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 α 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, eine 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 *Null*hypothese). 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: "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, S=7, n=55, s=2

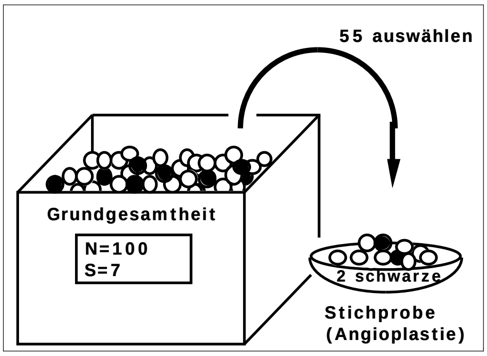


Fig. 6.1: Das Urnenmodell der Vierfeldertafel unter Annahme der Nullhypothese H_0.

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, ..., g}, d. h. eine Zahl g so berechnen, dass wir H_0 nur mit einer Irrtumswahrscheinlichkeit $\alpha \leq 5\%$ 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 α. Ergibt sich – wider Erwarten – ein Wert α > 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, ... , 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 α 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 α 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 $\alpha \leq 5\%$ in den Verwerfungsbereich

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

fallen würde.

Berechnen wir α:

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

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 α auf dem Taschencomputer berechnen (Fig. 6.2):

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

$$\sum_{s=0}^{2} \left(\frac{\mathrm{nCr}(7,s)\ \mathrm{nCr}(93,55-s)}{\mathrm{nCr}(100,55)} \right) \qquad 144\ldots E\text{-}1$$

Fig. 6.2: Irrtumswahrscheinlichkeit α

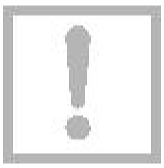
Die Rechnung bestätigt also die Ergebnisse der Simulation: Die Untersuchung spricht, immer aus der Sicht der Statistik, nicht signifikant zugunsten der Angioplastie.

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 α für den Verwerfungsbereich V={0, 1}:



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

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.