegenteil: Die Begriffe werden Ihnen ermöglichen, sich in neuen und anspruchsvollen Testsituationen besser zurecht zu finden.

## 5.1 Die Nullhypothese

Unter der Nullhypothese (Symbol:  $H_0$ ) versteht man bei einem statistischen Test die Behauptung, an der man zweifelt oder von der man glaubt, dass sie falsch sei. Man möchte also  $H_0$  eigentlich ablehnen. Man will jedoch nicht vorschnell entscheiden, sondern aufgrund von objektiven Fakten, sodass man auch Personen mit anderer Auffassung überzeugen kann. In unserem Knallkörperbeispiel formulieren wir aufgrund unseres Zweifels:

$H_0$ : „Der Ausschussanteil ist höchstens 4%“

Der (misstrauische) Einkäufer vermutet ja, dass der Ausschussanteil über 4% liegt, und in dieser Vermutung wird er bestärkt, wenn er  $H_0$  ablehnen kann.

Merken Sie sich:



Unter der **Nullhypothese  $H_0$**  versteht man bei einem statistischen Test die Behauptung (Vermutung), die man *widerlegen* möchte.

### Warum nicht direkt beweisen?

Möglicherweise macht Ihnen diese Definition Mühe, weil Sie verständlicherweise den Eindruck haben, man mache im Hinblick auf das Ziel einen Umweg. Warum nicht *direkt* beweisen, dass der Ausschuss grösser als 4% ist?

Beachten Sie, dass man mit Hilfe der Statistik nichts schlüssig *beweisen* kann. Man kann nur *Wahrscheinlichkeiten* berechnen und dann z.B. wie folgt entscheiden: Wenn unter der Voraussetzung von  $H_0$  das Testergebnis eine *kleine* Wahrscheinlichkeit hat, so spricht dies eben gegen diese Hypothese  $H_0$ . In diesem Fall verwerfen wir  $H_0$  zugunsten der sog. Alternativ- oder Gegenhypothese  $H_1$  (auf die wir im nächsten Abschnitt zu sprechen kommen) und wir nehmen die kleine Irrtumswahrscheinlichkeit in Kauf, einen Fehler 1. Art zu begehen.

Übrigens gilt nicht nur in der Statistik, sondern ganz allgemein, dass man eine Theorie anhand von Experimenten nur widerlegen, aber nicht beweisen kann. Wenn das Experiment mit den Voraussagen übereinstimmt, heisst das nicht, dass die Theorie richtig ist. Es kann ja mehrere verschiedene Theorien geben, die teilweise die gleichen Vorhersagen machen.

### Vergleich mit einem Widerspruchsbeweis

Das "Beweis"-Prinzip des statistischen Tests lässt sich mit dem Prinzip des Widerspruchsbeweises vergleichen, das Ihnen sicher schon in der Mathematik begegnet ist – vielleicht gerade in folgendem Beispiel:

Wenn Sie beweisen wollen, dass sich  $\sqrt{2}$  nicht als ein gewöhnlicher Bruch schreiben lässt (Behauptung  $B_1$ ), so können Sie probenhalber das Gegenteil (Behauptung  $B_0$ ) als richtig annehmen. Aus  $B_0$  folgt dann, dass sich eine gewisse natürliche Zahl auf zwei unterschiedliche Arten in Primfaktoren zerlegen lässt. Das steht in Widerspruch zur Eindeutigkeit der Primfaktorzerlegung. Der Widerspruch lässt sich nur lösen, indem Sie die Hypothese  $B_0$  zugunsten der Alternative  $B_1$  fallen lassen.

Dem Widerspruch im Beweis entspricht beim statistischen Test das Ereignis mit der *kleinen* Wahrscheinlichkeit ( 5%).



#### Aufgabe 5.1:

Max glaubt, daß eine von Moritz verwendete Münze beim Werfen öfter „Zahl“ zeigt als „Bild“. Er beschliesst, sich wie folgt „Klarheit“ zu verschaffen: Zeigt die Münze in 20 Würfeln mehr als beispielsweise 14-mal „Zahl“, so wird er seinen Verdacht als erhärtet betrachten.

Wie lautet in diesem Beispiel die Nullhypothese  $H_0$ ?

## 5.2

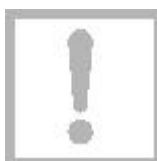
### Die Alternativhypothese (Gegenhypothese)

In unserem Knallkörper-Beispiel möchten wir  $H_0$  (= "höchstens 4% Ausschuss") ablehnen, weil wir „vermuten“ (die Gründe dafür sind, wie gesagt, unwesentlich), daß die Qualität schlechter ist als vereinbart, d. h. dass der Ausschussanteil grösser als 4% ist. Als Alternativhypothese (Symbol:  $H_1$ ) postulieren wir:

$H_1$ : „Der Ausschussanteil ist grösser als 4%“

Es ist nicht falsch, wenn Sie den Eindruck haben, dass eigentlich diese Alternativhypothese  $H_1$  am Anfang unserer Überlegungen steht und dass  $H_0$  mehr oder weniger als Gegenteil davon nachträglich formuliert wird.

Allgemein können wir sagen:



Unter der **Alternativ- oder Gegenhypothese  $H_1$**  versteht man bei einem statistischen Test die Behauptung, die man akzeptiert, wenn  $H_0$  verworfen wird.

Wenn Statistikfachleute  $H_0$  ablehnen können, fühlen Sie sich in der starken Position: Die *kleine* Wahrscheinlichkeit für das Eintreffen des Testergebnisses spricht gegen  $H_0$  und somit für  $H_1$ . Aber auf keinen Fall dürfen Sie sagen, es sei *bewiesen*, dass  $H_0$  falsch und  $H_1$  richtig sei.

Und wenn das Testergebnis *nicht* so ausfällt, dass Sie  $H_0$  verwerfen können, ist ebenfalls nicht bewiesen, dass  $H_0$  zutrifft. Man drückt das auch sprachlich aus, indem man neutraler sagt, man *behalte*  $H_0$  (vorläufig) bei, bzw. man *verwerfe*  $H_0$  *nicht*.

Die Alternativhypothese  $H_1$  beeinflusst zusammen mit  $H_0$  den sog. Verwerfungsbereich, den wir weiter unten besprechen werden.



### Aufgabe 5.2

*Wir kommen auf das Max&Moritz-Problem zurück. Bitte nehmen Sie dieses ebenso einfache wie lehrreiche Beispiel ernst und vor allem: Nehmen sie falsche Antworten nicht auf die leichte Schulter.....*

*Max glaubt, wie gesagt, daß eine von Moritz verwendete Münze beim Werfen öfter „Zahl“ zeigt als „Bild“. Er beschließt, sich wie folgt „Klarheit“ zu verschaffen: Zeigt die Münze in 20 Würfeln mehr als 14-mal „Zahl“, so wird er seinen Verdacht als erhärtet betrachten.*

*Wie lautet in diesem Beispiel die Alternativhypothese  $H_1$ ?*

## 5.3 Die Testgrösse

Unser Test ist ein Zufallsexperiment. Dabei achten wir nur auf die zufallsbedingte *Anzahl* der *defekten* Knallkörper in der Stichprobe. Diese Anzahl ist also eine sog. Zufallsgrösse, man nennt sie die *Testgrösse* (Symbol:  $Z$ ). Aufgrund des Wertes von  $Z$  entscheiden wir, ob wir  $H_0$  verwerfen oder beibehalten.

Merken wir uns:



Unter der **Testgrösse  $Z$**  versteht man die beim statistischen Test ermittelte Zahl, von deren Wert schliesslich abhängt, ob  $H_0$  verworfen oder eben nicht verworfen wird.



### Aufgabe 5.3

*Welche Testgrösse  $Z$  verwendet Max? Bitte  $Z$  möglichst präzis definieren.*

## 5.4 Der Verwerfungsbereich - der Annahmehereich

$H_0$  wird abgelehnt, wenn ein Ereignis eintritt, das eine *kleine* Wahrscheinlichkeit hat, falls  $H_0$  gelten würde. Bei unserem Knallkörperbeispiel wollen wir  $H_0$  verwerfen, falls wir mehr als 13 defekte Raketen in der Stichprobe finden, d.h. falls

$$Z > 13$$

ist oder m. a. W., falls die Testgrösse  $Z$  einen Wert aus dem Intervall in den ganzen Zahlen

$$V = \{14, 15, \dots, 199, 200\}$$

annimmt (Abb. 5.1). Wir nennen dieses Intervall den Verwerfungsbereich  $V$ ; üblich ist auch die Bezeichnung *kritischer Bereich K*.

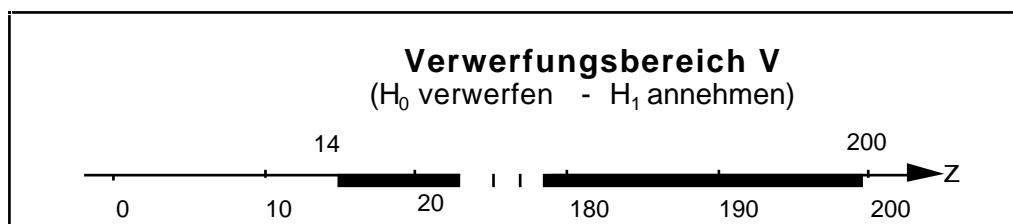


Fig. 5.1: Der Verwerfungsbereich beim Knallkörperbeispiel (es gehören nur die überstrichenen *ganzen* Zahlen dazu).

Merken wir uns:



Unter dem **Verwerfungsbereich  $V$**  versteht man bei einem statistischen Test die im *voraus* festgelegte Zahlenmenge mit der folgenden Eigenschaft:

Falls  $H_0$  zutrifft, so ist die Wahrscheinlichkeit, dass der Wert der Testgrösse  $Z$  in  $V$  liegt, klein. Wird der Test durchgeführt und kommt  $Z$  in  $V$  zu liegen ( $Z \in V$ ), so wird  $H_0$  zugunsten von  $H_1$  verworfen.

Analog ist der Annahmebereich definiert: Wenn der Wert der Testgrösse nicht in  $V$  ( $Z \notin V$ ), sondern in das Intervall

$$\{0, 1, \dots, 12, 13\}$$

fällt, wird  $H_0$  nicht verworfen, sondern  $H_0$  wird angenommen; besser ist, wie gesagt, das Wort „beibehalten“ (Fig. 5.2). Für den Annahmebereich führen wir *kein* eigenes Symbol ein.

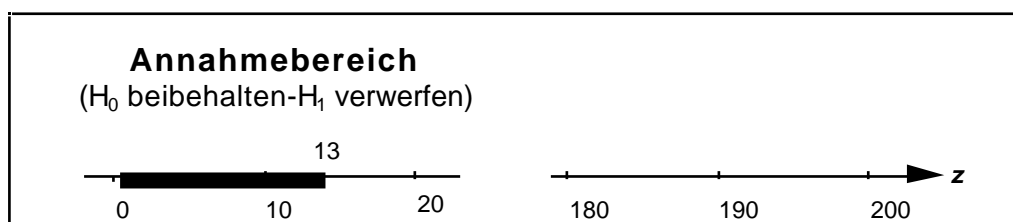


Fig. 5.2: Der Annahmebereich im Knallkörperbeispiel



#### Aufgabe 5.4

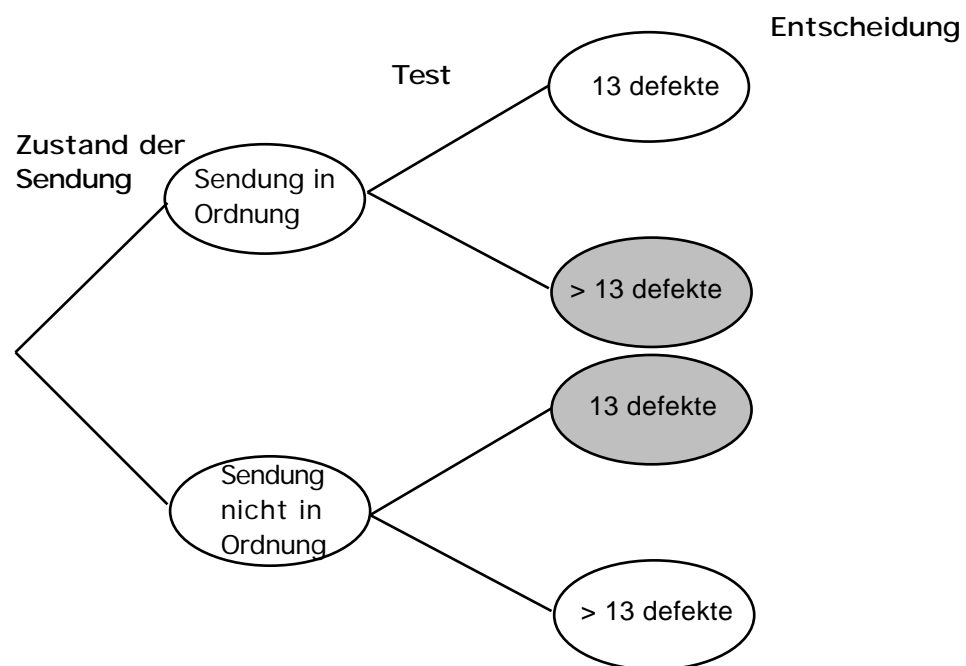
*Max&Moritz-Problem: Welchen Verwerfungsbereich  $V$  wählt Max? Zeichnen Sie!*

Versuchen Sie nun zum Schluss, die folgende Lernkontrolle vorerst ohne Zurückblättern zu lösen. Kontrollieren Sie so, wie zuverlässig Sie die neuen Begriffe und Zusammenhänge bereits intuitiv haben.



### Lernkontrolle: Die Begriffe in der Übersicht

Sie haben gesehen, wie man die (richtigen und falschen) Entscheidungen beim Test der Knallkörper übersichtlich in einem Baumdiagramm darstellen kann. Hier nochmals das Gerüst des Diagramms.



Ergänzen Sie nun die Skizze mit den neuen Fachbegriffen. Es wird erwartet, dass Sie die folgenden Symbole, Beziehungen und Begriffe an der richtigen Stelle eintragen :

$$H_0, H_1, Z \in V, Z \notin V,$$

Weiter sollen Sie die folgenden Entscheidungen und möglichen Fehler eintragen:

- Sendung annehmen, Sendung ablehnen,
- Fehler 1. Art, Fehler 2. Art.

Schliesslich sollen Sie auch die Wahrscheinlichkeiten

$$\alpha \text{ und } \beta,$$

mit denen diese Entscheidungen getroffen werden, im Diagramm (auf den entsprechenden Zweigen) unterbringen.

Wenn Sie mit Ihrer Leistung zufrieden sind, können Sie sich dem Kapiteltest zuwenden.



---

## Lösungen der Aufgaben zu Kapitel 5

---

### Lösung Aufgabe 5.1

Max möchte die folgende Nullhypothese widerlegen:

$$H_0: \text{„Die Münze zeigt ‚Zahl‘ und ‚Bild‘ gleich oft“}$$

oder:

$$H_0: \text{„Die Münze zeigt ‚Zahl‘ mit der Wahrscheinlichkeit } p = \frac{1}{2} \text{“}$$

Es ist sehr wichtig, dass Sie diese Aufgabe richtig gelöst haben. Wenn Sie auf eine wesentlich andere Formulierung gekommen sind, sollten Sie den ganzen Abschnitt nochmals durchlesen und sich klar machen, weshalb  $H_0$  eine Behauptung ist, welche man *ablehnen* möchte.

---

### Lösung Aufgabe 5.2

Max gedenkt die Nullhypothese  $H_0$  zugunsten der folgenden Alternativhypothese zu verwerfen:

$$H_1: \text{„Die Münze zeigt öfter ‚Zahl‘ als ‚Bild‘“}$$

oder:

$$H_1: \text{„Die Münze zeigt ‚Zahl‘ mit der Wahrscheinlichkeit } p > \frac{1}{2} \text{“}$$

---

### Lösung Aufgabe 5.3

Max verwendet folgende Testgrösse:

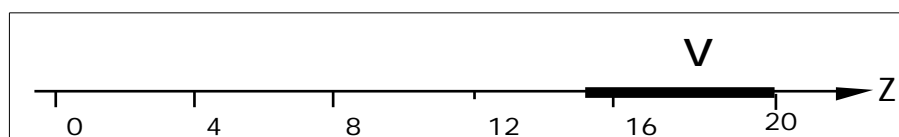
$$Z = \text{„Anzahl der Ergebnisse ‚Zahl‘ in 20 Würfeln“}$$

---

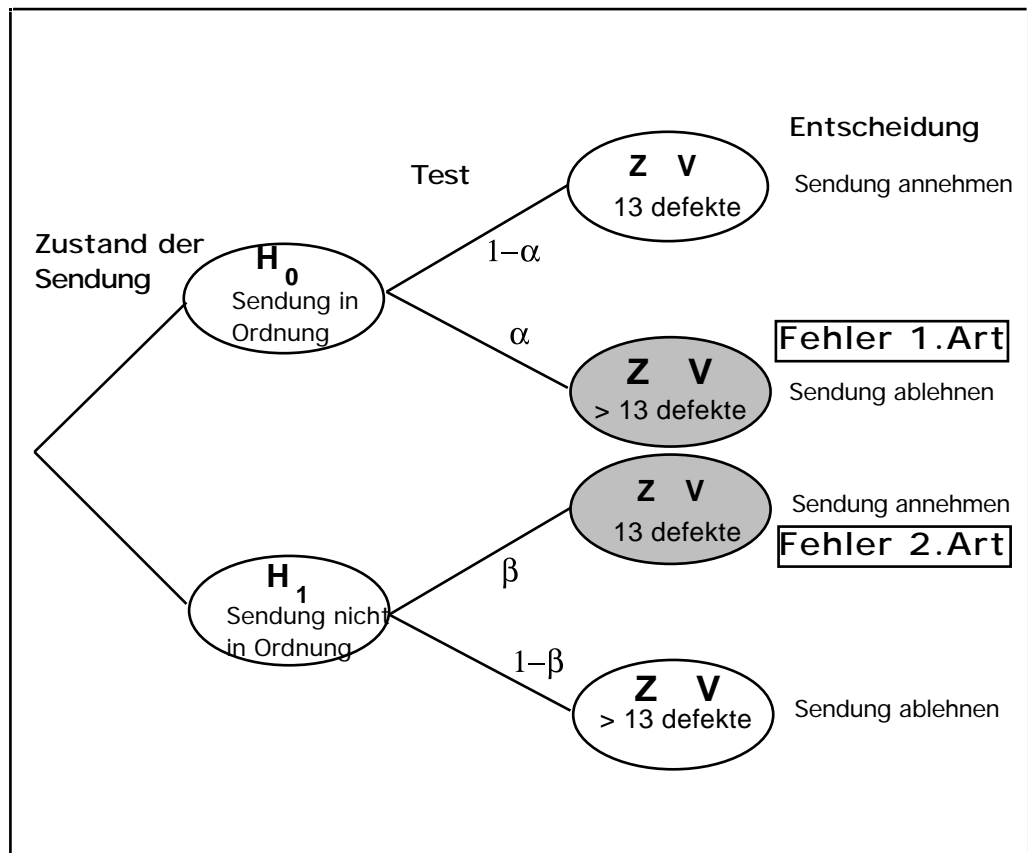
### Lösung Aufgabe 5.4

Max operiert mit dem folgenden Verwerfungsbereich:

$$V = \{15, 16, 17, 18, 19, 20\}$$



## Lösung der Lernkontrolle



## 6

## Exakter Test von Fisher: Wie einen Herzinfarkt behandeln?

### Die Fragestellung

Ein Herzinfarkt ist die Folge von verstopften Herzkranzarterien, welche den Herzmuskel mit Blut versorgen (sollten). Mehr als zwei Drittel aller Patienten überleben die unmittelbaren Folgen eines Herzinfarkts. Wenn sich ihr Zustand stabilisiert hat, gilt es, die verstopften Gefäße möglichst schnell wieder durchgängig zu machen, so dass der Schaden am Herzmuskel beschränkt bleibt. Zwei Methoden stehen im Vordergrund: die primäre Angioplastie, bei der im Spital innert Stunden nach dem Herzinfarkt mit Hilfe eines Ballonkatheters die verstopfte Stelle ausgeweitet wird, und die Behandlung mit Medikamenten, welche den Blutpfropf wieder auflösen. Eine kleine Studie hat die folgenden Ergebnisse ergeben:

Behandlungsmethode→ Wirkung ↓	Angioplastie	Medikamente	Summe
Patient überlebt	53	40	93
Patient stirbt	2	5	7
Summe	55	45	100



In diesem Kapitel wollen wir testen, ob man aus diesen Ergebnissen schließen kann, dass die Angioplastie der medikamentösen Behandlung *signifikant* überlegen ist. Der Unterschied scheint eklatant: Die Medikamente versagen in über 10% der Fälle, die Angioplastie dagegen in weniger als 5%.

### Vierfeldertafeln



Es gibt viele ähnliche Fragestellungen, die auf eine sog. Vierfeldertafel wie oben führen. In diesem Kapitel werden Sie lernen, solche Vierfeldertafeln von der Statistik her zu beurteilen. Insbesondere werden Sie dabei sehen, wie die eingeführten Fachbegriffe  $H_0$ ,  $H_1$ ,  $Z$ ,  $V$  und eingesetzt werden.

Wir gehen nach einer leicht verständlichen und elementaren Methode vor, die von R. A. Fisher, einem Pionier der mathematischen Statistik, eingeführt wurde.

### Zusatzinformation 1: Praktische Tragweite

Die praktische Tragweite des Vergleichs zwischen Angioplastie und Medikamenten: Kleinere Untersuchungen (etwa entsprechend obiger Tabelle) in Kanada schienen auf einen Vorteil der Angioplastie hinzuweisen, eine Methode, die einen erfahrenen Spezialisten in einem grösseren Spital erfordert.

Deshalb transportierte man frische Herzinfarktpatienten oft über sehr weite Distanzen in ein Zentrumsspital. In diesem Kapitel geht es um die Frage, ob – aus der Sicht der Statistik – diese Transporte zu rechtfertigen sind.

Das Beispiel ist auch deshalb sehr interessant, weil inzwischen die Frage durch eine Untersuchung in einer grossen Grundgesamtheit geklärt ist. Mehr darüber am Schluss des Kapitels.

### **Zusatzinformation 2: Doppelblindversuche**

Wie wählt man die Gruppe der 100 Patienten aus, und wie teilt man diese Gruppe für die beiden Behandlungen auf? Diese beiden Fragen sind absolut zentral, um brauchbare Ergebnisse zu erhalten. Überleben oder nicht hängt neben der Behandlung noch von vielen anderen Einflussgrössen ab wie Alter, Gewicht, Schwere des Infarktes etc. Man muss daher dafür sorgen, dass man möglichst gleich risikoreiche Patienten vergleicht. Dazu gibt man sich zunächst einmal relativ enge Vorgaben bezüglich der offensichtlichen Faktoren (z. B. Männer, Alter 45 - 55, normales Gewicht etc.). Um subtilere bewusste oder unbewusste Bevorzugungen zu vermeiden, lässt man danach das Los entscheiden, wer mit welcher Methode behandelt wird: Man wählt also z. B. 45 Zahlen zufällig aus 1, 2, 3, ..., 100 und behandelt alle Patienten mit diesen Nummern medikamentös. In der Fachsprache heisst das *Randomisieren*.

In medizinischen Versuchen verlangt man wenn möglich noch zusätzlich, dass weder der Arzt noch der Patient wissen, welche Methode angewendet wurde (doppelblind). Damit vermeidet man den sog. Placebo-Effekt, d. h. die "Heilung" durch Scheinmedikamente aufgrund von Suggestivkräften.

In unserem Beispiel ist es aber aus offensichtlichen Gründen nicht möglich, einen Doppelblindversuch durchzuführen.

## **6.1 Die Vierfeldertafel im Licht der Statistikbegriffe**

Eigentlich möchten wir ja beweisen, dass die Angioplastie die bessere Methode ist. Nun wissen Sie inzwischen, dass man solche Behauptungen in der Statistik *indirekt* "beweist". Man nimmt an, das Gegenteil treffe zu, nämlich die Wirkung der Angioplastie sei "um Null besser" als die Medikamentenbehandlung (daher der Name *Nullhypothese*). Wenn daraus folgt, dass die obige Vierfeldertafel mit ihren Zahlenwerten *unwahrscheinlich* ist, sind wir geneigt, die Überlegenheit der Angioplastie zu akzeptieren.

Der Gedankengang wird klarer, wenn wir die neuen Begriffe verwenden.

### **Nullhypothese und Alternativhypothese:**

In unserem Fall postulieren wir also die folgende Nullhypothese:

$$H_0: \text{"Beide Methoden sind gleich gut"}$$

Wenn die Untersuchungsergebnisse unter der Voraussetzung von  $H_0$  nur eine *geringe* Wahrscheinlichkeit haben, werden wir  $H_0$  –entsprechend unserer Beweisabsicht– zugunsten der folgenden Alternativhypothese verwerfen:

$H_1$ : "Die Angioplastie ist der Medikamentenbehandlung überlegen"

### **Grundgesamtheit, Stichprobe, Urnenmodell unter Annahme von $H_0$**

Der folgende zentrale Gedanke mag Sie befremden, er ist aber eine unmittelbare Konsequenz aus der Nullhypothese: Gemäss  $H_0$  wären so oder so 7 der 100 Patienten (= Grundgesamtheit) gestorben. Die 55 mit Angioplastie behandelten Patienten können wir als eine Stichprobe betrachten, in die *zufällig* 2 dieser 7 sterbenden Patienten geraten sind. Hier ist die oben erwähnte Randomisierung wichtig. Dadurch haben wir gesorgt, dass jeder Patient die gleiche Chance hat, in die Angioplastiegruppe zu kommen.

So betrachtet, können wir die Situation in das nebenstehende Urnenmodell (Fig. 6.1) übersetzen: Die Urne enthält 100 Kugeln, nämlich 7 schwarze und 93 weisse.

Die *Stichprobe* (die Gruppe Angioplastie) enthält 55 Kugeln, nämlich 2 schwarze und 53 weisse.

Zusammengefasst gilt:

$$N=100, \quad S=7, \quad n=55, \quad s=2$$

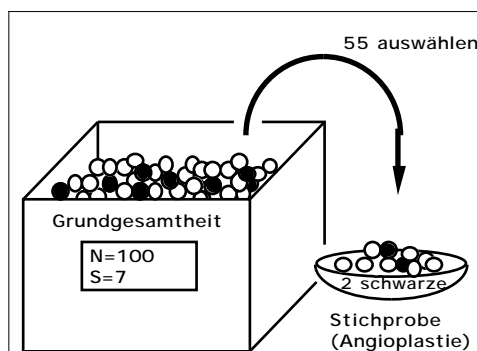


Fig. 6.1: Das Urnenmodell der Vierfeldertafel unter Annahme der Nullhypothese  $H_0$ .

### **Testgrösse und Verwerfungsbereich**

Die Testgrösse haben wir schon mehrmals umgangssprachlich erwähnt:

$Z$  = „Anzahl sterbender Patienten in der Stichprobe (Angioplastie)“

Mögliche Werte von  $Z$  sind 0, 1, ..., 7. Wenn  $H_0$  zutrifft, sollte der Wert von  $Z$  etwa 3 oder 4 sein (die beiden Gruppen sind ja ungefähr gleich gross). Grosse Abweichungen nach unten sind verdächtig und sprechen gegen  $H_0$  und für  $H_1$ . Was heisst "grosse" Abweichungen? Würden wir wie bis anhin vorgehen, so würden wir jetzt (z. B. durch planmässiges Probieren) einen Verwerfungsbereich  $V = \{0, 1, \dots, g\}$ , d. h. eine Zahl  $g$  so berechnen, dass wir  $H_0$  nur mit einer Irrtumswahrscheinlichkeit 5% verwerfen würden.

Wenn man –wie hier– den Wert von  $Z$  schon kennt ( $Z = 2$ ), geht man oft umgekehrt vor, und man berechnet zum Verwerfungsbereich

$$V = \{0, 1, 2\}$$

die Irrtumswahrscheinlichkeit. Ergibt sich –wider Erwarten– ein Wert  $> 5\%$ , so könnten wir nicht mehr von einem signifikanten Ergebnis sprechen.

Wie entscheidet eine Computersimulation unsere Frage?

## 6.2 Die Ergebnisse einer Computersimulation

### Was wurde simuliert?

Eine Urne enthält 100 Kugeln (Fig. 6.1), 93 weisse und 7 schwarze. Es wird eine Stichprobe vom Umfang 55 gezogen und die Anzahl  $z$  der schwarzen Kugeln gezählt ( $z$  = Wert der Testgrösse  $Z$ ).

### Wie wurde simuliert?

Die Simulation dieser Stichprobe ist etwas schwieriger zu programmieren als im ersten Fall. Sie finden eine ausführliche Darstellung zweier Simulationsprogramme (Taschenrechner und EXCEL-Makro) im Kap. II des Anhanges S. Ihre Lehrperson wird Sie orientieren, ob bzw. wie weit Sie sich damit befassen sollen.

### Die Häufigkeitstabelle

Die Testgrösse  $Z$  kann die Werte  $z=0, 1, \dots, 7$  annehmen. Bei einer Serie von 100 Simulationen mit dem Taschenrechner (vgl. Kap. II.1) sind diese Werte mit den folgenden absoluten Häufigkeiten  $f_z$  aufgetreten:

$z=$	0	1	2	3	4	5	6	7
$f_z=$	0	1	11	22	29	26	10	1

### Auswertung der Simulation



Mit einer gewissen Ernüchterung müssen wir feststellen, dass Stichproben mit 2 oder weniger schwarzen Kugeln überhaupt nicht selten sind, sondern mit der relativen Häufigkeit 12% auftreten. Mit dem Verwerfungsbereich

$$V=\{0, 1, 2\}$$

würden wir also die Nullhypothese  $H_0$  mit einer Irrtumswahrscheinlichkeit von schätzungsweise über 10% verwerfen. In der Statistik betrachtet man ein solches Resultat als *nicht signifikant*. Man würde vorläufig nichts entscheiden und weiter untersuchen.

Gleichzeitig müssen wir erkennen, dass man sich bei Vierfeldertafeln nicht zu stark auf den sog. gesunden Menschenverstand verlassen kann.

### Klartext für die Ärzte

Wenn beide Methoden (Angioplastie, Medikamente) gleich gut sind, müssen die Ärzte mit einer relativ grossen Wahrscheinlichkeit damit rechnen, dass obige Untersuchungsergebnis („nur“ zwei Tote bei Angioplastie) oder ein noch besseres zu beobachten. Es spricht mit anderen Worten nicht so stark für die Angioplastie und gegen die Medikamente. Allein aufgrund dieser (umfangmässig eher kleinen) Untersuchung lässt sich – immer aus der Sicht

der Statistik – wohl kaum vertreten, Herzinfarktpatienten über evtl. grosse Strecken zu transportieren, damit eine Angioplastie vorgenommen werden kann anstelle einer überall möglichen medikamentösen Behandlung.



### Aufgabe 6.1

Gehen Sie davon aus, dass in der obigen Vierfeldertafel die Grössen der beiden Behandlungsgruppen unverändert sind (55 und 45). Wie müssten aufgrund der Computersimulation die 7 toten Patienten verteilt sein, damit wir auf dem 5%-Niveau die Angioplastie als signifikant besser als die Medikamentenbehandlung akzeptieren würden? Wie muss mit anderen Worten der Verwerfungsbereich gewählt werden?

## 6.3 Wie gross ist $\alpha$ wirklich?

Aufgrund der Simulation sind wir nun wohl skeptisch, ob wir die Nullhypothese

 $H_0$ : "Beide Methoden sind gleich gut"

zugunsten der Alternativhypothese:

 $H_1$ : "Die Angioplastie ist der Medikamentenbehandlung überlegen"

verwerfen können. Voraussetzung wäre, dass die Testgrösse  $Z$  mit einer Irrtumswahrscheinlichkeit 5% in den Verwerfungsbereich



$V = \{0, 1, 2\}$

fallen würde.

Berechnen wir :

$$\begin{aligned}
 &= P(Z \in V) \\
 &= P(Z = 0 \text{ oder } Z = 1 \text{ oder } Z = 2) \\
 &= \sum_{s=0}^2 P(Z=s)
 \end{aligned}$$

$P(Z=s)$  ist die Wahrscheinlichkeit, dass in die Stichprobe mit 55 Kugeln gerade  $s$  der 7 schwarzen Kugeln und  $55-s$  der 93 weissen Kugeln geraten sind. Dieses Problem haben wir im Kapitel 3 behandelt; es gilt:

$$P(Z = s) = \frac{\binom{7}{s} \binom{93}{55-s}}{\binom{100}{55}}$$

Jetzt können wir auf dem Taschencomputer berechnen (Fig. 6.2):

$$= \sum_{s=0}^2 \frac{\binom{7}{s} \binom{93}{55-s}}{\binom{100}{55}} = 14.4\%$$

$$\sum_{s=0}^2 \frac{nCr(7,s) \cdot nCr(93,55-s)}{nCr(100,55)} = 1.44...E-1$$

Fig. 6.2: Irrtumswahrscheinlichkeit

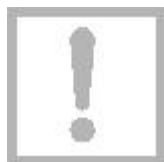
Die Rechnung bestätigt also die Ergebnisse der Simulation: Die Untersuchung spricht, immer aus der Sicht der Statistik, nicht signifikant zugunsten der Angioplastie.

### **Zusatzinformation: Beide Methoden sind gleich gut**

Eine spätere grosse Studie mit über 3000 Patienten hat den Vorteil der Angioplastie tatsächlich *nicht* bestätigen können (New England Journal of Medicine 335, 1253-1260; 1996). Während des Spitalaufenthalts nach dem Infarkt starben 5.6% der mit Medikamenten und 5.5% der mit der Angioplastie behandelten Patienten, was kein signifikanter Unterschied ist.

### **Der Verwerfungsbereich für das 5%-Niveau**

Abschliessend wollen wir nun doch noch abklären, bei welchen Untersuchungsergebnissen wir die Überlegenheit der Angioplastie als erwiesen erachtet hätten. Dazu berechnen wir die Irrtumswahrscheinlichkeit für den Verwerfungsbereich  $V=\{0, 1\}$ :



$$= \frac{\binom{7}{0} \binom{93}{55}}{\binom{100}{55}} + \frac{\binom{7}{1} \binom{93}{54}}{\binom{100}{55}} = 3.08\%$$

Auf einen signifikanten Unterschied könnten wir also nur schliessen, wenn von den mit Angioplastie behandelten Patienten höchstens einer gestorben wäre. Auch dieses Resultat konnten wir bereits aufgrund der Simulation vermuten.

### **Schlussfolgerungen**

Dieses ausführlich besprochene Beispiel zeigt mit einiger Deutlichkeit, dass der sog. gesunde Menschenverstand nicht ausreicht, um Vierfeldertafeln mit relativ kleinen Zahlen schlüssig zu beurteilen. Die Unterschiede müssen ziemlich extrem ausfallen, damit sie signifikant sind. In der abschliessenden Lernkontrolle sehen Sie, wie die Entscheidung ausfällt, wenn die Anzahl der Patienten wesentlich grösser ist. Aber meistens stehen den Ärzten eines (auch grossen) Spitals nicht beliebig viele Patienten "zur Verfügung"....

Prüfen Sie anhand der folgenden Lernkontrolle, ob Sie fähig sind, eine Vierfeldertafel statistisch auszuwerten. Melden Sie sich anschliessend zum Kapiteltest.





### **Lernkontrolle: Noch mehr Herzinfarkte**

**a**

*Multiplizieren Sie in der Vierfeldertafel auf Seite 6.1 alle Zahlen mit 10 und prüfen Sie nach, ob jetzt der Unterschied zwischen den beiden Methoden signifikant ist.*

**b**

*Gehen Sie davon aus, dass in der Vierfeldertafel von Teilaufgabe a die Grössen der beiden Behandlungsgruppen unverändert sind (550 und 450). Wie könnten im Extremfall die 70 verstorbenen Patienten verteilt sein, damit wir auf dem 5%-Niveau die Angioplastie noch als signifikant besser als die Medikamentenbehandlung akzeptieren würden?*

## Lösungen der Aufgaben zu Kapitel 6

### Lösung Aufgabe 6.1

Die Simulationsergebnisse deuten darauf hin, dass eine Vierfeldertafel mit *höchstens* einem Toten unter den mit Angioplastie behandelten Patienten signifikant für die Angioplastie sprechen würde. Zum Verwerfungsbereich  $V=\{0, 1\}$  gehört nämlich ein von ungefähr 1%.

### Lösung der Lernkontrolle

a

Es geht um die statistische Auswertung der folgenden (hypothetischen) Vierfeldertafel

Behandlungsmethode→ Erfolg ↓	Angioplastie	Medikamente	Summe
Patient überlebt	530	400	930
Patient stirbt	20	50	70
Summe	550	450	1000

Weil die relativen Anteile in den beiden Vierfeldertafeln gleich gross sind, könnte man vermuten, dass auch die Wahrscheinlichkeiten unverändert bleiben. Wie sich gleich zeigen wird, stimmt dies überhaupt nicht.

$H_0$ ,  $H_1$  und  $Z$  sind gleich oder analog definiert wie im Text. Wir bestimmen nun für den Verwerfungsbereich  $V=\{0, 1, \dots, 20\}$  die Wahrscheinlichkeit für den Fehler 1. Art:

$$= \sum_{s=0}^{20} \frac{\binom{70}{s} \binom{930}{550-s}}{\binom{1000}{550}} = 0.000003$$

$$\sum_{s=0}^{20} \frac{\binom{nCr(70,s)}{nCr(930,550-s)}}{\binom{nCr(1000,550)}} = 0.000003$$

Wenn also beide Methoden gleichwertig wären, würde sich diese oder eine noch extremere Vierfeldertafel nur mit verschwindend kleiner Wahrscheinlichkeit ergeben. Die Vierfeldertafel spricht hochsignifikant für die Angioplastie.

b

Es ist nur eine Angelegenheit der Geduld, den grössten Verwerfungsbereich  $V$  zu bestimmen, für den gerade noch 5% gilt. Die folgenden Ergebnisse erhalten Sie, indem Sie jeweils im Taschenrechner den oberen Summationsindex ändern:

$$V = \{0, 1, \dots, 30\} \quad = 2.33\%$$

$$V = \{0, 1, \dots, 35\} \quad = 22.7\%$$

Die ersten beiden Versuche zeigen, dass 30 sterbende Patienten in der Angioplastiegruppe "viel" zu klein und 35 "viel" zu gross sind. Wir probieren mit Zahlen dazwischen:

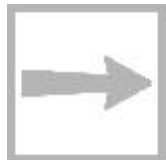
$$V = \{0, 1, \dots, 31\} \quad = 4.09\%.$$

$$V = \{0, 1, \dots, 32\} \quad = 6.78\%.$$

Offenbar erfüllt der Verwerfungsbereich  $V = \{0, 1, \dots, 31\}$  die gesuchten Anforderungen. Die folgende Vierfeldertafel würde mit anderen Worten gerade noch einen signifikanten Unterschied zwischen den beiden Methoden signalisieren:

Behandlungsmethode	Angioplastie	Medikamente
Erfolg		
Patient überlebt	519	411
Patient stirbt	31	39

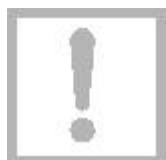
## 7

**Test mit Binomialverteilung: Meinungsumfragen****Problemstellung: Wie viele stimmen ja?**

Der Vorstand einer grossen Gewerkschaft möchte deren Statuten in einem wichtigen Punkte ändern. Über die Änderungen müssen die Mitglieder abstimmen (sog. Urabstimmung). Der Änderungsvorschlag gilt als angenommen, wenn mehr als zwei Drittel der Stimmen diese Änderung akzeptieren (Ja-Stimmen).

Mit einer Meinungsumfrage möchte der Vorstand zunächst die Chancen seines Antrages abklären. In der Mitgliederkartei werden völlig zufällig 600 Mitglieder ausgewählt und telefonisch befragt. Nur 480 davon sind an der Frage überhaupt interessiert und würden abstimmen, und zwar 340 mit ja und 140 mit nein.

Soll der Vorstand in dieser Situation die grossen Kosten und Umtriebe der Abstimmung „riskieren“? Er überlegt: Zwei Drittel von 480 sind 320. Der Anteil der Ja-Stimmen in der Stichprobe liegt nur wenig über diesem erforderlichen Minimum. Man muss durchaus damit rechnen, dass nicht ganz  $\frac{2}{3}$  der Mitglieder die Änderung bejahen und nur zufällig mehr als 320 Ja-Stimmer in die Stichprobe geraten sind.

**Hier müssen die StatistikerInnen entscheiden!**

Die Frage lautet: Wie gross ist die Wahrscheinlichkeit, dass der Anteil  $p$  der Ja-Stimmen (in einer zukünftigen Abstimmung) zwar höchstens  $\frac{2}{3}$  ist, sich aber in der Meinungsumfrage trotzdem ein positives Ergebnis der obigen Art ergibt? Wenn  $p$  klein ist (z.B. 5%), so wird man eben die "vielen" Ja-Stimmen nicht diesem unwahrscheinlichen Ereignis zuschreiben, sondern umgekehrt annehmen, dass  $p > \frac{2}{3}$ ...

Der Vorstand wird dann die Abstimmung durchführen, und – wie wir inzwischen wissen – mit der Wahrscheinlichkeit das Risiko eingehen, einen Fehler 1. Art zu begehen...

**Die Lernziele dieses Kapitels**

In diesem Kapitel erfahren Sie, wie Ergebnisse von Meinungsumfragen und ähnlich gearteten Stichproben aus der Sicht der Statistik beurteilt werden können. Insbesondere lernen Sie, die zugrunde liegenden Grundgesamtheiten in ein *neues Urnenmodell* zu übersetzen. Für die Berechnung von  $p$  und  $q$  müssen Sie Ihre Kenntnisse über die *Binomialverteilung* auffrischen.

Die Frage lautet: Wie gross ist die Wahrscheinlichkeit, dass unter 480 Stimmende wenigstens 340 Ja-Stimmer geraten, obwohl in einer zukünftigen Abstimmung höchstens  $\frac{2}{3}$  ein Ja einlegen würden.

## 7.1 Ein neues Urnenmodell

### Die Grundgesamtheit

In unserem Beispiel geht es um diejenigen Gewerkschaftsmitglieder, die an der Abstimmung teilnehmen (werden?, wollen?). Wie früher denken wir uns jedes dieser Mitglieder durch eine Kugel in der Urne repräsentiert: Sie bilden die Grundgesamtheit. Den Ja-Stimmern ordnen wir schwarze Kugeln zu.

Sicher fällt Ihnen ein grosser Unterschied gegenüber den bisherigen Beispielen auf: Die Anzahl  $N$  der Kugeln in der Urne ist nicht (oder nur höchst ungefähr) bekannt. Deshalb macht es auch keinen Sinn, wie bisher nach der Anzahl  $S$  der schwarzen Kugeln zu fragen. Wir müssen das Urnenmodell und die Fragestellung ein wenig modifizieren.

### Der Ja-Stimmer-Anteil

Statt für die Anzahl  $S$  interessieren wir uns für den *Anteil*  $p=S/N$  der schwarzen Kugeln in der Grundgesamtheit, also gerade für diejenige Zahl, die eben entscheidet, ob die Vorlage des Vorstandes bei der Urabstimmung angenommen wird.

### Die Meinungsumfrage als Stichprobe

In der Meinungsumfrage wurden 480 Mitglieder gefunden, die abstimmen wollen. In unserem Urnenmodell entspricht dem eine Stichprobe mit dem Umfang  $n=480$ . Sie enthält  $s=340$  schwarze und 140 weisse Kugeln; der Anteil der schwarzen beträgt:

$$\frac{s}{n} = \frac{340}{480} = 0.71.$$

Es stellt sich nun die Frage, ob man aufgrund dieses Ergebnisses mit grosser Wahrscheinlichkeit auf einen Anteil  $p > 2/3 = 0.67$  schliessen kann (Fig. 7.1).

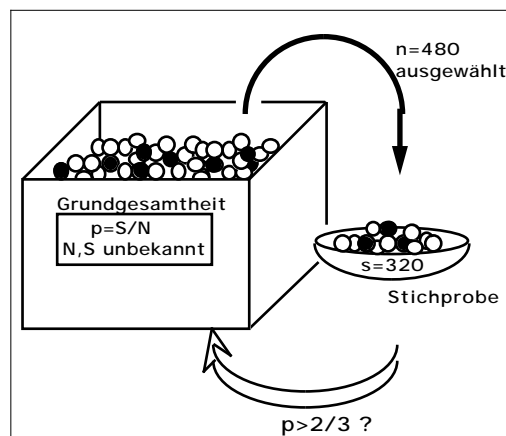


Abb. 7.1: Urnenmodell für die Meinungsumfrage.

### Wahrscheinlichkeitsrechnung im neuen Urnenmodell

Wir haben vorerst die folgende Grundaufgabe zu lösen: Wie gross ist die Wahrscheinlichkeit für eine Stichprobe mit  $s$  schwarzen Kugeln, wenn  $p$  gegeben ist, z. B.  $p = 2/3$ ?

Weil die Anzahl  $N$  der Kugeln in der Urne *unbekannt* ist, lässt sich das Problem nicht exakt (mit den früher besprochenen Formeln) lösen. Weil jedoch  $N$  als *gross* vorausgesetzt wird, können wir es *näherungsweise* aufgrund der folgenden Idee lösen: Wenn wir die *erste* Kugel der Stichprobe ziehen, so wird sie mit der Wahrscheinlichkeit  $p = 2/3$  schwarz sein. Bei den folgenden

Zügen ist die Wahrscheinlichkeit nicht mehr exakt  $2/3$ , sondern geringfügig grösser oder kleiner, je nach dem Ausgang der vorherigen Züge. Wir machen einen kleinen Fehler, wenn wir annehmen, dass in *jedem* Zug die Wahrscheinlichkeit für eine schwarze Kugel  $p = 2/3$  ist:

$$P(\text{"schwarze Kugel im } k\text{-ten Zug"}) = p = \frac{2}{3}$$

### **Wie gross ist der Fehler dieser Approximation?**

Anhand eines Zahlenbeispiels wollen wir uns ein Vorstellung von der "Qualität" der Näherung machen: Sei  $N=90'000$ ,  $S=60'000$  und seien bereits 479 Kugeln, nämlich 320 schwarze und 159 weisse gezogen worden. Dann gilt für die letzte Ziehung:

$$P(\text{"schwarze Kugel im 480. Zug"}) = \frac{60'000-320}{90'000-479} = 0.666659.$$

Der Wert unterscheidet sich also um weniger als  $1/100'000$  vom exakten Wert  $2/3 = 0.6666\dots$ . Solche Fehler sind im Hinblick auf die Fragestellung bedeutungslos.

Gestützt auf diese offenbar vertretbare Näherung gelingt es uns in den folgenden Abschnitten, unsere Fragestellung sowohl durch Simulation als auch rechnerisch zu bewältigen.



### **Aufgabe 7.1**

*Der Gewerkschaftsvorstand ist natürlich daran interessiert, dass möglichst viele Mitglieder abstimmen. Kann er aufgrund der oben beschriebenen Meinungsumfrage hoffen, dass mehr als 75% aller Mitglieder an die Urne gehen?*

*Illustrieren Sie diese Fragestellung anhand eines gezeichneten Urnenmodells analog zu Fig. 7.1.*

## **7. 2      Statistische Begriffe bei der Meinungsumfrage**

### **Die Nullhypothese**

Der Vorstand möchte eigentlich beweisen, dass das Umfrageergebnis von 340 Ja unter 480 Stimmen auf  $p > 2/3$  hinweist. Nun kann man bekanntlich in der Statistik nur *indirekt* „beweisen“, indem man versucht, eine Nullhypothese zu verwerfen. Hier möchte der Vorstand widerlegen, dass der Anteil  $p$  der schwarzen Kugeln höchstens  $2/3$  ist. Die Nullhypothese lautet also:

$$H_0: „p \leq \frac{2}{3}“$$

### **Die Alternativhypothese**

Wenn es gelingt,  $H_0$  zu verwerfen, wird die folgende Alternativhypothese akzeptiert:

$$H_1: „p > \frac{2}{3}“$$

### Die Testgrösse

Entscheidend für unser Urteil ist die Anzahl der schwarzen Kugeln in der Stichprobe. Unsere Testgrösse ist also:

$$Z = \text{"Anzahl der schwarzen Kugeln in der Stichprobe"}$$

### Der Verwerfungsbereich

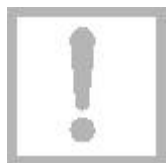
Wenn die Nullhypothese zutrifft, sollte der Wert von  $Z$  nicht „wesentlich“ grösser als  $2/3 \cdot 480 = 320$  sein. Bei unserem Test möchte der Vorstand  $H_0$  verwerfen, wenn  $Z$  einen Wert von 340 oder grösser annimmt. Er „liebäugelt“ mit anderen Worten mit dem Verwerfungsbereich

$$V = \{ 340, 341, \dots, 480 \}$$

Diese Wahl lässt sich aber nur rechtfertigen, wenn der Fehler 1. Art eine genügend kleine Wahrscheinlichkeit hat.

### Der Fehler 1. Art - Irrtumswahrscheinlichkeit $\alpha$

Wenn man  $H_0$  *irrtümlicherweise* verwirft, spricht man, wie Sie wissen, vom Fehler 1. Art. Er ergibt sich dann, wenn der Wert von  $Z$  so gross ist, dass er in den Verwerfungsbereich  $V$  fällt, obwohl  $p < 2/3$  ist. Die Wahrscheinlichkeit dafür hängt natürlich vom tatsächlichen Wert von  $p$  ab; unter der Irrtumswahrscheinlichkeit versteht man bekanntlich den grösstmöglichen Wert. Hier gilt offenbar:



$$= P(Z \in V, \text{ wenn } p = \frac{2}{3})$$

oder

$$= P(Z \geq 340, \text{ wenn } p = \frac{2}{3})$$

### Der Fehler 2. Art

Wenn man  $H_0$  *irrtümlicherweise akzeptiert*, spricht man bekanntlich vom Fehler 2. Art. Er ergibt sich immer dann, wenn der Wert von  $Z$  so klein ist, dass er nicht in den Verwerfungsbereich  $V$  fällt, obwohl  $p > 2/3$  ist, d. h. obwohl die Alternativhypothese  $H_1$  zutrifft. Die Wahrscheinlichkeit dafür hängt ebenfalls von  $p$  ab. Wir können sie z. B. für  $p=75\%$  angeben:



$$= P(Z \notin V, \text{ wenn } p = 75\%)$$

oder:

$$= P(Z < 340, \text{ wenn } p = 75\%)$$

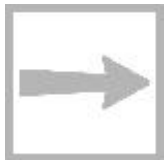


### Aufgabe 7.2

Der Gewerkschaftsvorstand möchte wissen, ob er aufgrund der Meinungsumfrage mit einer Stimmbeteiligung von mehr als 75% rechnen kann.

Formulieren Sie für diese ganz neue und anspruchsvolle Testaufgabe die Bedeutungen oder Definitionen von  $H_0$ ,  $H_1$ ,  $Z$ ,  $V$ ,  $\alpha$  und  $\beta$ .

## 7.3 Bestimmung von $\alpha$ und $\beta$ durch Simulation



Die Simulation ist wieder eine sehr bequeme Methode, um  $\alpha$  und  $\beta$  zu ermitteln. Darüber hinaus zeigt sie uns unmittelbar die Bedeutung dieser Grössen, beispielsweise von  $\alpha$ : Wir simulieren 100 Stichproben mit dem Umfang 480 und zählen, wie viele davon 340 oder mehr schwarze Kugeln enthalten ( $Z \geq V$ ). Diese Anzahl entspricht gerade  $\alpha$  in Prozenten.

### Wie wurde simuliert?

Wenn Sie sich für die Programmierung dieser einfachen Simulation interessieren, können Sie das Kapitel III im Anhang S studieren. Ihre Lehrperson hat Sie orientiert, ob Sie sich damit befassen müssen.

Nachstehend präsentieren wir Ihnen die Ergebnisse einer Simulation in übersichtlicher grafischer Aufbereitung

### Das Histogramm von $Z$ für 100 Simulationen

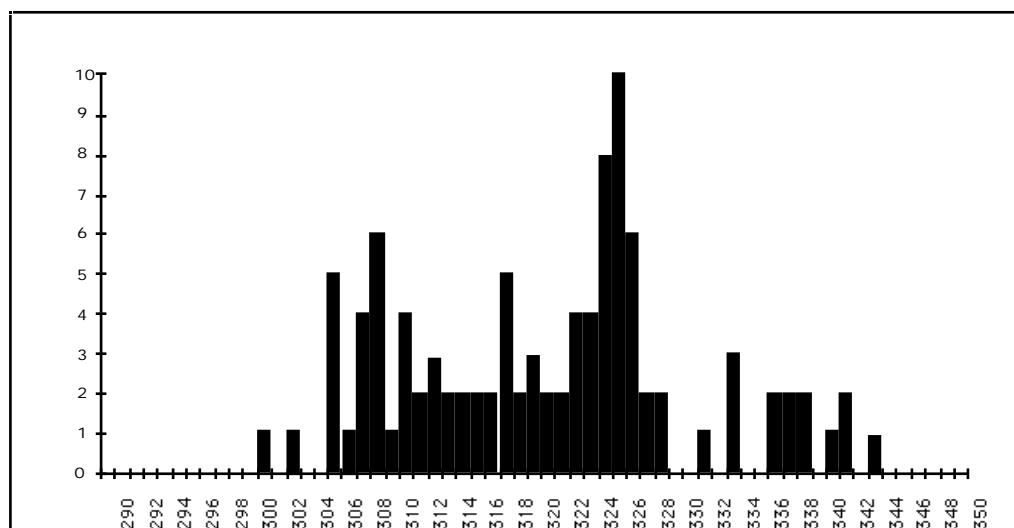


Abb. 7.2: Histogramm von  $Z$  für 100 Simulationen

### Ein Schätzwert für $\alpha$



Das Histogramm zeigt eindrücklich, dass die Werte von  $Z$  nur in der „Nähe“ von 320 verschieden von 0 sind. Insbesondere erkennen wir, dass nur 3 oder 4 der 100 Stichproben 340 oder mehr schwarze Kugeln enthielten, also zu einem Fehler 1. Art geführt hätten. Es ergibt sich somit für  $\alpha$  der Schätzwert

4%

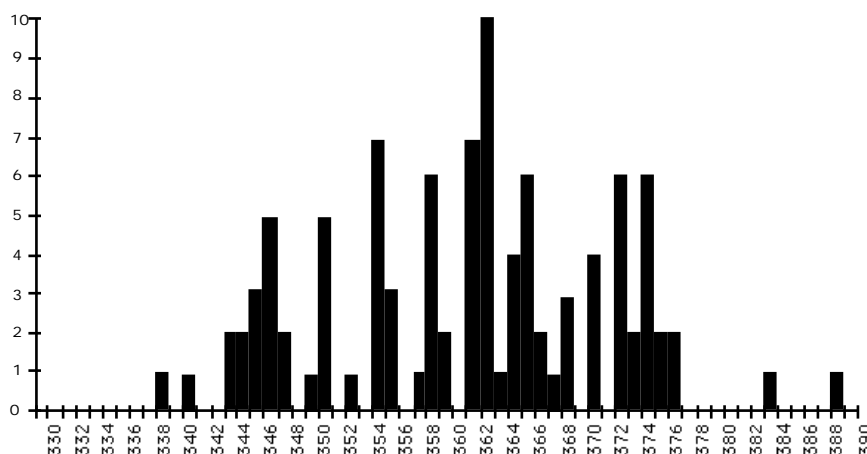
Einen Fehler 2. Art macht man, wenn man aufgrund des Tests glaubt, die Vorlage werde abgelehnt, obwohl der Anteil der Ja-Stimmer z. B. 75% ist. In der folgenden Aufgabe ermitteln Sie, wie gross schätzungsweise die Wahrscheinlichkeit für diesen Fehlentscheid ist.





### Aufgabe 7.3: Bestimmung von $\beta$

Es wurden 100 Stichproben mit dem Umfang 480 simuliert und zwar aus einer Urne mit einem Anteil von  $p=75\%$  an schwarzen Kugeln. Im nachstehenden Histogramm sind die Ergebnisse grafisch zusammengefasst. Ermitteln Sie daraus einen Schätzwert für  $\beta$  bei  $p=75\%$ .



## 7.4

### Die Binomialverteilung

Im nächsten Abschnitt wollen wir uns überlegen, wie und formelmässig vom Stichprobenumfang  $n$  und von der Wahrscheinlichkeit  $p$  abhängen. Solche Formeln ermöglichen es, einen statistischen Test zu planen, z. B. bei einem gegebenen  $p$  den Wert  $n$  so zu wählen, dass 5% ist.

Zu diesem Zweck müssen wir zunächst ein wichtiges Ergebnis aus der Wahrscheinlichkeitsrechnung in Erinnerung rufen: Die Gesetzmässigkeit der sog. *Binomialverteilung*. Dabei geht es um die folgende

#### Grundaufgabe

Gegeben ist eine Urne mit einer unbekannten, sehr grossen Anzahl von Kugeln. Der Anteil der schwarzen beträgt exakt:

$$p = \frac{S}{N} = \frac{2}{3}$$

Es wird eine Stichprobe von  $n=480$  Kugeln entnommen (Fig. 7. 3).

Wie gross ist die Wahrscheinlichkeit, dass sie  $s$  schwarze Kugeln enthält:

$$P(\text{"s schwarze"}) = ?$$

für  $s=0, 1, \dots, 480$ .

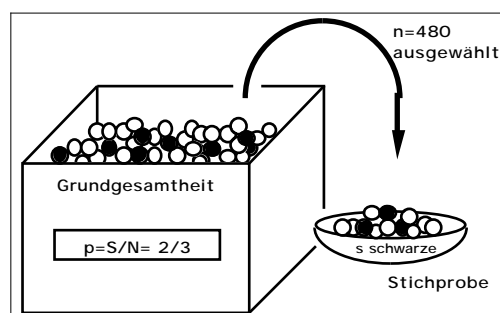


Fig. 7.3: Die Ausgangslage für die Grundaufgabe: Der Anteil  $p=S/N=2/3$  der schwarzen Kugeln wird als bekannt vorausgesetzt; die Grössen  $S$  und  $N$  sind unbekannt.

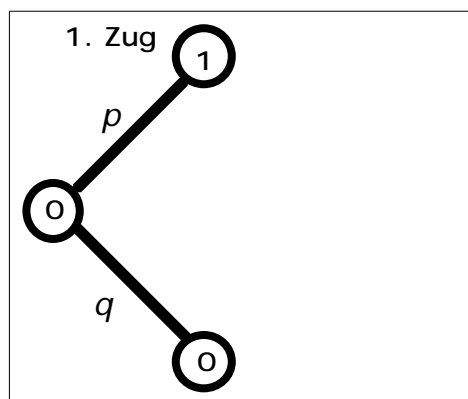
### Nochmals zur Lösungsidee

Wir nehmen an, dass die Wahrscheinlichkeit für eine schwarze Kugel stets gleich bleibt:

$$P(\text{"schwarze Kugel im k. Zug"}) = p = \frac{2}{3}$$

### Das Baumdiagramm der Stichprobenziehung

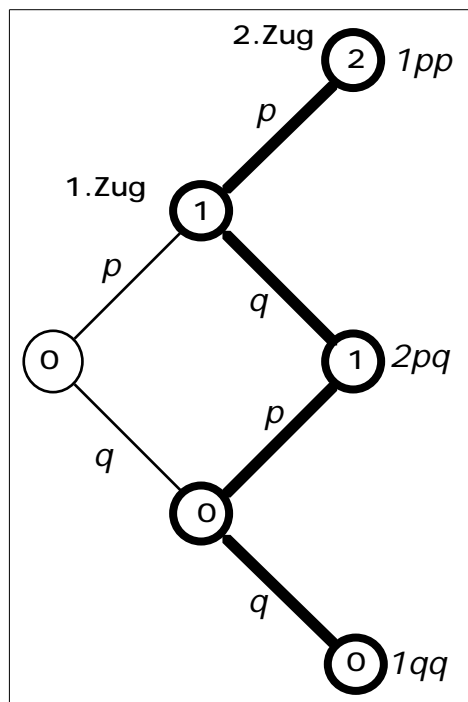
Mit Hilfe des nachfolgenden Baumdiagramms gelingt es uns, die oben formulierte Grundaufgabe zu lösen. Alle nach rechts oben weisenden Zweige entsprechen der Ziehung einer *schwarzen* Kugel (Wahrscheinlichkeit  $p$ ). Alle nach rechts unten verlaufenden Zweige gehören zu Ziehungen einer *weißen* Kugel (Wahrscheinlichkeit  $q = 1-p$ ).



#### 1. Zug

Wir *starten* mit 0 *schwarzen* Kugeln.

- Mit der Wahrscheinlichkeit  $p$  ziehen wir eine schwarze Kugel (Zweig nach oben rechts), und wir haben am Schluss 1 schwarze.
- Mit der Wahrscheinlichkeit  $q=1-p$  ziehen wir eine weiße, sodass wir schliesslich 0 schwarze haben.

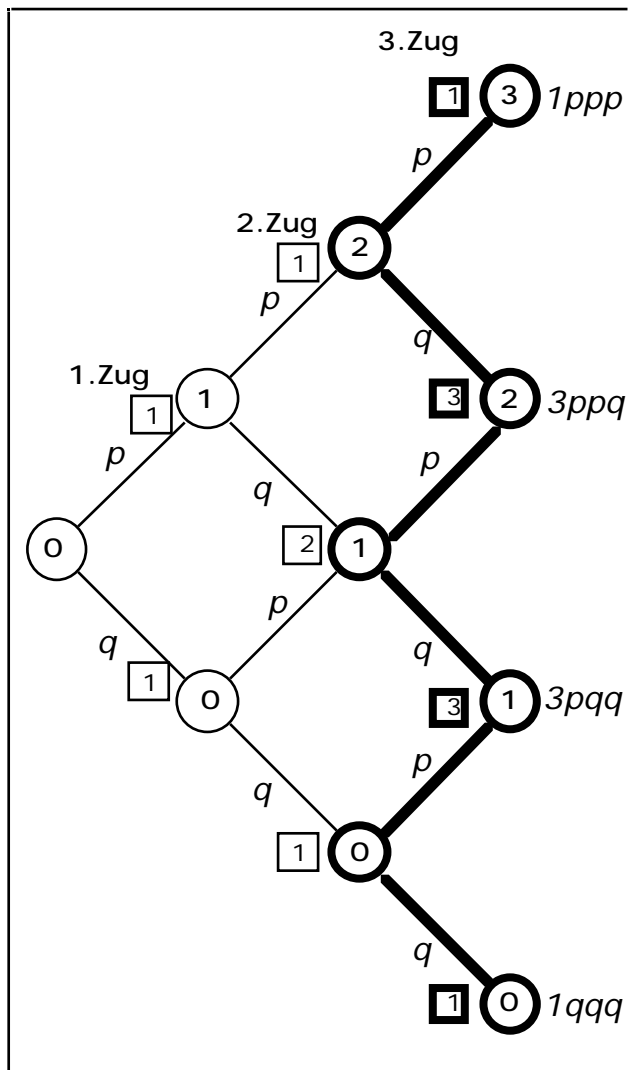


#### Zwei Züge

- Mit der Wahrscheinlichkeit  $p^2$  haben wir 2 schwarze Kugeln (1. Pfadregel für den Pfad nach oben rechts, der im Knoten 2 endet).

- Es gibt zwei verschiedene "gewinkelte" Pfade, die auf 1 schwarze Kugel führen (zum mittleren Knoten 1 rechts). Auf beiden Pfaden ist die Wahrscheinlichkeit gemäss 1. Pfadregel  $pq$ ; somit hat das Ergebnis „1 schwarze“ die Wahrscheinlichkeit  $2pq$ .

- Für „0 schwarze“ ist die Wahrscheinlichkeit  $q^2$  (Pfad nach rechts unten zum Knoten 0).



### Drei Züge

In diesem Diagramm sind zusätzlich die Anzahl Pfade angegeben, die zu einem Knoten führen (quadratisch eingerahmt).

Betrachten wir den zweitobersten Knoten rechts, welcher für das Schlussergebnis „2 schwarze“ steht:

- Es gibt 1 Pfad, der von links oben zu diesem Knoten führt...
- ... und von links unten die *Verlängerung* von 2 Pfaden, die zum vorangehenden Knoten führten.

Im ganzen gibt es also 3 (ausschliesslich nach rechts führende) Pfade zu diesem Knoten.

Auf allen drei Pfaden beträgt die Wahrscheinlichkeit gemäss 1. Pfadregel  $p^2q$ , d. h. das Ergebnis „2 schwarze in 3 Zügen“ hat die Wahrscheinlichkeit  $3p^2q$ .

Ganz analog können die angeschriebenen Wahrscheinlichkeiten für die Ergebnisse bei den drei anderen Knoten begründet werden.

### Die Anzahl Pfade als Binomialkoeffizienten

Ist Ihnen schon aufgefallen, dass die Anzahl der kürzesten Pfade (im Diagramm quadratisch eingerahmt), die zu den einzelnen Knoten führen, das Bauprinzip des Pascaldreieckes aufweisen? Diese Zahlen lassen sich, wie Sie vermutlich noch wissen, als Binomialkoeffizienten schreiben. Von oben nach unten gelesen, können wir sie also wie folgt schreiben:

$$1 = \binom{3}{3} \quad 3 = \binom{3}{2} \quad 3 = \binom{3}{1} \quad 1 = \binom{3}{0}$$

Möglicherweise gefällt Ihnen die folgende Begründung für das Auftreten der Binomialkoeffizienten besser:

- Damit man zum *obersten* Knoten rechts kommt, muss man sich in 3 möglichen Entscheidungen 3 mal für einen Zweig nach oben entscheiden.
- Damit man zum *zweitobersten* Knoten rechts kommt, muss man sich in 3 möglichen Entscheidungen 2 mal für einen Zweig nach oben entscheiden.
- etc.

### Wahrscheinlichkeiten mit Binomialkoeffizienten schreiben

Setzen wir nun die Binomialkoeffizienten bei den im Baumdiagramm berechneten Wahrscheinlichkeiten ein, so erhalten wir, immer von oben nach unten gelesen:

- $P(\text{"3 schwarze Kugeln in 3 Zügen"}) = \binom{3}{3} p^3 = \binom{3}{3} p^3 q^{3-3}$
- $P(\text{"2 schwarze Kugeln in 3 Zügen"}) = \binom{3}{2} p^2 q = \binom{3}{2} p^2 q^{3-2}$
- $P(\text{"1 schwarze Kugeln in 3 Zügen"}) = \binom{3}{1} p q^2 = \binom{3}{1} p^1 q^{3-1}$
- $P(\text{"0 schwarze Kugeln in 3 Zügen"}) = \binom{3}{0} q^3 = \binom{3}{0} p^0 q^{3-0}$

Wir können alle vier Formeln wie folgt in eine einzige zusammenfassen:

$$P(\text{"s schwarze Kugeln in 3 Zügen"}) = \binom{3}{s} p^s q^{3-s}$$

### Die Lösung der Grundaufgabe

Für die Lösung der Grundaufgabe müssen wir das dreistufige Baumdiagramm auf 480 Stufen erweitern. Wir erhalten die folgende Formel, in welcher – nota bene – die unbekannte Anzahl N nicht vorkommt :

$$P(\text{"s schwarze Kugeln in 480 Zügen"}) = \binom{480}{s} p^s q^{480-s}$$

### Binomialverteilte Zufallsgrößen

Wenn die Wahrscheinlichkeiten für die möglichen Werte einer Zufallsgrösse Z gemäss dieser Formel verteilt sind, spricht man von einer *binomialverteilten* Zufallsgrösse oder von einer *Binomialverteilung*.

#### Binomialverteilung

Ein bestimmtes Zufallsexperiment wird n-mal durchgeführt. *Jedesmal* ist die Wahrscheinlichkeit für einen sog. *Treffer* gleich p. Dann ist die Anzahl Z der Treffer in n Durchführungen binomialverteilt und es gilt mit  $q=1-p$ :

$$P(\text{"k Treffer in n Durchführungen"}) = \binom{n}{k} p^k q^{n-k}$$

oder kurz:

$$P(Z=k) = \binom{n}{k} p^k q^{n-k} \quad k=0, 1, \dots, n$$

Sie finden diese Formel in Ihrer Formelsammlung vermutlich unter dem Stichwort "Ziehen mit Zurücklegen" und dies aus folgendem Grund: Wenn Sie nach jedem Zug die gezogene Kugel wieder in die Urne zurücklegen würden, wäre das Postulat von der konstanten Wahrscheinlichkeit für eine schwarze Kugel erfüllt, und die obige Formel würde exakt gelten....



#### Aufgabe 7.4

Berechnen Sie die Wahrscheinlichkeit, in einer Stichprobe aus der Urne gemäss Fig. 7.3....

- a) ... unter insgesamt 4 Kugeln 2 schwarze zu finden (Handrechnung).
- b) ...unter insgesamt 480 Kugeln 320 schwarze zu finden (mit Taschenrechner).

## 7.5 Die Berechnung von $\alpha$ und $\beta$

Mit der eben hergeleiteten Formel können wir berechnen:



$$\begin{aligned}
 &= P(Z \leq V, \text{ wenn } H_0 \text{ gilt}) \\
 &= P(Z \leq 340, \text{ wenn } p = \frac{2}{3}) \\
 &= \sum_{s=0}^{340} P(Z = s) \\
 &= \sum_{s=0}^{340} \binom{480}{s} p^s q^{480-s}
 \end{aligned}$$

Der Taschenrechner liefert innerhalb weniger Minuten (evtl. auf approximatives Rechnen einstellen!):

$$= 0.028505 = 2.85\%$$

Damit wird das Ergebnis der Simulation bestätigt: Der Ja-Stimmenanteil ist signifikant grösser als  $2/3$ .



#### Aufgabe 7.5

Wie gross ist bei diesem Test  $\beta$ , falls der tatsächliche Ja-Stimmenanteil  $p=75\%$  beträgt?

Wenn Sie die letzte Aufgabe richtig gelöst haben, sollten Sie ohne grosse Mühe die abschliessende Lernkontrolle bewältigen können. Melden Sie sich danach zum Kapiteltest an.



### **Lernkontrolle: Voraussichtliche Stimmbeteiligung**

Die Meinungsumfrage des Gewerkschaftsvorstandes hat ergeben, dass 480 von 600 befragten Mitgliedern, in der Stichprobe also 80%, abstimmen würden.

**a**

Entscheiden Sie durch einen regelgerechten statistischen Test: Ist die Stimmbeteiligung signifikant grösser als 75%?

**b**

Bei welchen Umfrageergebnissen kann man beim Signifikanzniveau auf eine Stimmbeteiligung von mehr als 75% schliessen? Formulieren Sie Ihre Antwort mittels Verwerfungsbereich  $V$ .

**c**

Wie gross ist  $\beta$  für den eben berechneten Verwerfungsbereich  $V$ , wenn die hypothetische Stimmbeteiligung 80% betragen würde?

## Lösungen der Aufgaben zu Kapitel 7

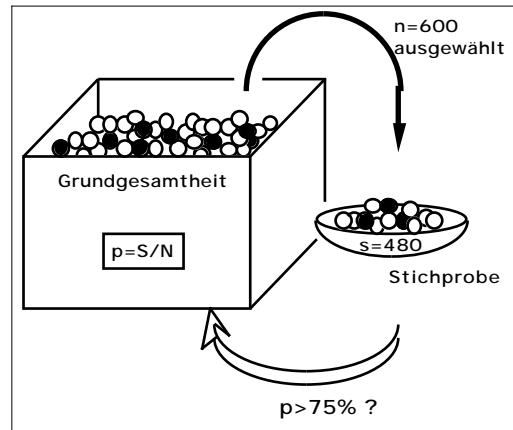
### Lösung Aufgabe 7.1

Die Grundgesamtheit bilden die Mitglieder der Gewerkschaft. Jedem Mitglied entspricht eine Kugel in der Urne (Anzahl  $N$ ; ist in der Praxis nicht unbedingt exakt bekannt).

Mitgliedern, die abzustimmen gedenken, ordnen wir schwarze Kugeln zu (Anzahl  $S$ ; gesuchter Anteil  $p=S/N$ ).

Die Stichprobe hat den Umfang 600 und enthält 480 schwarze Kugeln.

Gefragt ist, ob man daraus auf  $p > 75\%$  schliessen kann.



### Lösung Aufgabe 7.2

Das Urnenmodell für den Test haben wir in Aufgabe 7.1 bereits gezeichnet. Mit  $p = S/N$  bezeichnen wir die unbekannte Stimmbeteiligung.

Der Vorstand möchte die Nullhypothese:

$$H_0: "p = 75\%"$$

widerlegen zugunsten der Alternativhypothese:

$$H_1: "p > 75\%"$$

Entscheidend ist der Wert der Testgröße:

$$Z = \text{Anzahl der schwarzen Kugeln in der Stichprobe vom Umfang 600}$$

Als Verwerfungsbereich zieht der Vorstand

$$V = \{480, 481, \dots, 600\}$$

in Betracht, was allerdings nur sinnvoll ist, wenn daraus ein 5% resultiert:

$$= P(Z \geq 480, \text{ wenn } p=75\%)$$

Bei dieser Wahl von  $V$  beträgt die Wahrscheinlichkeit für den Fehler 2. Art bei z. B.  $p=80\%$ :

$$= P(Z < 480, \text{ wenn } p=80\%)$$

**Lösung Aufgabe 7.3**

Es gilt:

$$= P(Z < 340, \text{ wenn } p=75\%)$$

Das Histogramm zeigt, dass nur 1 oder 2 (die Beschriftung der horizontalen Achse ist nicht eindeutig) von 100 Stichproben weniger als 340 schwarze Kugeln enthalten. Es ergibt sich so der Schätzwert:

$$2\% \text{ für } p=75\%$$

**Lösung Aufgabe 7.4**

$$\text{a) } P(\text{"2 schwarze in 4 Zügen"}) = \frac{4}{2} \cdot \frac{2}{3}^2 \cdot \frac{1}{3}^2 = 6 \cdot \frac{4}{9} \cdot \frac{1}{9} = 29.6\%$$

$$\text{b) } P(\text{"320 schwarze in 480 Zügen"}) = \frac{480}{320} \cdot \frac{2}{3}^{320} \cdot \frac{1}{3}^{160} = 3.86\%$$

**Lösung Aufgabe 7.5**

$$\begin{aligned} &= P(Z \leq V, \text{ wenn } H_1 \text{ gilt}) \\ &= P(Z < 340, \text{ wenn } p=75\%) \\ &= \sum_{s=0}^{339} P(Z=s) \\ &= \sum_{s=0}^{339} \binom{480}{s} 0.75^s 0.25^{480-s} \end{aligned}$$

Nach relativ langer Rechenzeit liefert der Taschenrechner:

$$= 1.65\%$$

Die Rechenzeit kann halbiert werden, indem die Wahrscheinlichkeit des Gegenereignisses berechnet wird.

Interpretation: Die Macht des Tests bei einem Ja-Stimmenanteil von 75% beträgt über 98%, d. h. in durchschnittlich 98 von 100 Fällen würde der Vorstand aufgrund dieses Tests richtig entscheiden.



### Lösung der Lernkontrolle

a

Die statistischen Begriffe für diesen Test haben wir bereits in Aufgabe 7.2 geklärt:

$$H_0: "p = 75\%"$$

$$H_1: "p > 75\%"$$

$Z$  = Anzahl der schwarzen Kugeln in der Stichprobe vom Umfang 600

$$V = \{480, 481, \dots, 600\}$$

Damit ergibt sich für :

$$\begin{aligned} &= P(Z \geq 480, \text{ wenn } p = 75\%) \\ &= \sum_{s=480}^{600} P(Z=s) \\ &= \sum_{s=480}^{600} \binom{600}{s} 0.75^s 0.25^{600-s} \\ &= 0.2\% \end{aligned}$$

Die hypothetische Stimmbeteiligung ist also *hochsignifikant* grösser als 75%.

b

Durch Probieren ermitteln wir den Verwerfungsbereich so, dass gerade noch 5%:

Für  $V = \{465, \dots, 600\}$  ergibt sich  $\alpha = 8.47\%$ .

Für  $V = \{468, \dots, 600\}$  ergibt sich  $\alpha = 4.80\%$ .

Für  $V = \{467, \dots, 600\}$  ergibt sich  $\alpha = 5.85\%$ .

Offenbar sind Ergebnisse  $Z \geq 468$  signifikant für eine Stimmbeteiligung von über 75%.

c

Für den Verwerfungsbereich  $V = \{468, \dots, 600\}$  hat bei  $p = 80\%$  den Wert:

$$\begin{aligned} &= P(Z < 468, \text{ wenn } p = 80\%) \\ &= \sum_{s=0}^{467} P(Z=s) \\ &= \sum_{s=0}^{467} \binom{600}{s} 0.8^s 0.2^{600-s} \\ &= 10.2\% \end{aligned}$$

## 8

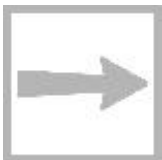
## Ein zweiseitiger Vorzeichentest: Welche Sorte anpflanzen?

### Problemstellung

In vielen entwickelten Ländern erzeugt die Landwirtschaft Überschüsse bei gewissen Produkten (Milch, Getreide, Früchte etc.), deren Vermarktung kostspielig sein kann. Deshalb versuchen weitsichtige Bauern auf neue Erzeugnisse umzustellen. Beispielsweise ist die aus China stammende Kiwifrucht um 1960 in Neuseeland eingeführt und bald schon grossmasstäblich kultiviert worden. Inzwischen wird die beliebte Frucht auch in Europa erfolgreich angepflanzt.

Wie Sie sich denken können, gedeiht eine Pflanze nicht ohne weiteres in einem neuen Klima mit einem anderen Boden. In der Regel stehen aber verschiedene Sorten einer bestimmten Pflanze zur Auswahl und man kann in Vorversuchen abklären, welche sich für die neue Umgebung überhaupt eignen.

Wir denken uns nun, dass die zwei vielversprechendsten Sorten A und B bezüglich ihres Ertrages untersucht werden sollen. Zu diesem Zwecke werden insgesamt 20 Parzellen jeweils halb mit Sorte A und halb mit Sorte B bepflanzt. Nach der Ernte soll entschieden werden, ob sich die Erträge der beiden Sorten signifikant unterscheiden.



### Die Lernziele dieses Kapitels

In diesem Kapitel erfahren Sie, wie die statistische Auswertung dieses Versuches *geplant* werden kann. Insbesondere lernen Sie, was man unter einem *Vorzeichentest* - eine Anwendung der Binomialverteilung - versteht.

Erstmals begegnen Sie einem sog. *zweiseitigen* Test, bei welchem der Verwerfungsbereich nicht wie bis anhin aus *einem*, sondern aus *zwei getrennten* Bereichen besteht.

Der Sortenvergleich ist *ein* Beispiel eines häufigen und wichtigen Test-Schemas: Gegeben sind zwei natürlich *gepaarte* Stichproben, hier die Erträge der beiden Sorten. Man möchte abklären, ob sich die beiden Stichproben signifikant unterscheiden.

Sie haben nach jedem Abschnitt in diesem Kapitel die Gelegenheit, anhand einer weiteren Testsituation dieser Art zu prüfen, ob Sie die Erklärungen verstanden haben. Nachstehend die Aufgabenstellung, die Sie parallel zum Sortenvergleich bearbeiten werden.



### Aufgabe 8.0: Geburtsgewicht und Intelligenz

Zwei Forscher stellten sich die Frage, ob die Ernährung im Mutterleib einen Einfluss auf die spätere Intelligenz hat. Sie untersuchten zu diesem Zwecke 14 Paare von eineiigen Zwillingen, indem sie Jahre nach der Geburt deren IQ (=Intelligenzquotient) ermittelten. Hier die Untersuchungsergebnisse:

Nr. des Paares	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IQ des schweren Zwilling: X	97	79	10 0	100	10 0	124	95	80	91	108	91	90	10 4	119
IQ des leichten Zwilling: Y	97	70	10 1	106	85	123	84	70	84	106	97	90	92	104

Vergleichen Sie dieses neue Testproblem mit dem Sortenvergleich und notieren Sie die grundsätzlichen Gemeinsamkeiten. (Aus A. Engel: Stochastik, 1987, Ernst Klett Verlag)

## 8.1 Die Nullhypothese beim Sortenvergleich

In den folgenden Abschnitten werden wir die statistischen Fachbegriffe  $H_0$ ,  $H_1$  etc. für den Test "Sortenvergleich" definieren. Parallel dazu werden Sie den gleichen Begriff für den Test "Geburtsgewicht und IQ" bestimmen.

Wir beginnen mit der Nullhypothese: Der Test hat zum Ziel, einen allfälligen Ertragsunterschied zwischen den beiden Sorten aufzudecken. Die zukünftigen Testergebnisse sollten nach Möglichkeit die folgende Nullhypothese widerlegen:

$H_0$ : "Sorten A und B sind gleich ertragsreich"



### Aufgabe 8.1

Im Beispiel "Geburtsgewicht und IQ" interessiert uns, ob es einen signifikanten Unterschied zwischen den IQ's des schweren und des leichten Zwillinges gibt.

Formulieren Sie für diesen Test die Nullhypothese.

## 8.2 Die Alternativhypothese beim Sortenvergleich

Vermutlich erwarten Sie aufgrund der bisher durchgeführten Tests, dass wir allenfalls die Nullhypothese verwerfen zugunsten einer Hypothese folgender Art: "Sorte A ist ertragsreicher als Sorte B" oder umgekehrt. Beachten Sie aber, dass wir *im voraus* keinerlei Hinweis auf einen solchen Tatbestand haben. Uns interessiert nur, ob es *überhaupt einen Unterschied* zwischen den beiden Sorten gibt. Deshalb arbeiten wir mit der folgenden Alternativhypothese:

$H_1$ : "Die Erträge der Sorten A und B sind verschieden"

Sollten wir aufgrund der Testergebnisse schliesslich  $H_1$  akzeptieren, so wird in der Regel auch klar sein, welche der beiden Sorten wir als ertragsreicher zu betrachten haben.



### Aufgabe 8.2

Formulieren Sie für den Test "Geburtsgewicht und IQ" eine Alternativhypothese.

## 8.3 Die Testgrösse beim Sortenvergleich

Wir werden nach der Ernte die Erträge auf den 20 Parzellen wägen und in einer Liste festhalten, etwa so:

Parzellennummer:	1	2	...	...	20
Ertrag der Sorte A: X	5.2	3.9	...	...	7.7
Ertrag der Sorte B: Y	4.3	4.9	...	...	8.8
<b>Vorzeichen von X-Y</b>	<b>+</b>	<b>-</b>	<b>...</b>	<b>...</b>	<b>-</b>

In der letzten Zeile haben wir die Vorzeichen der Differenzen X-Y notiert.

Wir argumentieren nun so:



Wenn gemäss  $H_0$  die Erträge der beiden Sorten *gleich* sind, so werden diese Vorzeichen in den einzelnen Parzellen zufällig auftreten und es sollten etwa gleich viele Plus- wie Minuszeichen vorkommen.

Weicht aber die Anzahl der Pluszeichen wesentlich von 10 ab, so glauben wir nicht mehr daran, dass  $H_0$  zutrifft. In diesem Fall verwerfen wir  $H_0$  und akzeptieren  $H_1$ .

Als Testgrösse verwenden wir hier offenbar:

$$Z = \text{"Anzahl der positiven Differenzen von X-Y"}$$

oder salopper:

$$Z = \text{"Anzahl der Pluszeichen"}$$

Deshalb spricht man von einem *Vorzeichentest*.



### Aufgabe 8.3

- Definieren Sie für den Test "Geburtsgewicht und IQ" die Testgrösse Z.
- Welchen Wert hat Z? Beachten Sie: Gewisse der Zwillingspaare haben den gleichen IQ, d. h. die Differenz der beiden Zufallsgrössen X und Y ist 0. Es ist üblich und zweckmässig, solche Paare beim Vorzeichentest wegzulassen, weil sie nichts zur Entscheidung beitragen.

## 8.4 Der Verwerfungsbereich beim Sortenvergleich

Wir werden, wie gesagt,  $H_0$  verwerfen, wenn die Anzahl  $Z$  der Pluszeichen, auf die eine oder die andere Seite, zu stark von der Mitte (10) abweicht. Was wir unter "zu stark" verstehen, ist uns zunächst freigestellt. Um eine *signifikante* Entscheidung treffen zu können, müssen wir die maximale Abweichung so festlegen, dass sie in höchstens 5% der Fälle überschritten wird, wenn  $H_0$  gilt.



Probehalber legen wir fest: Weicht  $Z$  um mehr als 4 vom Wert 10 ab, so betrachten wir das als signifikante Abweichung. Dies führt auf den Verwerfungsbereich:

$$V = \{0, 1, \dots, 5, 15, \dots, 20\}$$

In Fig. 8.1 ist  $V$  grafisch dargestellt:  $V$  besteht aus zwei getrennten Intervallen.

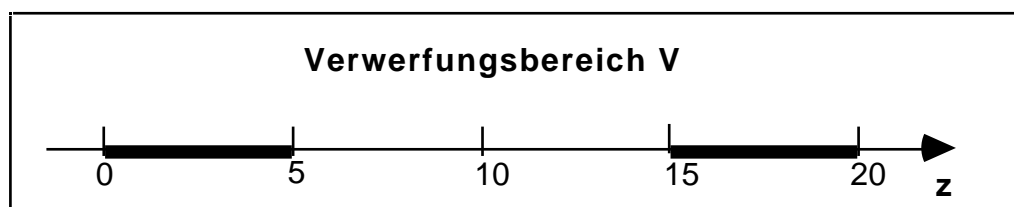


Fig. 8.1: Der Verwerfungsbereich  $V$  besteht aus zwei getrennten Intervallen.

Unausgesprochen sind wir bei dieser Festlegung davon ausgegangen, dass auf keiner Parzelle die Erträge genau gleich sind, d. h. dass wir keine Parzelle weglassen müssen, weil die Differenz 0 nichts zur Entscheidung beiträgt.

### Einseitig - zweiseitig

In den früheren Testbeispielen war der Verwerfungsbereich  $V$  meistens ein einziges Intervall in den ganzen Zahlen. Hier ist, wie gesagt,  $V$  die Vereinigung von zwei getrennten Bereichen und man spricht deshalb von einem *zweiseitigen* Test. Im Unterschied dazu heißen die früheren Test *einseitig*.



### Aufgabe 8.4

Legen Sie einen Verwerfungsbereich für den Test "Geburtsgewicht und IQ" fest. Begründen Sie Ihre Wahl mit den Untersuchungsergebnissen.

Beachten Sie bitte die Ergänzungen zu der gedruckten Lösung.

## 8.5 Berechnung von $\alpha$

Wir machen einen Fehler 1. Art, wenn wir die Nullhypothese  $H_0$  verwerfen, obwohl sie zutrifft, d. h. wenn wir zum Schluss kommen, dass die beiden Sorten nicht den gleichen Ertrag abwerfen ( $H_1$ ), obwohl das Gegenteil ( $H_0$ ) wahr ist. Wie gross ist die Wahrscheinlichkeit für diesen Irrtum?

Wenn  $H_0$  zutrifft, wird auf jeder der 20 Parzellen mit der Wahrscheinlichkeit  $p=0.5$  ein "+"-Zeichen erzeugt. Unsere Zufallsgrösse  $Z$  ist also binomisch verteilt mit  $n=20$  und  $p=0.5$ . Mit der im letzten Kapitel hergeleiteten Formel können wir somit wie folgt berechnen:



$$\begin{aligned}
 &= P(Z \in V, \text{ wenn } H_0 \text{ gilt}) \\
 &= P(0 \leq Z \leq 5 \text{ oder } 15 \leq Z \leq 20, \text{ wenn } p=0.5) \\
 &= 1 - P(6 \leq Z \leq 14, \text{ wenn } p=0.5) \\
 &= 1 - \sum_{k=6}^{14} P(Z=k) \\
 &= 1 - \sum_{k=6}^{14} \binom{20}{k} p^k q^{20-k} \\
 &= 1 - 0.5^{20} \sum_{k=6}^{14} \binom{20}{k}
 \end{aligned}$$

Der Taschenrechner liefert:

$$= 0.041 = 4.1\%$$

Mit dem gewählten Verwerfungsbereich werden wir also im nächsten Oktober entscheiden können, ob sich die beiden Sorten signifikant unterscheiden.



### Aufgabe 8.5

a

Beschreiben Sie umgangssprachlich, wie beim Test "Geburtsgewicht und IQ" ein Fehler 1. Art entsteht.

b

Berechnen Sie  $\alpha$  für den Verwerfungsbereich  $V=\{0, 1, 2, 3, 9, 10, 11, 12\}$ .

c

Hängt der IQ signifikant vom Geburtsgewicht ab?

d

Bei welchem zweiseitigen Verwerfungsbereich sind signifikante Entscheidungen möglich?

### Zusatzinformation zum Test "Geburtsgewicht und IQ"

Mit dem Vorzeichentest kann also aufgrund der Untersuchungsergebnisse *nicht* geschlossen werden, dass der IQ signifikant vom Geburtsgewicht abhängt. Der Vorzeichentest unterscheidet leider nicht, wie *stark* die IQ der beiden Zwillinge jeweils auseinander liegen.

Es gibt einen verbesserten Vorzeichentest (Rangsummentest von Wilcoxon), der auch die *Beträge* der Differenzen auswertet. Bei diesem Test ergibt sich, dass der IQ signifikant vom Geburtsgewicht abhängt.

Das verwirrt zunächst: Hängt nun der IQ vom Geburtsgewicht ab oder nicht? Beachten Sie aber: Eine Nullhypothese, die beibehalten wurde, ist noch nicht bewiesen, und verschiedene Tests führen notwendigerweise in bestimmten Situationen zu anderen Entscheidungen. Der verbesserte Vorzeichentest hat in den meisten Situationen die grössere Macht als der hier besprochene. Man ist daher versucht zu sagen, dass die Nullhypothese nicht zutrifft und der gewöhnliche Vorzeichentest einen Fehler 2. Art begeht. Wenn man diesen Gedanken aber weiterführt, wird man so lange verschiedene Tests durchführen, bis man einmal die gewünschte Entscheidung erhält. Das ist jedoch nicht statthaft. Das Niveau wird nur eingehalten, wenn man genau einen Test durchführt, und zwar muss man sich im Voraus für diesen einen Test entscheiden, ohne die Daten anzuschauen.

## 8.6 Berechnung von $\beta$

Der Test soll ja entscheiden, ob die Erträge der beiden Sorten A und B verschieden sind. Es ist nun bekanntlich denkbar, dass die Erträge tatsächlich ungleich sind, dass es aber der Test nicht merkt, weil Z trotzdem nicht in den Verwerfungsbereich V fällt (Fehler 2. Art mit der Wahrscheinlichkeit  $\beta$ ).

Berechnen wir  $\beta$  für den Fall, dass die Sorte A nur auf 30% der Parzellen höhere Erträge liefert als Sorte B. Unter dieser Voraussetzung ergibt sich mit der Wahrscheinlichkeit  $p=0.3$  auf jeder der 20 Parzellen ein "+"-Zeichen, die Zufallsgrösse Z ist mit anderen Worten binomialverteilt mit  $n=20$  und  $p=0.3$ . Mit diesen Überlegungen erhalten wir für  $\beta$  den Wert:

$$\begin{aligned}
 &= P( "Z \leq V, \text{ wenn } H_1 \text{ gilt" } ) \\
 &= P( "6 \leq Z \leq 14, \text{ wenn } p=0.3" ) \\
 &= \sum_{k=6}^{14} P(Z=k) \\
 &= \sum_{k=6}^{14} \binom{20}{k} p^k q^{20-k} \\
 &= 58.4\%
 \end{aligned}$$

Die *Macht* des Tests, die beschriebene Ungleichheit aufzudecken, beträgt also nur 41.6%. Man kann diese enttäuschend kleine Macht vergrössern, indem man entweder V oder – noch besser – die Parzellenzahl n vergrössert.



### Aufgabe 8.6

a

Beschreiben Sie umgangssprachlich, worin beim Test "Geburtsgewicht und IQ" der Fehler 2. Art besteht.

b

Berechnen Sie für den Verwerfungsbereich  $V = \{0, 1, 2, 3, 9, 10, 11, 12\}$  die Wahrscheinlichkeit  $\beta$  für den Fehler 2. Art, wenn Sie annehmen, dass zu 70% der bei der Geburt schwerere Zwilling den höheren IQ hat.

Wenn Sie die bisherigen Aufgaben einigermaßen richtig gelöst haben, sind Sie reif für die Lernkontrolle. Wenn Sie diese mit Erfolg bewältigt haben, können Sie sich zum letzten Kapitteltest dieses Leitprogramms anmelden.



### Lernkontrolle: Zwei Messmethoden vergleichen

Es stehen zwei Methoden zur Wahl, den Stärkegehalt von Kartoffeln zu bestimmen. Die beiden Methoden wurden verglichen, indem man 16 Kartoffeln halbierte und jeweils beide Methoden anwendete. Es ergaben sich die folgenden Unterschiede des Stärkegehaltes (in %) zwischen den beiden Hälften:

2	0	0	1	2	2	3	-3	1	2	3	0	-1	1	-2	1
---	---	---	---	---	---	---	----	---	---	---	---	----	---	----	---

Es sind zwei verschiedene Fragen denkbar, die je mit Hilfe eines entsprechenden Vorzeichentests beantwortet werden können.

a

Man hat das Gefühl, die beiden Methoden seien nicht gleichwertig. Prüfen Sie mit einem Vorzeichentest, ob es einen signifikanten Unterschied zwischen den beiden Methoden gibt.

b

Man hat das Gefühl, die erste Methode liefere in der Regel höhere Stärkegehalte als die zweite. Prüfen Sie mit einem (einseitigen!) Vorzeichentest, ob Methode 1 signifikant höhere Werte liefert als Methode 2.

c

Wenn Sie richtig gerechnet haben, sind Sie auf (scheinbar) widersprüchliche Ergebnisse gekommen. Versuchen Sie diesen Widerspruch zu klären.



## Lösungen der Aufgaben zu Kapitel 8

### Lösung Aufgabe 8.0

In beiden Beispielen geht es um die Beobachtung *zweier* zufälliger Grössen, wir nennen sie X und Y, die in natürlicher Weise *gepaart* sind:

Vergleich der Sorten A und B	Geburtsgewicht und IQ
X= Ertrag der Sorte A	X= IQ des schweren Zwillings
Y= Ertrag der Sorte B	Y= IQ des leichten Zwillings

### Lösung Aufgabe 8.1

Auch im Test "Geburtsgewicht und IQ" sollen allfällige *Unterschiede* zwischen dem IQ des schweren und des leichten Zwillings aufgedeckt werden. Zu diesem Zweck versucht man mit Hilfe der Versuchsergebnisse die folgende Nullhypothese zu widerlegen:

$H_0$ : "Der IQ ist unabhängig vom Geburtsgewicht"

oder:

$H_0$ : "Der schwere und der leichte Zwilling haben den gleichen IQ"

### Lösung Aufgabe 8.2

$H_1$ : "Die IQ's des schweren und des leichten Zwilling sind verschieden"

### Lösung Aufgabe 8.3

Wir wiederholen die Versuchsergebnisse des Tests, wobei wir in einer zusätzlichen Zeile die Vorzeichen der Differenzen X und Y anfügen:

Nr. des Paares	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IQ des schweren Zwilling: X	97	79	100	10	10	12	95	80	91	10	91	90	10	11
			0		0	4				8			4	9
IQ des leichten Zwilling: Y	97	70	10	106	85	12	84	70	84	10	97	90	92	10
			1			3				6				4
Vorzeichen von X-Y	/	+	-	-	+	+	+	+	+	+	-	/	+	+

a Wir definieren:

 $Z = \text{"Anzahl der Pluszeichen in der obigen Tabelle"}$ 

b

Bei 2 der 14 Zwillingspaare stimmen die IQ's überein. Deshalb lassen wir diese beiden Paare weg. Von den restlichen 12 Differenzen sind 9 positiv, d. h. für die obigen Daten gilt:

 $Z = 9$ 

### Lösung Aufgabe 8.4

Wenn  $H_0$  zutrifft, sollten etwa 6 der 12 betrachteten Differenzen positiv sein. Wenn wir im vorliegenden Fall  $H_0$  verwerfen möchten, müssten wir wegen  $Z=9$  Abweichungen von wenigstens 3 als signifikant betrachten können. Pro behalber legen wir deshalb fest:

 $V = \{0, 1, 2, 3, 9, 10, 11, 12\}$ 

Die nachfolgende Berechnung wird uns zeigen, ob wir damit signifikante Entscheidungen ( 5%) treffen können.

*Ergänzung:* Wir haben hier also den Verwerfungsbereich so gewählt, dass  $H_0$  gerade noch verworfen wird. Wir wollen dann das zugehörige berechnen; man bezeichnet diesen Wert auch als *p-Wert*.

### Lösung Aufgabe 8.5

a Beim Test "Geburtsgewicht und IQ" macht man den Fehler 1. Art, wenn man aufgrund der Daten irrtümlicherweise schliesst, dass der IQ vom Geburtsgewicht abhängt, d. h. dass die beiden Zwillinge verschiedene IQ haben.

b Wenn  $H_0$  gilt, so ist  $Z$  binomisch verteilt mit  $n=12$  und  $p=0.5$ . Für den Verwerfungsbereich  $V = \{0, 1, 2, 3, 9, 10, 11, 12\}$  ergibt sich demzufolge:

$$\begin{aligned}
 &= P(Z \in V, \text{ wenn } H_0 \text{ gilt}) \\
 &= P(0 \leq Z \leq 3 \text{ oder } 9 \leq Z \leq 12, \text{ wenn } p=0.5) \\
 &= 1 - P(4 \leq Z \leq 8, \text{ wenn } p=0.5) \\
 &= 1 - \sum_{k=4}^8 P(Z=k) \\
 &= 1 - \sum_{k=4}^8 \binom{12}{k} p^k q^{12-k} \\
 &= 1 - 0.5^{12} \sum_{k=4}^8 \binom{12}{k}
 \end{aligned}$$

Der Taschenrechner liefert:

$$= 0.146 = 14.6\%$$

c

Wenn  $H_0$  zutrifft, müsste man also durchschnittlich in jeder 7-ten solchen Untersuchung mit dem vorliegenden Untersuchungsergebnis rechnen. Man kann mit anderen Worten nicht schliessen, dass der IQ signifikant vom Geburtsgewicht abhängt.

d

Es ist naheliegend, zu hoffen, dass signifikante Entscheidungen mit dem verkleinerten Verwerfungsbereich  $V = \{0, 1, 2, 10, 11, 12\}$  möglich sind. Die Rechnung liefert in der Tat:

$$\begin{aligned} &= P(Z \in V, \text{ wenn } H_0 \text{ gilt}) \\ &= P(0 \leq Z \leq 2 \text{ oder } 10 \leq Z \leq 12, \text{ wenn } p=0.5) \\ &= 1 - P(3 \leq Z \leq 9, \text{ wenn } p=0.5) \\ &= 1 - \sum_{k=3}^9 P(Z=k) \\ &= 1 - \sum_{k=3}^9 \binom{12}{k} p^k q^{12-k} \\ &= 1 - 0.5^{12} \sum_{k=3}^9 \binom{12}{k} \\ &= 3.86\% \end{aligned}$$

### Aufgabe 8.6

a

Einen Fehler 2. Art macht man, wenn man nicht merkt, dass der IQ vom Geburtsgewicht abhängt, d. h. dass schwerer und leichter Zwilling einen unterschiedlichen IQ haben.

b

Wenn bei 70% der Zwillingspaare der schwerere Zwilling den höheren IQ hätte, so würde mit der Wahrscheinlichkeit  $p=0.7$  bei jedem der 12 Zwillingspaare ein Pluszeichen auftreten, d. h.  $Z$  wäre eine binomisch verteilte Zufallsgrösse mit  $p=0.7$  und  $n=12$ . Folglich ergäbe sich für den nebenstehend berechnete Wert:

$$= 50.6\%$$

$$\begin{aligned} &= P(Z \in V, \text{ wenn } H_1 \text{ gilt}) \\ &= P(4 \leq Z \leq 8, \text{ wenn } p=0.7) \\ &= \sum_{k=4}^8 P(Z=k) \\ &= \sum_{k=4}^8 \binom{12}{k} p^k q^{12-k} \\ &= 50.6\% \end{aligned}$$

Der Test hätte also die sehr geringe Macht von  $1 - \beta = 49.4\%$ , den 70%-Anteil zu erkennen (durchschnittlich nur in jedem 2. Fall).

### Lösung der Lernkontrolle

Beim Vorzeichentest betrachten wir nur die 13 Kartoffeln, bei welchen die beiden Methoden einen unterschiedlichen Stärkegehalt angezeigt haben. In der Tabelle gibt es 10 Pluszeichen und 3 Minuszeichen.

a

Es soll getestet werden, ob es einen signifikanten *Unterschied* zwischen den beiden Methoden gibt.

Nullhypothese:

$H_0$ : "Die beiden Methoden liefern den gleichen Stärkegehalt"

Alternativhypothese:

$H_1$ : "Die beiden Methoden liefern *nicht* den gleichen Stärkegehalt"

Testgrösse:

$Z = \text{"Anzahl der Pluszeichen in der Tabelle"}$

Verwerfungsbereich:

Wenn  $H_0$  zutrifft, sollte die Tabelle nicht wesentlich mehr oder weniger als 6 oder 7 "+"-Zeichen enthalten. Wenn wir aufgrund der vorliegenden Daten (10 "+"-Zeichen)  $H_0$  verwerfen wollen, müssen wir den folgenden Verwerfungsbereich in Betracht ziehen:

$V = \{0, 1, 2, 3, 10, 11, 12, 13\}$

Bei diesem Verwerfungsbereich ergibt sich das folgende Signifikanzniveau:

$$\begin{aligned}
 &= P(Z \in V, \text{ wenn } H_0 \text{ gilt}) \\
 &= P(0 \leq Z \leq 3 \text{ oder } 10 \leq Z \leq 13, \text{ wenn } p=0.5) \\
 &= 1 - P(4 \leq Z \leq 9, \text{ wenn } p=0.5) \\
 &= 1 - \sum_{k=4}^9 P(Z=k) \\
 &= 1 - \sum_{k=4}^9 \binom{13}{k} p^k q^{13-k} \\
 &= 1 - 0.5^{13} \sum_{k=4}^9 \binom{13}{k} \\
 &= 9.23\%
 \end{aligned}$$

Wir sehen, dass die vorliegenden Daten *nicht* signifikant gegen die Nullhypothese sprechen: Wir werden die beiden Methoden (vorläufig) als gleichwertig ansehen.

b

Es soll getestet werden, ob Methode 1 signifikant *höhere* Werte liefert als Methode 2. (Beachten Sie den entscheidenden Unterschied zu vorherigen Aufgabe a, wo wir untersucht haben, ob die Werte *unterschiedlich* sind.)

Nullhypothese:

$H_0$ : "Die beiden Methoden liefern den *gleichen* Stärkegehalt"

Alternativhypothese:

$H_1$ : "Methode 1 liefert *höhere* Werte als Methode 2"

Testgrösse:

$Z$  = "Anzahl der Pluszeichen in der Tabelle"

Verwerfungsbereich:

Wenn  $H_0$  zutrifft, sollte die Tabelle *nicht wesentlich mehr* als 7 "+"-Zeichen enthalten. Wenn wir aufgrund der vorliegenden Daten  $H_0$  verwerfen wollen, müssen wir den folgenden *einseitigen* Verwerfungsbereich in Betracht ziehen:

$$V = \{10, 11, 12, 13\}$$

Bei diesem Verwerfungsbereich können wir auf dem folgenden Signifikanzniveau entscheiden:

$$\begin{aligned} &= P(Z \in V, \text{ wenn } H_0 \text{ gilt}) \\ &= P(10 \leq Z \leq 13, \text{ wenn } p=0.5) \\ &= \sum_{k=10}^{13} P(Z=k) \\ &= \sum_{k=10}^{13} \binom{13}{k} p^k q^{13-k} \\ &= 4.61\% \end{aligned}$$

Wir sehen: Methode 1 liefert signifikant höhere Werte als Methode 2.

**c**

Die StatistikerInnen fühlen sich sicher, wenn sie die Nullhypothese verwerfen können, denn dieser Entscheid beruht auf einem Ereignis, das nur eine geringe Wahrscheinlichkeit hätte, wenn  $H_0$  zutreffen würde. Sie neigen dann eben (siehe Aufgabe b) dazu, die Alternativhypothese zu akzeptieren, die das Ereignis viel einfacher erklärbar macht.

In unserem Fall steht diese Hypothese "Methode 1 liefert höhere Werte als Methode 2" in offenem Widerspruch zu "Die beiden Methoden liefern den gleichen Stärkegehalt", die wir aufgrund der gleichen Daten in Aufgabe a akzeptiert haben. Wir haben sie dort akzeptiert oder beibehalten, weil eben das "Instrument" des angewendeten Tests nicht ausgereicht hat, sie zu verwerfen. Ob dies zu Recht geschehen ist oder ob ein "besseres Instrument" zu anderen Schlüssen gekommen wäre, müssen wir dahin gestellt lassen. *Auf jeden Fall gilt eine bereits früher gemachte wichtige Bemerkung: Sie müssen sich im voraus entscheiden, welchen Test Sie anwenden wollen und Sie müssen das evtl. nicht willkommene Ergebnis akzeptieren.*

***Anhang S***

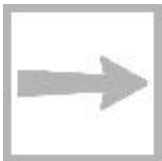
# ***Simulationen***



## **Computersimulationen zu Kapitel 2**

Im Kap. 2 haben Sie gesehen, wie man *grundsätzlich* mit Hilfe eines Ikosaeders Stichproben auslosen, oder wie wir sagen, *simulieren* kann. Mit Recht haben Sie den Eindruck, dass dieses Verfahren nicht praktikabel ist. In diesem Kapitel erfahren Sie, wie die Aufgabe elegant und schnell vom Computer erledigt wird.

Natürlich gibt es professionelle Anwenderprogramme, die Ihnen diese Aufgabe abnehmen. Hier geht es darum, dass Sie eine oder beide von zwei präsentierten Simulationstechniken studieren und selber programmieren. Besprochen werden:

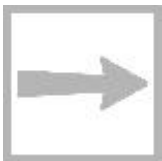


- Die Simulation als Tabellenkalkulation (auf EXCEL)
- Die Simulation auf einem Taschencomputer.

Ihre Lehrperson hat Sie instruiert, welche der beiden Techniken Sie bearbeiten sollen.

### **I.1**

#### **Simulation als Tabellenkalkulation**



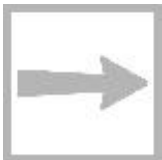
Wir setzen voraus, dass rudimentäre Erfahrungen auf EXCEL vorliegen. Trotzdem sind die Erklärungen ziemlich ausführlich. Wenn Sie diesen Abschnitt durchgearbeitet haben, sind Sie in der Lage, innerhalb weniger Minuten auf EXCEL eine Simulationen zu programmieren, wie Sie sie aus Kap. 2 kennen.

Mit wenig zusätzlichem Zeit- und Arbeitsaufwand werden Sie fähig sein, das Ergebnis der Simulation in Form einer *Häufigkeitstabelle* und einem *Histogramm* auszuwerten.

Für die nachfolgenden Ausführungen spielt es praktisch keine Rolle, mit welcher Version von EXCEL Sie arbeiten müssen. Die präsentierten Ergebnisse wurden durchweg mit EXCEL 4.0 auf Mac erzeugt.

### **I.1.1**

#### **Simulation von 10-er Stichproben**



Wir erlernen in diesem Abschnitt das Prinzip der Simulation auf EXCEL anhand von übersichtlichen 10-er Stichproben. Erst im nächsten Abschnitt proben wir den Ernstfall.

### Auftrag 1: EXCEL öffnen

Öffnen Sie auf Ihrem Computer das Programm EXCEL. Auf dem Bildschirm erscheint die nachstehend teilweise abgebildete „Tabelle 1“. Die Zelle A1 ist umrahmt, weil sie *aktiviert* ist und damit eine Eingabe irgendwelcher Art in dieser Zelle gespeichert würde.

	Tabelle1			
	A	B	C	D
1				
2				
3				
4				
5				
6				

### Auftrag 2: Eine Zufallszahl erzeugen

Geben Sie in die Zelle A1 den Befehl `=ZUFALLSZAHLC()` ein, entweder durch Eintippen oder über das Menü **FORMEL / FUNKTION EINFÜGEN** (mit der **ENTER**-Taste abschliessen). Dieser Befehl bewirkt, dass – bildlich gesprochen – der Computer 6-mal den Ikosaeder wirft und das Ergebnis als einen 6-stelligen Dezimalbruch präsentiert, z. B. so:

	Tabelle1			
	A	B	C	D
1	0.012988			
2				
3				
4				
5				

Wenn wir uns nur für 5-stellige Zufallszahlen interessieren, schneiden wir – in Gedanken – einfach die hinterste Dezimale ab und nehmen nur die ersten fünf; in diesem Fall hat EXCEL die Nummer `1298` ausgewählt.

### Zusatzinformation 1

Man nennt die vom Computer erzeugten Zufallszahlen *Pseudozufallszahlen*, weil sie in Wirklichkeit nicht durch den Zufall bestimmt sind, sondern durch ein Programm, einen sog. Zufallsgenerator. Gute Zufallsgeneratoren erzeugen Pseudozufallszahlen, die man nicht von richtigen Zufallszahlen unterscheiden kann.

### Zusatzinformation 2

Sie haben sicher festgestellt, dass Sie das Eintippen des Befehls



`ZUFALLSZAHL()` in der Eingabezeile unmittelbar oberhalb der Tabelle kontrollieren können. In der Zelle A1 steht also nicht dieser Befehl, sondern dessen *Ergebnis*. Der Befehl selber ist aber nicht etwa verloren gegangen, sondern wird sozusagen in einer Schicht unter der Tabelle konserviert. Sie können sich davon überzeugen, indem Sie die Zelle mit dem Cursor aktivieren und die Eingabezeile oberhalb der Tabelle konsultieren.

### Auftrag 3: Schwarze oder weisse Kugel

Der oben gewürfelten Nummer 1298 entspricht eine schwarze Kugel. Alle zufälligen Dezimalbrüche, die kleiner als 0.04 sind, ergeben Nummern der Form

0, 1, 2, ..., 3998, 3999,

also solche von schwarzen Kugeln, welche wir in unserem früheren 0-1-Protokoll einfach als 1 notiert haben. Diese Notation können wir auf EXCEL wie folgt erzeugen. Geben Sie in die Zelle A1 den Befehl:

`=WENN( ZUFALLSZAHL( ) < 0.04; 1 ; 0)`

ein. Was passiert? Es wird, wie oben erwähnt, ein zufälliger 6-stelliger Dezimalbruch erzeugt ( „ZUFALLSZAHL( )“ ) und geprüft, ob dieser kleiner als 0.04 ist ( „WENN( .....< 0.04; 1; 0“ ). Im Ja-Fall schreibt EXCEL eine „1“ in die Zelle A1, im Nein-Fall eine „0“. Das sieht dann möglicherweise so aus:

Tabelle1					
	A	B	C	D	
1	1				
2					
3					
4					
5					

### Auftrag 4: Spaltenbreite verkleinern

Damit wir die Simulation einer 10-er Stichprobe auf den Bildschirm kriegen, müssen wir die Spaltenbreite verkleinern. Aktivieren Sie dazu die Zeile 1 (Tabellenrand links) und befehlen Sie über das Menü

`FORMAT /SPALTENBREITE...`

die Spaltenbreite 2. Ihre Tabelle sieht nun so aus:

Tabelle1															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	1														
2															
3															
4															
5															

### Auftrag 5: Simulation einer 10-er Stichprobe

Nun sind wir soweit: Wir können eine 10-er Stichprobe simulieren, indem wir den in der „Unterschicht“ von Zelle A1 liegenden Befehl

`=WENN( ZUFALLSZAHL( ) < 0.04; 1 ; 0)`

auch in die rechts benachbarten Zellen B1 bis J1 in der obersten Zeile der Tabelle schreiben. In EXCEL ist das mit einem einfachen Kopierbefehl möglich:

- Aktivieren Sie mit dem Cursor diese Zellen A1 bis J1;
- Wählen Sie im Menü

BEARB. / RECHTS AUSFÜLLEN

Damit wird nicht nur der Befehl in die „Unterschicht“ jeder dieser Zellen kopiert, sondern diese Befehle werden zugleich ausgewertet. Deshalb sieht Ihre Tabelle nun etwa so aus:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	0	0	0	0	0	0	0	0	0	0							
2																	
3																	
4																	
5																	

In obiger Stichprobe befinden sich drei schwarze Kugeln.

### Auftrag 6: Anzahl schwarzer Kugeln

Wir können die Zeile einfach so ergänzen, dass die Anzahl der schwarzen Kugeln automatisch berechnet wird. Dazu schreiben wir in die Zelle L1 den Befehl:

`=SUMME(A1:J1)`

Die Tabelle mag nun wie folgt aussehen:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	0	0	1	0	0	0	1	0	0	0		2					
2																	
3																	
4																	
5																	

Sie sehen, dass neu nicht nur die Summe der schwarzen, nämlich 2, berechnet wurde, sondern dass offensichtlich vorher alle Nummern neu ausgelost wurden. Bei jeder Änderung in der Tabelle wird diese vollständig neu berechnet, wenn Sie nicht ausdrücklich eine gegenteilige Option wählen.

**Auftrag 7: 25 Simulationen auf einmal**

Und so unglaublich einfach ist es, 25 Stichproben zu simulieren:

- Zellen A1 bis L25 (also diagonal) mit dem Cursor aktivieren und ...
- .... mit dem Menü **BEARB. / UNTEN AUSFÜLLEN** nach unten kopieren.

Hier das Ergebnis:

Tabelle																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	0	0	0	0	0	0	0	0	0	0	0				
2	0	0	0	0	0	0	0	0	0	0	0	0				
3	0	0	0	0	0	0	0	0	0	0	0	0				
4	0	1	0	0	1	0	0	0	0	0	0	2				
5	0	0	0	0	0	0	0	0	0	0	0	0				
6	0	0	0	0	0	0	0	0	0	1	1	1				
7	0	0	0	0	0	0	0	0	0	0	0	0				
8	0	0	0	0	0	0	0	0	0	0	0	0				
9	0	0	0	0	0	0	0	0	0	0	0	0				
10	0	0	0	0	1	0	0	0	0	0	1	1				
11	0	0	0	0	0	0	0	0	0	1	1	0				
12	0	0	0	0	0	0	0	0	0	0	0	0				
13	0	0	0	1	0	0	0	0	0	0	0	1				
14	0	0	0	0	0	0	0	0	0	0	0	0				
15	0	0	0	0	0	0	0	0	0	0	0	0				
16	0	0	0	0	0	0	0	0	0	0	0	0				
17	0	0	0	0	0	1	0	0	0	0	0	1				
18	0	0	0	0	0	0	0	0	0	0	0	0				
19	0	0	0	0	0	0	0	0	0	0	0	0				
20	0	0	0	0	0	0	0	1	0	0	1	1				
21	0	0	0	0	0	0	0	0	0	0	0	0				
22	0	0	0	0	0	0	1	0	0	1	2	0				
23	0	0	0	0	0	0	0	0	0	0	0	0				
24	0	1	0	0	0	1	0	0	0	0	2	0				
25	0	1	0	0	0	0	0	0	0	0	1	0				
26																
27																

Soweit die Simulation der 25 Stichproben. Sie könnten nun die Tabelle von Hand auswerten. In der gleichen Zeit schaffen Sie aber auch die programmierte Auswertung.

**Ein Schönheitsfehler**

In unserer Lösung haben wir "vergessen" zu kontrollieren, ob eine eben ausgeloste Nummer nicht früher schon einmal vorgekommen ist. Wir haben in Kap. 2 erwähnt, dass dieses Ereignis sehr selten eintritt und dass deshalb der Fehler vernachlässigbar ist.

**Auftrag 8: Die Häufigkeitstabelle**

Auf einem neuen Blatt wollen wir das Ergebnis dieser 25 Simulationen in Form einer Häufigkeitstabelle zusammenstellen. Hier das Ergebnis vorweg:

Zeile 1: Anzahl der schwarzen Kugeln

**Tabelle3**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	1	2	3	4	5	6	7	8	9	10		
2	15	7	3	1	0	0	0	0	0	0	0		
3													
4													
5													

Zeile 2: Absolute Häufigkeiten (von den 25 Stichproben enthielten z. B. 15 keine schwarze Kugel)

In der Zeile 1 stehen die möglichen Anzahlen 0, 1, ..., 10 schwarzer Kugeln, in der Zeile 2 die absolute Häufigkeit, mit welcher diese Werte in der Simulation der 25 Stichproben vorkommen.

Wenn Sie die fünf Punkte der nachfolgenden Anweisung gewissenhaft befolgen, sollten Sie schon nach einer Viertelstunde fähig sein, diese Tabelle aus dem Kopf zu erzeugen. In den meisten Fällen können die Teilziele über Menübefehle *erheblich* bequemer, als unten beschrieben, erreicht werden. Da diese Menübefehle jedoch von Version zu Version variieren, ist es nicht möglich, in den Anweisungen detailliert darauf einzugehen. Konsultieren Sie allenfalls Ihr Handbuch oder Ihren Nachbar.

1

Neue Tabelle: Eröffnen oder aktivieren Sie - ohne die Tabelle 1 zu schliessen - eine neue Tabelle (in der Fig. oben „Tabelle 3“).

2

Erzeugen Sie in der Zeile 1 die Zahlenfolge 0, 1, ..., 10, entweder durch Eintippen oder über Menübefehle.

3

Schreiben Sie in die Zelle A2 den Befehl:

`=HÄUFIGKEIT(Tabelle1!$L$1:$L$25;A1)`

Er bewirkt das folgende: In der Kolonne L1 bis L25 der Tabelle 1 wird gezählt, wie oft ein Wert vorkommt, *der kleiner oder gleich* dem Wert 0 in der Zelle A1 der Tabelle 3 ist.

4

Schreiben Sie in die Zelle B2 den Befehl:

`=HÄUFIGKEIT(Tabelle1!$L$1:$L$25;B1)-HÄUFIGKEIT(Tabelle1!$L$1:$L$25;A1)`

Er bewirkt das folgende: In der Kolonne L1 bis L25 der Tabelle 1 wird gezählt, wie oft ein Wert vorkommt, *der kleiner oder gleich* dem Wert 1 in der Zelle B1

der Tabelle 3 ist und davon wird die unter  berechnete Häufigkeit subtrahiert.

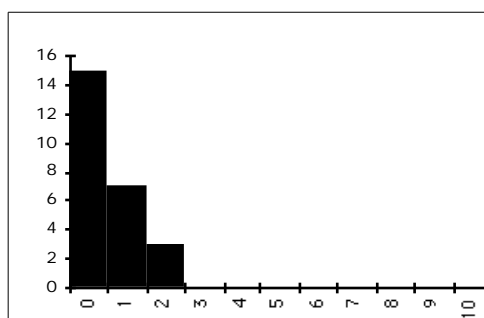
In die weiteren Zellen C2 bis J3 der zweiten Zeile müssen wir die analogen Befehle eingeben. Wir erreichen dies, indem wir die Zellen B2 bis J2 aktivieren und über das Menü  den analogen Befehl in B2 nach rechts kopieren.

### Auftrag 9: Das Histogramm der Häufigkeitstabelle

Zuletzt geht es noch um das Histogramm. Aktivieren Sie dazu mit dem Cursor die Häufigkeitstabelle und befehlen Sie je nach Version über das Menü  oder  das Erstellen eines Diagramms.

In einem neuen Fenster erscheinen Optionen: Wichtig ist, dass es Ihnen nun gelingt, mitzuteilen, dass von der Tabelle ein xy-Diagramm zu erstellen ist, und dass in der ersten Reihe die x-Werte stehen.

Sie erhalten ein Diagramm, das sich von demjenigen in der nebenstehenden Fig. mehr oder weniger unterscheidet; über Menübefehle können Sie das Muster des Diagramms fast nach Belieben ändern.



#### Aufgabe I.1.1

Nehmen Sie an, die Urne enthalte 8000 statt nur 4000 schwarze Kugeln. Es sollen wieder 25 Stichproben simuliert werden.

Prüfen Sie Ihr statistisches Gespür und schreiben Sie aus dem Kopf eine Häufigkeitstabelle auf.

Ändern Sie jetzt Ihr Simulationsprogramm in Tabelle 1 so ab, dass EXCEL für die neue Situation die Stichproben simuliert.

Drucken Sie die neue Häufigkeitstabelle und das zugehörige Histogramm, die ohne weiteres Zutun automatisch berechnet wurden.



#### Aufgabe I.1.2

Stellen Sie sich vor, dass sich unser Einkäufer mit dem Lieferanten auf den folgenden einfachen Testplan geeinigt hätte:

Es werden zufällig 10 Knallkörper ausgewählt und gezündet. Wenn diese Stichprobe mehr als einen defekten Knallkörper enthält, wird die Sendung nicht akzeptiert.

a

Worin besteht der Fehler 2. Art bei diesem Test?

b

In der Aufgabe I.1.1 haben wir 25 Stichproben simuliert für den Fall, dass die Urne 8000 schwarze Kugeln enthält. Wie oft wäre bei diesen 25 Simulationen der Fehler 2. Art gemacht worden? (Entnehmen Sie die gesuchte Anzahl sowohl aus der Häufigkeitstabelle, wie auch aus dem zugehörigen Histogramm.)

c

Wie gross ist etwa die Wahrscheinlichkeit  $\beta$  für den Fehler 2. Art für den Fall, dass die Sendung doppelt so viele defekte Knallkörper enthält, als erlaubt sind?

## I.1.2 Hundert Simulationen einer 200-er Stichprobe

Unser eigentliches Ziel ist es, die Wahrscheinlichkeiten  $\alpha$  und  $\beta$  für die Fehler 1. und 2. Art beim Qualitätstest mit einer 200-er Stichprobe zu schätzen. Dazu müssen wir einfach unsere Tabelle 1 nach rechts auf 200 Plätze ausdehnen. Unsere Schätzwerte werden besser, wenn wir statt nur 25 z. B. 100 Simulationen durchführen, also unsere Tabelle nach unten auf 100 Zeilen ausdehnen.

### Auftrag 1:

Programmieren Sie auf EXCEL die Simulationen von Hundert 200-er Stichproben. Orientieren Sie sich an der nachstehenden Vorlage, welche die vier Ecken der gesuchten Tabelle zeigt. Beachten Sie auch die anschliessenden Erklärungen.

	A	B	C	D	E	F	G	H
1	0.04							
2	1	2	3	4	5	6	7	
3	0	0	0	0	0	0	0	
4	0	0	0	0	1	0	1	
5	1	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	

Tabelle1										
GL	GM	GN	GO	GP	GQ	GR	GS	GT	G	
194	195	196	197	198	199	200				
0	0	0	0	0	0	0	0	4		
0	0	0	0	0	0	0	0	9		
0	0	0	0	0	0	0	0	8		
0	0	0	0	0	0	0	0	7		
0	0	0	0	0	0	0	0	10		
0	0	0	0	0	0	0	0	2		

- Zelle A1 enthält den 4000 schwarzen Kugeln entsprechenden Wert 0.04.

- =WENN(ZUFALLSZAHL()<\$A\$1;1;0)

- Die Zeile 2 enthält zur Kontrolle die Zahlenfolge 1, 2, ..., 200. Das Programm funktioniert auch ohne diese Zeile.

- =SUMME(A3:GR3) .

Wenn Sie die Spalte GT prüfen, sehen Sie, dass die Anzahl schwarzer Kugeln in den einzelnen Stichproben im Bereich zwischen 4 und 14 schwanken. Damit Sie auf der sicheren Seite stehen, bereiten Sie eine Häufigkeitstabelle vor mit den Werten 0, 1, ..., 19, 20 für die Anzahl der schwarzen Kugeln.

**Erläuterungen:**

- Zeile 1 enthält die Zahlenfolge 0, 1, ..., 20 für die Anzahl schwarzer Kugeln.
- In der Zelle A2 steht der Befehl:

```
=HÄUFIGKEIT(Tabelle1!$GT$3:$GT$102;A1)
```

- In der Zelle B2 steht der Befehl:

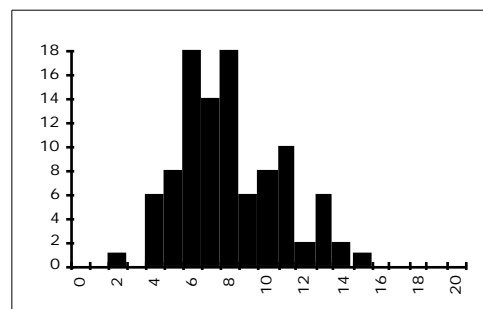
```
=HÄUFIGKEIT(Tabelle1!$GT$3:$GT$102;B1)-HÄUFIGKEIT(Tabelle1!$GT$3:$GT$102;A1)
```

- In den weiteren Zellen dieser Zeile stehen entsprechende (nach rechts kopierte) Befehle.

**Auftrag 3: Das Histogramm**

Erstellen Sie das Histogramm der Häufigkeitstabelle.

Für unsere Simulation ergibt sich das Diagramm rechts.

**Aufgabe I.2.1**

Sie müssen in Ihrem Programm eine einzige Zelle ändern, um die folgende Aufgabe zu lösen.

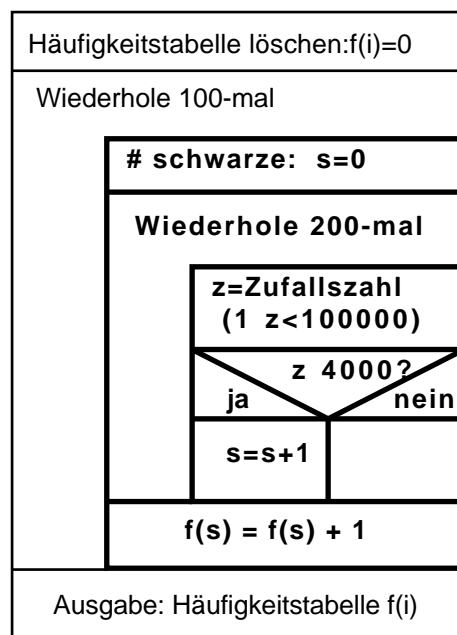
Nehmen Sie an, die Knallkörper-Sendung enthalte statt nur 4000 genau 8000 defekte Exemplare und der Einkäufer führe den besprochenen Qualitätstest durch. Welchen Schätzwert  $\beta$  erhalten Sie aufgrund von 100 Simulationen für die Wahrscheinlichkeit des Fehlers 2. Art?

**I.2****Simulation auf dem Taschencomputer**

Unten finden Sie für das Simulationsprogramm das Struktogramm und den Programmcode für den Taschenrechner vom Typ TI-92. Die fettgedruckten Teile beziehen sich auf das Ziehen *einer* Stichprobe mit 200 Kugeln und ihrer Auswertung in der Häufigkeitstabelle  $f(0)$ ,  $f(1)$ , ...,  $f(20)$ . (Wir hoffen mit anderen Worten, dass nicht mehr als 20 schwarze Kugeln gezogen werden.)

Bei den dünngedruckten Partien geht es um das *hundertmalige* Ziehen einer solchen Stichprobe.





```

simulier()
Prgm
newMat(2,21) f
For i, 1, 21, 1
    i-1 f[1, i]
EndFor
ClrIO
For j, 1, 100, 1
    0→s
    For i, 1, 200, 1
        If rand(100000) > 4000 Then
            s+1→s
        EndIf
    EndFor
    f[2,s+1]+1→f[2,s+1]
EndFor
Disp f
EndPrgm

```

Der Rechner präsentiert nach etwa 20 Minuten Rechenzeit die rechtsstehende Häufigkeitstabelle.

0	1	2	3	4	5	6	7	8	9
0	1	0	4	6	7	17	10	13	13

Nachstehend die Tabelle, soweit die Häufigkeiten von 0 verschieden sind.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1	0	4	6	7	17	10	13	13	14	6	5	2	1	1	0	0

### Interpretation

2 von 100 simulierten Stichproben enthalten mehr als 13 schwarze Kugeln, d. h. hier ergibt sich für den Schätzwert:

2%

### Simulation von $\beta$

Wenn wir im obigen Programm die Zahl 4000 durch 8000 ersetzen und die Häufigkeitstabelle auf 31 Plätze vergrössern (weil mit deutlich grösseren Anzahlen an schwarzen Kugeln gerechnet werden muss), erhalten wir wiederum nach ungefähr einer Viertelstunde die folgenden Häufigkeitstabelle:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	0	0	0	0	0	0	0	1	1	2	4	7	5	13	11	10	10

20 von 100 simulierten Stichproben enthalten 13 oder weniger schwarze Kugeln, d. h. hier ergibt sich für den Schätzwert:

20%

## Lösungen der Aufgaben zu Kapitel I

### Lösung Aufgabe I.1.1

a) Vergleichen Sie Ihre Lösung mit c).

b)

Das Programm kann in drei kleinen Schritten geändert werden:

- Der Befehl in der Zelle A1 in der Tabelle 1 lautet neu:

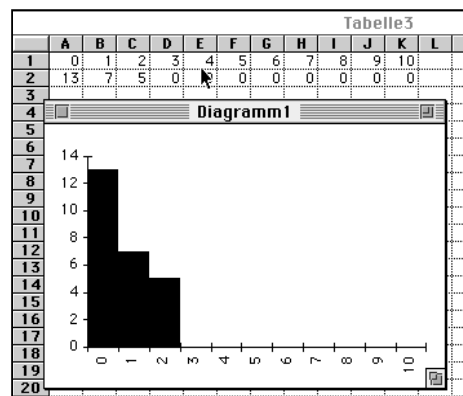
`=WENN(ZUFALLSZAHL()<0.08 ;1;0)`

- Befehl in A1 über Menü **BEARB. / RECHTS AUSFÜLLEN** nach rechts kopieren (bis J1).

- Befehle in A1 bis J1 über Menü **BEARB./ UNTEN AUSFÜLLEN** nach unten kopieren bis Zeile 25.

c)

Kaum haben Sie Ihr Simulationsprogramm geändert, können Sie auch schon die neue Häufigkeitstabelle und das Histogramm „abholen“. Der Wert 2 sollte deutlich öfter als bisher vorkommen.



### Lösung Aufgabe I.1.2

a)

Den Fehler 2. Art begeht man, wenn man die Sendung akzeptiert, obwohl sie mehr als 4000 defekte Knallkörper enthält, d.h. wenn die 10-er Stichprobe weniger als zwei defekte Knallkörper enthält.

b)

Von den 25 Stichproben in der Aufgabe 2.2.1 (siehe oben) enthalten 13 keine und 7 genau eine schwarze Kugel, d. h. aufgrund von 20 der 25 Stichproben würde man den Fehler 2. Art machen.

Im Histogramm entsprechen diese Stichproben den beiden linken Säulen über 0 und 1.

C

Wenn wir die relative Häufigkeit  $\frac{20}{25}$  als Schätzwert für die Wahrscheinlichkeit des Fehler 2. Art betrachten, müssen wir mit 80% rechnen.

---

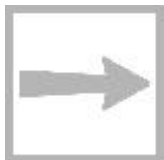
**Aufgabe 1.2.1**

Wenn man den Wert in der Zelle A1 von 0.04 auf 0.08 ändert, werden bei der Simulation Kugeln mit Nummern unter 8000 als schwarz betrachtet. Nach dieser Änderung wird Ihr ganzes Programm neu ausgewertet und Sie erhalten vollautomatisch eine neue Häufigkeitstabelle und ein neues Histogramm. Die Ergebnisse sollten sich nicht allzu sehr von denjenigen in der Lernkontrolle des Kap. 2 unterscheiden.

Dort finden Sie auch einen Hinweis auf die Schätzung von  $\beta$ .

---

## II

**Computersimulationen zu Kapitel 6****Was wird simuliert?**

Eine Urne enthält 100 Kugeln (Fig. 6.3), 93 weisse und 7 schwarze. Es wird eine Stichprobe vom Umfang 55 gezogen und die Anzahl der schwarzen Kugeln gezählt.

**Vergleich mit der früheren Computersimulation**

*Grundsätzlich* macht es keinen Unterschied, ob wir, wie früher, Stichproben mit dem Umfang 200 aus einer Urne mit 100'000 Kugeln ziehen oder, wie jetzt, Stichproben mit dem Umfang 55 aus einer Urne mit 100 Kugeln.

Wir numerieren die Kugeln (diesmal) mit den Nummern 1, 2, ..., 100, wobei wir uns vorstellen, dass die sieben schwarzen Kugeln die Nummern 1, ..., 7 tragen. Mit dem Zufallsgenerator lösen wir 55 *verschiedene* Kugeln aus und zählen die Anzahl der schwarzen.

Wichtig und entscheidend ist der folgende Unterschied: Früher haben wir einfachheitshalber *nicht* untersucht, ob nicht dieselbe Kugel zweimal in die Stichprobe geriet. Diese Nachlässigkeit war angesichts der grossen Kugelzahl und dem verhältnismässig kleinen Stichprobenumfang vertretbar. Diese Argumente treffen hier jedoch offensichtlich nicht mehr zu und deshalb werden die Programme etwas komplizierter.

**Wie wird simuliert?**

In diesem Kapitel finden Sie wieder zwei grundverschiedene Techniken:

- Simulation auf dem Taschencomputer
- Simulation mit Hilfe eines sog. EXCEL-Makros

## II.1

**Simulation auf dem Taschenrechner**

Im Vergleich zum Programm in Kap I.2 müssen wir hier „Protokoll führen“ über die bereits gezogenen Kugeln.

Die unten stehende Fig. II.1 (links) zeigt das Struktogramm eines möglichen Programms. Die fettgedruckte Partie bezieht sich auf das Simulieren der Ziehung *einer* Stichprobe vom Umfang 55. Sie wird protokolliert auf der einzeiligen Protokollmatrix  $p$  mit den 100 Komponenten  $p(1), \dots, p(100)$ : Wenn die Kugel mit der Nummer  $z$  gezogen wird, setzen wir  $p(z)=1$ .

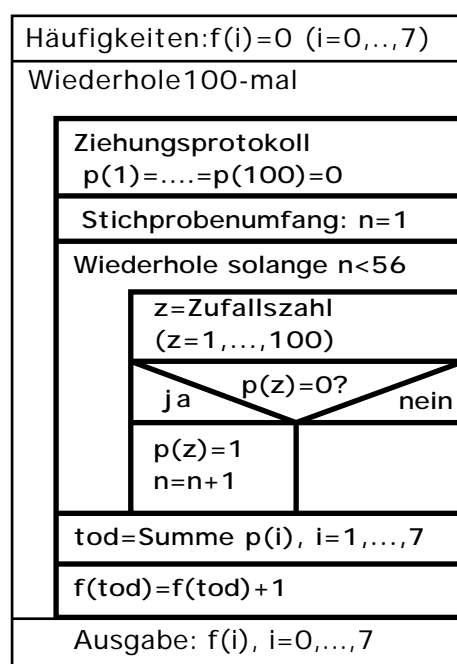
Die Variable  $n$  zählt die Anzahl bereits gezogener Kugeln.

Ob und wie viele der 7 Todeskandidaten sich schliesslich in der Stichprobe befinden, kann man anhand der Protokollplätze  $p(1), \dots, p(7)$  nachprüfen. Ihre Anzahl wird mit der Variablen  $tod$  gezählt:

$$tod = \sum_{i=1}^7 p(i).$$

Sie wird in der Häufigkeitstabelle  $f$  kumuliert.

Zusammen mit dem Struktogramm dürfte der rechts stehende Programmcode (in der Sprache des TI-92) ohne weitere Erläuterung klar sein.



```

Herz()
Prgm
newMat(2,8) ff
For i, 0, 7, 1
    i ff[1,i+1]
EndFor
For i, 1, 100, 1
    newMat(1,100)→p
    1→n
    While n<56
        rand(100)→z
        If p[1,z]=0 Then
            1→p[1,z]
            n+1→n
        EndIf
    EndWhile
    (p[1,i],i,1,7)→tod
    ff[2,tod+1]+1→ff[2,tod+1]
EndFor
Disp ff
EndPrgm

```

Abb. II.1: Struktogramm und Programm für das Simulieren von 100 Stichproben vom Umfang 55 aus einer Urne mit 93 weissen und 7 schwarzen Kugeln. Beachten Sie, dass die Häufigkeiten  $f(0), f(1), \dots$  im Programm auf den Matrixplätzen  $ff[2,1], ff[2,2], \dots$  stehen.

Der Taschenrechner liefert nach ca. einer Viertelstunde die Häufigkeitstabelle  $f$ , beispielsweise wie Fig. II.2.

0	1	2	3	4	5	6	7
0	3	16	30	24	21	3	3

Bei dieser Testserie fällt  $Z$  sogar in 19 von 100 Stichproben in den Verwerfungsbereich  $V=\{0, 1, 2\}$ , d. h.

19%.

Fig. II.2: Die Häufigkeitstabelle auf dem Taschenrechner

## II.2 Simulation mit einem EXCEL-Makro



### Eine neue Idee

Die Simulation auf EXCEL gestalten wir nach einer grundsätzlich neuen Idee: Wir würfeln zu jeder Kugel in der Urne eine Zufallszahl. Die Stichprobe besteht aus denjenigen 55 Kugeln mit den 55 kleinsten Zufallszahlen.

Wir werden diesen Vorgang nur einmal programmieren und lassen ihn dann durch ein sog. MAKRO von EXCEL 100-mal wiederholen. Sie werden alles rasch begreifen, wenn Sie die nachfolgenden Befehle auf Ihrem Computer eintippen und ausführen (lassen).

### Auftrag 1: Startposition der Tabelle

Öffnen Sie in Excel eine neue Tabelle und bearbeiten Sie diese gemäss der nachstehenden Vorlage (Fig. II.3):

- Die Spalte A spielt die Rolle der Urne: Schreiben Sie 7-mal eine „1“ und 93-mal eine „0“ (mittels des Befehls `BEARB / UNTEN AUSFÜLLEN`).
- In der Spalte B stehen 100 Zufallszahlen (in Zelle B1 `=ZUFALLSZAHN()` schreiben und nach unten kopieren).
- Wenn später die Zufallszahlen in der Spalte B samt den zugehörigen „Kugeln“ in der Spalte A geordnet sind, bilden die Kugeln auf den Plätzen A1 bis A55 die Stichprobe. Die Anzahl der schwarzen Kugeln in der Stichprobe ist dann gleich der Summe dieser 55 Zelleninhalte. Wir speichern die Summe in Zelle C1; dazu schreiben wir den Befehl `=SUMME(A1:A55)` in Zelle C1. Momentan sind es 7.
- Die Zellen E1 bis L1 bilden die erste Zeile der Häufigkeitstabelle: Die Zahlen 0, 1, ..., 7 sind die möglichen Anzahlen schwarzer Kugeln in der Stichprobe.
- Die Zellen E3 bis M3 halten die *momentane* Anzahl schwarzer Kugeln als 0-1-Protokoll fest (schreiben Sie in Zelle E3 `=WENN($C$1=E1;1;0)` und kopieren sie diesen Befehl nach rechts). Momentan muss die „1“ auf L3 stehen.
- Zelle N3 enthält schliesslich die Konstante 1 (Erklärung folgt).

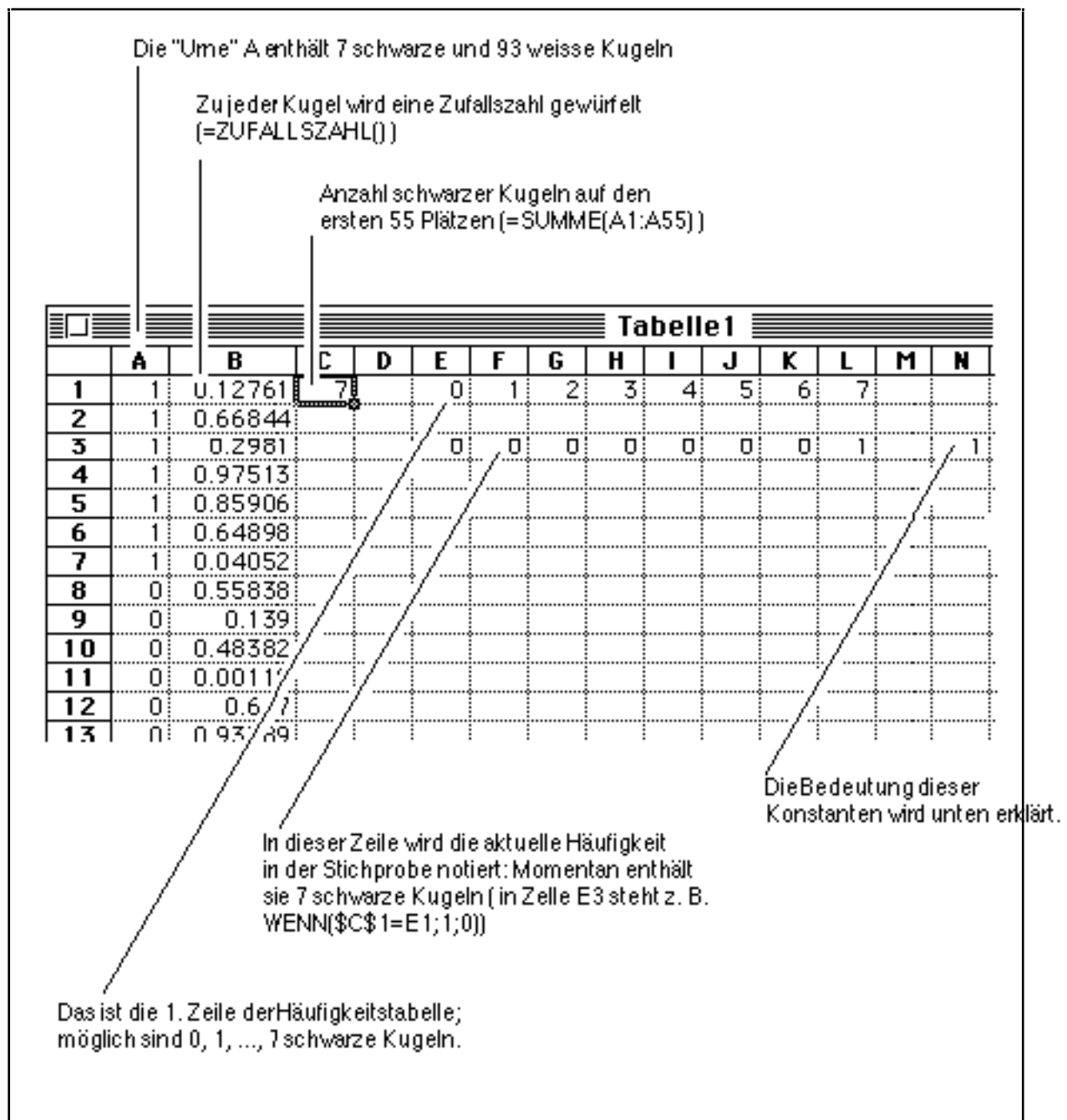


Fig. II.3: Die EXCEL-Tabelle in Startposition

**Auftrag 2: Simulation einer Stichprobe**

In die Stichprobe kommen die Kugeln mit den 55 kleinsten Zufallszahlen in der Spalte B. Wir konstruieren die Stichprobe, wie schon angetönt, mittels des folgenden Tricks: Wir *ordnen* die Zufallszahlen in der Spalte B der Grösse nach, wobei wir die „Kugeln“ in Spalte A parallel umordnen. Im Ergebnis haben wir eine zufällige Verteilung der 7 schwarzen Kugeln auf die 100 Plätze in Spalte A und insbesondere auf die ersten 55 Plätze.

Fig. II.4 zeigt die Tabelle nach einer solchen Umordnung: Die erste schwarze Kugel befindet sich auf Platz 6 und Zelle C1 zeigt, dass 5 schwarze in die Stichprobe geraten sind. Entsprechend steht eine „1“ auf Zelle J3 und alle anderen Plätze der Zeile E3 bis L3 sind mit „0“ belegt.

In der Fig. 6.5 sind die Zufallszahlen gar nicht (mehr) geordnet. Das rührt daher, das EXCEL jedesmal wieder neue Zufallszahlen berechnet, wenn irgendwo eine Zelle, z. B. C1, verändert wird.

Kugeln umgeordnet

Die Stichprobe umfasst nur noch 5 schwarze Kugeln

			Tabelle1												
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	0	0.4968	5		0	1	2	3	4	5	6	7			
2	0	0.1391													
3	0	0.2766			0	0	0	0	0	1	0	0		1	
4	0	0.251													
5	0	0.6371													
6	1	0.858													
7	0	0.5123													
8	0	0.9987													
9	0	0.9785													

Fig. II.4: Tabelle nach Umordnung der Zufallszahlen und Kugeln.

Die Umordnung erreichen Sie in zwei einfachen Schritten:

- Aktivieren Sie die Tabelle der Zufallszahlen und Kugeln, indem Sie mit dem Cursor die Diagonale B1 bis A100 überstreichen (also nicht A1-B100) ....
- ... und nun das Menü **DATEN / SORTIEREN / OK** wählen

### **Hundert Stichproben simulieren**

Die nachstehende Fig. II.5 zeigt das Ergebnis von 100 derartigen Simulationen. Neu, und vielleicht überraschend, ist für Sie die schwarz unterlegte Zeile E2 bis L2, in welcher offensichtlich festgehalten ist, wie oft Stichproben mit 0, 1, ..., 7 schwarzen Kugeln vorkamen. Diese Häufigkeiten ergaben sich nach und nach, indem nach jeder Simulation die unmittelbar darunterliegenden Werte (0 oder 1) in dieser Zeile kumuliert (addiert) wurden.

In der Zelle N2 wurden so gleichzeitig die Anzahl der schon durchgeführten Simulationen gezählt (jedesmal wurde die darunterliegende Konstante 1 addiert).

Anschliessend an die Figur finden Sie die Instruktionen, mit welchen Sie EXCEL mit einem sog. Makro zum Simulieren bringen.



Anzahl der Simulationen

Häufigkeitstabelle

	A	B	C	E	F	G	H	I	J	K	L	M	N	O
1	0	0.3047	4	0	1	2	3	4	5	6	7			
2	0	0.2646		0	2	5	26	32	24	7	4	100		
3	0	0.3273		0	0	0	0	1	0	0	0		1	
4	1	0.675												
5	0	0.73												
6	0	0.7308												

Fig. II.5: Das Ergebnis von 100 Stichproben-Simulationen.

**Auftrag 3: Ein Makro programmieren**

Befolgen Sie gewissenhaft die folgenden Instruktionen:

1

Wählen Sie das Menü **MAKRO / AUFZEICHNUNG BEGINNEN...** Taufen Sie das Makro beispielsweise „Herzinfarkt“.

2

Überstreichen Sie mit dem Cursor die Diagonale von B1 nach A100.

3

Wählen Sie das Menü **DATEN / SORTIEREN / OK**

4

Aktivieren Sie mit dem Cursor die Zeile E3 bis L3.

5

Wählen Sie das Menü **BEARB. / KOPIEREN**

6

Aktivieren Sie mit dem Cursor die Zeile E2 bis L2 (in Fig. II.5 schwarz unterlegt).

7

Wählen Sie das Menü

**BEARB. / INHALTE EINFÜGEN / WERTE / ADDIEREN / OK**

8

Wählen Sie das Menü **MAKRO. / AUFZEICHNUNG BEENDEN**

Die vorstehenden Schritte wurden von EXCEL auf einem eigenen Datenblatt festgehalten, wahrscheinlich unter dem Namen MAKRO 1. Sehen wir uns das mal an!

9

Wählen Sie das Menü **FENSTER. / 1 MAKRO 1**

In einem neuen Fenster erscheint die aufgezeichnete Befehlsfolge (Fig. II.6). Sie sollte ohne weitere Erklärungen verständlich sein.

Makro1'	
	A
1	Herzinfarkt (a)
2	=AUSWÄHLEN("Z1S1:Z100S2";"Z1S2")
3	=ORDNEN(1;"Z1S2";1)
4	=AUSWÄHLEN("Z3S5:Z3S14")
5	=KOPIEREN()
6	=AUSWÄHLEN("Z2S5:Z2S14")
7	=INHALTE.EINFUGEN(3;2;FALSCH;FALSCH)
8	=RUCKSPRUNG()
9	
10	

Fig. II.6: Das automatisch aufgezeichnete EXCEL-Programm (Makro).

#### Auftrag 4: Wiederholung in EXCEL programmieren

Mit dem zusätzlichen Befehl in Zelle A8

=WENN(Tabelle1!\$N\$2=100; GEHEZU(A9);GEHEZU(A2))

bringen wir das Programm dazu, dass die Befehlsfolge 100-mal wiederholt wird, d.h. so oft, bis Zelle N2 der Tabelle1 den Wert 100 hat (Fig. II.7).

Makro1'	
	A
1	Herzinfarkt (a)
2	=AUSWÄHLEN("Z1S1:Z100S2";"Z1S2")
3	=ORDNEN(1;"Z1S2";1)
4	=AUSWÄHLEN("Z3S5:Z3S14")
5	=KOPIEREN()
6	=AUSWÄHLEN("Z2S5:Z2S14")
7	=INHALTE.EINFUGEN(3;2;FALSCH;FALSCH)
8	=WENN(Tabelle1!\$N\$2=100; GEHEZU(A9);GEHEZU(A2))
9	=RUCKSPRUNG()
10	
11	

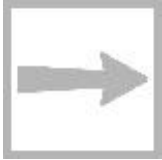
Fig. II.7: Eine ganz einfach programmierte EXCEL-Scheife

**Auftrag 5: Start des Makros**

Nach diesen Vorbereitungen können wir das Makro starten:

- 1 Löschen Sie, falls nötig, von Hand die Zellen E2 bis L2 in der Tabelle 1.
- 2 Wählen Sie in Tabelle 1 das Menü MAKRO / AUSFÜHREN
- 3 Es erscheint ein Dialogfenster. Wählen Sie durch Doppelklick das Makro „Herzinfarkt“. Wenn alles gut geht, simuliert jetzt das Programm, etwa im Rhythmus Ihres Herzschlages, die Stichproben....

## III

**Simulationen zu Kapitel 7****Was wird simuliert?**

Eine Urne enthält sehr viele Kugeln mit einem Anteil von exakt  $p = 2/3$  an schwarzen. Es werden nacheinander 480 Kugeln gezogen. Wir postulieren, dass die Wahrscheinlichkeit für eine schwarze Kugel in jedem Zug gleich gross ist:

$$P(\text{„schwarze Kugel im k. Zug“}) = p$$

**Wie wird simuliert?**

Wir besprechen in den folgende beiden kurzen Abschnitten Lösungen auf EXCEL und auf dem Taschenrechner.

## III.1

**Simulation mit EXCEL**

Das Postulat

$$P(\text{„schwarze Kugel im k. Zug“}) = p$$

lässt sich wie folgt simulieren: Wir erzeugen eine Zufallszahl  $x$  zwischen 0 und 1. Wenn  $x \leq p$ , interpretieren wir dies als Zug einer schwarzen Kugel, andernfalls als Zug einer weissen. 480 solche Zufallszahlen entsprechen somit einer simulierten Stichprobe.

In dieser Weise haben wir schon Stichproben beim Knallkörperstest simuliert (Kap. I). Deshalb sollte hier ein kurzer Hinweis für das Verständnis des Simulationsprogramms genügen: Die Fig.III.1 zeigt einen Ausschnitt aus einer realisierten Simulationstabelle für die Ermittlung von  $p$  (Zelle B1 beinhaltet  $p = 2/3$ ).

Tabelle1													
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	p=	0.7											
2	1	2	3	4	5	6	7	8	9	10	11	12	13
3	325	329	331	330	335	321	324	337	327	300	318	316	324
4													
5	1	1	1	1	1	1	1	0	1	0	1	1	1
6	1	1	0	1	1	1	1	0	1	0	1	1	1
7	0	0	0	0	1	0	1	0	1	1	1	1	1
8	0	0	0	0	1	1	1	0	0	1	1	0	1
9	1	1	0	0	0	1	0	0	1	1	0	1	0
10	0	1	1	1	1	1	1	0	1	1	1	1	0

Fig. III.1: Ein Ausschnitt (oben links) aus der EXCEL-Simulationstabelle.

Die simulierten Stichproben sind nicht wie in Kap. I zeilen- sondern spaltenweise angeordnet, weil in der verwendeten EXCEL-Version eine Zeile nur 256 Zellen umfasst. Sie erkennen von den ersten 13 Stichproben den aller-obersten Teil der Simulation (ab Zeile 5 nach unten) und in der 3. Zeile die Anzahl der schwarzen Kugeln in diesen Stichproben (beispielsweise enthält die 3. Stichprobe 331 schwarze Kugeln).

Letzte Klarheit über den Aufbau der Simulationstabelle vermittelt das Befehlsgerüst (Fig. III.2).

Tabelle1				
	A	B	C	D
1	p=	=2/3		
2	1	2	3	4
3	=SUMME(A5:A484)	=SUMME(B5:B484)	=SUMME(C5:C484)	=SUMME(D5:D484)
4				
5	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)
6	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)
7	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)
8	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)
9	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)	=WENN(ZUFALLSZAHL()<=\$B\$1;1;0)

Fig. III.2: Bauplan des EXCEL-Simulationsprogrammes

Vollständig analog zum früheren Vorgehen kann man aus der Zeile 3 der Simulationstabelle die Häufigkeitstabelle zusammenstellen und davon ein Histogramm zeichnen lassen. Konsultieren Sie allenfalls das Kap. I.

## III.2

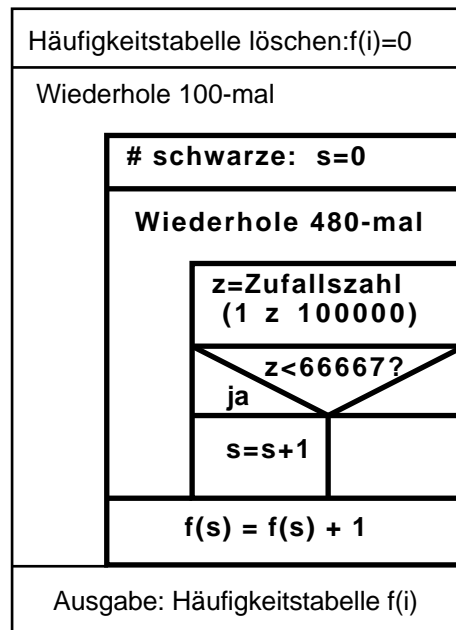
### Simulation auf dem Taschencomputer

Auch auf dem Taschenrechner lässt sich das Postulat

$$P(\text{„schwarze Kugel im k. Zug“}) = p = \frac{2}{3}$$

folgendermassen simulieren: Wir erzeugen eine Zufallszahl  $x$  zwischen 1 und 100'000 (Grenzen eingeschlossen). Wenn  $x < 66'667$ , interpretieren wir das als Zug einer schwarzen Kugel, andernfalls als Zug einer weissen. 480 solche Zufallszahlen entsprechen somit einer simulierten Stichprobe.

In dieser Weise haben wir bereits Stichproben beim Knallkörper-test simuliert. Deshalb sollte das folgende Struktogramm und der zugehörige Programmcode (für den TI-92) ohne weitere Hinweise verständlich sein (Fig. III.3):



```

simulier()
Prgm
newMat(2,481) f
For i, 0, 480, 1
    i f[1,i+1]
EndFor
ClrIO
For j, 1, 100,1
    0→s
    For i, 1, 480, 1
        If rand(100000)<66667 Then
            s+1→s
        EndIf
    EndFor
    f[2,s+1]+1→f[2,s+1]
EndFor
Disp f
EndPrgm
  
```

Fig. III.3: Struktogramm und Programm für die Simulation von 100 Stichproben mit dem Umfang 480 bei konstanter Wahrscheinlichkeit für eine schwarze Kugel.

Der Rechner präsentiert nach sehr langer Rechenzeit die nebenstehende Häufigkeitstabelle. Leider ist es etwas mühsam, sich bis zum interessanten Teil in der Umgebung von 320 „durchzukämpfen“.

```

0  1  2  3  4  5  6  7  8  9
0  0  0  0  0  0  0  0  0  0 ▶
  
```

***Anhang K***  
***für die Lehrperson***

***Kapiteltests***

## Kapiteltest 1: Pilzkontrolle

Verdorbene Pilze können giftig sein. Deswegen möchte ein Grossverteiler weitgehend sicherstellen, dass alle verkauften Pilze einwandfrei sind. Die Pilze werden lastwagenweise in jeweils 1500 Kistchen importiert. In einer ersten Kontrolle werden 10 zufällig ausgewählte Kistchen geprüft. Wenn auch nur ein Kistchen verdorbene Pilze enthält, wird die ganze Sendung zurückgewiesen.

a

Weshalb kontrolliert man nicht einfach *alle* Kistchen?

b

Beschreiben Sie für dieses Beispiel die Grundgesamtheit und die Stichprobe.

c

Zeichnen Sie ein Urnenmodell für diese Untersuchung. Welche Rolle spielen die schwarzen Kugeln? Welche der Grössen  $N$ ,  $S$ ,  $n$ ,  $s$  ist unbekannt und gesucht? Welche Bedeutungen und Werte haben die anderen Grössen?

d

Was ist hier zum möglichen Fehler 1. Art zu sagen? Wie gross ist seine Wahrscheinlichkeit ?

e

Worin besteht hier ein Fehler 2. Art?



## Kapiteltest 2: Gewicht von Tomatenkonserven

Auf einer Verpackungsanlage werden stündlich 3600 Dosen mit Tomaten gefüllt. Es kann nicht ganz verhindert werden, dass gewisse Dosen ein zu kleines Abtropfgewicht aufweisen. Der Produktionsprozess müsste jedoch gestoppt werden, wenn der Anteil an zu leichten Dosen über 2% steigt. Um diesen Qualitätsstandard zu halten, werden stündlich 100 zufällig ausgewählte Dosen kontrolliert.

a

Der Test soll mittels einer Computersimulation untersucht werden. Erklären Sie anhand eines Urnenmodells, was und wie grundsätzlich simuliert werden soll, wenn es darum geht, den Fehler 1. Art herauszufinden. (Wie viele Kugeln enthält die Urne? Wie viele schwarze? Welchen Umfang haben die Stichproben? Was wird notiert? etc)

b

Ab welcher Anzahl zu leichter Dosen sollte der Produktionsprozess unterbrochen werden, wenn die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 5% betragen soll? Entscheiden Sie aufgrund der nachstehend zusammengestellten Simulationsergebnisse.

(100 Simulationen,  $x$ =Anzahl schwarzer Kugeln,  $f_x$ = Anzahl Stichproben mit  $x$  schwarzen Kugeln).

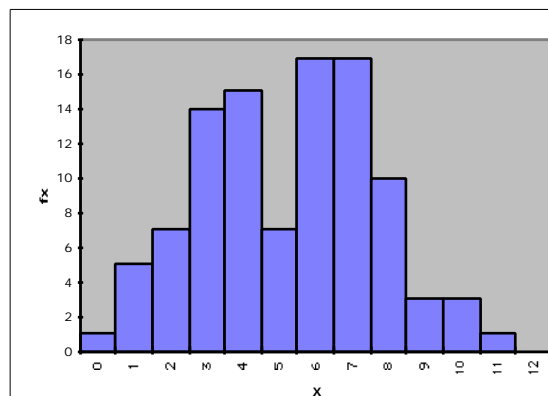
$x=$	0	1	2	3	4	5	6	7	8	9	10	11	12
$f_x=$	14	26	33	22	4	0	1	0	0	0	0	0	0

c

Der Test soll mittels einer Computersimulation auf den Fehler 2. Art untersucht werden. Erklären Sie anhand des Urnenmodells, was und wie grundsätzlich simuliert werden soll, wenn man sich für einen Schätzwert von bei einem Anteil von 5% zu leichter Dosen interessiert?

d

Aufgrund Ihrer Angaben erhält man bei der Simulation z. B. eine Häufigkeitstabelle mit dem nebenstehenden Histogramm. Welcher Schätzwert ergibt sich daraus für , wenn der Produktionsprozess gemäss Ergebnis in Aufgabe b unterbrochen wird ?



## Kapiteltest 3: Therapien nach Herzinfarkt

Ein Herzinfarkt ist die Folge von verstopften Arterien, welche den Herzmuskel mit Blut versorgen (sollten). Mehr als zwei Drittel aller Patienten überleben die unmittelbaren Folgen eines Herzinfarkts. Wenn sich ihr Zustand stabilisiert hat, gilt es, die verstopften Gefässe möglichst schnell wieder durchgängig zu machen, so dass der Schaden am Herzmuskel beschränkt bleibt. Zwei Methoden stehen im Vordergrund: Die Behandlung mit Medikamenten, welche den Blutpfropf wieder auflösen und die primäre Angioplastie. Bei letzterer wird im Spital innert Stunden nach dem Herzinfarkt durch einen kleinen operativen Eingriff ein Ballonkatheter eingeführt, mit welchem die verstopfte Stelle ausgeweitet wird.

Eine kleine Studie hat die folgenden Ergebnisse ergeben:

Behandlungsmethode→ Erfolg ↓	Angioplastie	Medikamente	Summe
Patient überlebt	53	40	93
Patient stirbt	2	5	7
Summe	55	45	100

Für diese Aufgabe postulieren wir, die beiden Behandlungsmethoden seien exakt gleich wirksam. Die insgesamt sieben Patienten in der zweiten Zeile wären also so oder so gestorben. Unter dieser Voraussetzung können wir die 100 untersuchten Patienten als eine Grundgesamtheit betrachten, aus der zufällig eine Stichprobe mit 55 Patienten für die Behandlung durch Angioplastie ausgewählt wurde.

a

Entwerfen Sie ein Urnenmodell für die geschilderte Situation.

b

Wie gross ist die Wahrscheinlichkeit, dass gerade zwei der sieben Todeskandidaten in diese Gruppe kamen?

c

Berechnen Sie analog: Wie gross ist die Wahrscheinlichkeit, dass in die zweite Gruppe mit den 45 Patienten gerade fünf Todeskandidaten geraten?

d

Was stellen Sie fest? Wie ist Ihre Beobachtung zu erklären?

Wir werden in einem eigenen Kapitel auf dieses Beispiel zurückkommen; dort werden Sie dann auch noch zusätzliche Informationen erhalten.

## **Kapiteltest 4: Produktionskontrolle bei Kugelschreiberminen**

Auf einer Produktionsanlage werden stündlich 3600 Kugelschreiberminen angefertigt. Es kann nicht verhindert werden, dass gewisse Exemplare nicht funktionieren. Der Produktionsprozess müsste jedoch gestoppt werden, wenn der Ausschussanteil 2% übersteigt. Um diesen Qualitätsstandard zu halten, werden stündlich unter 200 zufällig ausgewählte Minen die Anzahl der defekten ermittelt.

a

Bei welcher Anzahl defekter Minen sollte der Produktionsprozess unterbrochen werden, damit der Test mit 90% Sicherheit einen Ausschussanteil von 5% entdeckt?

Lösen Sie diese neue Aufgabe durch sinnvolles Probieren.

b

Wie gross ist dann die Wahrscheinlichkeit für den Fehler 1. Art?

## Kapiteltest 5: Max & Moritz werfen eine Münze

**Eine Variante des Max&Moritz-Problems! Achtung: Der kleine Unterschied hat weitreichende Konsequenzen für die Antworten!**

Max glaubt, daß eine von Moritz verwendete Münze **nicht gleich oft** "Zahl" und "Bild" zeigt. Er beschliesst, die Münze mit 20 Würfeln zu testen.

a

Machen Sie einen Vorschlag: Welche Entscheidungsregel soll sich Max im *voraus* zurechtlegen? Formulieren Sie vielleicht zuerst umgangssprachlich. Aber versuchen Sie dann, mit Hilfe des Begriffs "Testgrösse" zu antworten.

b

Auf welchen Verwerfungsbereich  $V$  wird sich also Max abstützen? Skizzieren Sie  $V$  auf einem Zahlenstrahl.

c

Wie lauten bei diesem Test die Nullhypothese  $H_0$  und die Alternativhypothese  $H_1$ ?

## Kapiteltest 6: Kriminelle Zwillinge

Es gibt zwei Sorten von Zwillingen. Die eineiigen Zwillinge sind identische Kopien voneinander. Sie haben dasselbe genetische Material. Dagegen haben zweieiige Zwillinge nicht mehr gemeinsames genetisches Material als gewöhnliche Geschwister.

Ist Kriminalität durch Umwelteinflüsse oder durch genetische Faktoren bedingt? Hier ein paar Fakten:

- Von 13 Kriminellen, die *eineiige* Zwillinge waren, hatten 10 davon Zwillingenbrüder oder -schwestern, die ebenfalls verurteilt waren, während in 3 Fällen der Zwillingspartner anscheinend nicht kriminell war.
- Von 17 Kriminellen, die *zweieiige* Zwillinge (gleichen Geschlechts!) waren, hatten 2 von ihnen verurteilte Zwillingenpartner, während die anderen 15 nicht als kriminell bekannt waren (zitiert nach Engel, [2]).

a

Fassen Sie die Daten in einer Vierfeldertafel zusammen.

b

Eineiige Zwillinge und zweieiige Zwillinge gleichen Geschlechts stehen unter den gleichen Umwelteinflüssen. Wenn sie verschiedenes soziales Verhalten aufweisen, dann könnte dies genetisch bedingt sein. Formulieren Sie die Nullhypothese für dieses Datenmaterial und führen Sie einen geeigneten statistischen Test durch. Interpretieren Sie das Ergebnis im Hinblick auf eine genetisch bedingte Ursache der Kriminalität.

## **Kapiteltest 7: ESP = Extrasensory Perception**

Es gibt immer wieder Personen, die durch ihre (scheinbare?) ausserordentliche Wahrnehmungsfähigkeit auffallen. Schon oft wurde mit einem statistischen Test überprüft, ob die Leistungen dieser ESP-Personen das Prädikat "ausserordentlich" wirklich verdienen.

Hier der Bericht über einen tatsächlich durchgeführten Versuch:

Die ESP-Person musste erraten, welchen von vier Videofilmen sich eine bestimmte Person, der sog. Sender, im hermetisch abgeriegelten Nebenzimmer gerade anschaute. In 61 von 165 Versuchen nannte die ESP-Person den richtigen Videofilm.

a

Ist dieses Testergebnis erklärbar mit dem Verdacht, dass die ESP-Person jeweils einfach zufällig geraten hat oder spricht es signifikant zugunsten von ESP-Fähigkeit?

Lösen Sie die Aufgabe regelkonform mit Hilfe der Ihnen nun sicher geläufigen statistischen Begriffe.

b

Welche Versuchsergebnisse würden signifikant für ESP-Fähigkeiten sprechen? Antworten Sie bitte durch Angabe eines Verwerfungsbereiches.

c

Wie gross ist beim gewählten Verwerfungsbereich die Macht des Tests, zu "erkennen", dass die ESP-Person fähig ist, durchschnittlich jeden zweiten Film richtig zu erraten.

## Kapiteltest 8: Der Einfluss des rooming-in

Sollen Spitäler die Eltern kranker Kinder mit aufnehmen (rooming-in Modell), um die Verweildauer abzukürzen?

Diese Frage versuchte man im Jahre 1983 aufgrund der folgenden Untersuchung zu beantworten: Aus 50 jugendlichen Patienten wurden 25 Paare gebildet, die in den wichtigen Belangen (Geschlecht, Alter, Krankheit, Behandlung etc.) vergleichbar waren. Von einem Kind jedes Paares wurden die Eltern mit aufgenommen, vom anderen nicht. Die folgende Tabelle zeigt die von jedem Kind in der Klinik verbrachten Tage (zitiert nach Engel, [2]).

Paar-Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13
ohne rooming-in	29	44	18	15	15	7	15	11	12	12	7	11	7
mit rooming-in	29	32	16	8	8	7	14	8	7	12	7	8	7

Paar-Nr.	14	15	16	17	18	19	20	21	22	23	24	25	
ohne rooming-in	22	7	14	28	31	18	16	10	11	29	8	23	
mit rooming-in	17	6	19	12	8	8	6	10	8	21	17	15	

Je nachdem, welche Vermutung man hat, wird man *einen* der beiden folgenden Tests durchführen. Die beiden Teilaufgaben sind daher völlig unabhängig von einander.

a

Klären Sie mit einem Vorzeichentest ab, ob es einen signifikanten *Unterschied* in der Verweildauer der beiden Kindergruppen gibt.

b

Entscheiden Sie mit einem neuen Vorzeichentest, ob die Verweildauer der rooming-in Patienten signifikant *kleiner* ist als bei der Kontrollgruppe.

## Quellen

### **Lehrbücher**

- [1] Engel, Arthur: Wahrscheinlichkeitsrechnung und Statistik 1 & 2.  
Verlag Ernst Klett, Stuttgart; 1. Auflage 1975. ISBN 2-12-983160-6.
- [2] Engel, Arthur: Stochastik.  
Verlag Ernst Klett, Stuttgart; 1. Auflage 1987. ISBN 2-12-983110-X.
- [3] Ineichen, Robert: Elementare Beispiele zum Testen statistischer Hypothesen. Verlag Orell Füssli, Zürich; 2. Auflage 1978.  
ISBN 3 280 00900 6.
- [4] Keiser, Otto M.: Wahrscheinlichkeitsrechnung und Statistik 1 bis 8.  
Verlag AKAD, Zürich; 1. Auflage 1993.
- [5] Kreyszig, Erwin: Statistische Methoden und ihre Anwendungen.  
Verlag Vandenhoeck & Rupprecht, Göttingen; 6. Auflage 1977.  
ISBN 3-525-40717-3.
- [6] Sobol, I.M.: Die Monte-Carlo-Methode.  
Verlag DVW, Berlin; 4. Auflage 1991. ISBN 3-326-00635-7.

### **Manuskripte, Artikel, Vorträge**

- [7] Ineichen, Robert: Wie könnte man Tests in der Statistik einführen?  
Vortrag an der ETH; 1981/82
- [8] Künsch, Hans-Ruedi: Simulation. Vorlesungsmanuskript
- [9] NZZ: Behandlung nach dem Herzinfarkt. 8. Januar 1997, Seite 51.
- [10] Simon, Julian L. u. a.: THE NEW BIOSTATISTICS OF RESAMPLING.  
Manuskript

### **Handbücher**

- [11] Boonin Elisabeth: Using Excel Visual Basic for applications.  
Indianapolis, IN ; 1996. ISBN 0-7897-0325-4.



