3c) Einführung in das Testen von Hypothesen

Eine Krankheit wird mit bisher bekannten Mitteln mit einer Wahrscheinlichkeit von 60 % geheilt. Von einem neuen Medikament wird nach Vortesten behauptet, dass es höhere Heilungsquoten habe.

Anhand eines Tests an 80 (freiwilligen) Patienten soll nun über die Markteinführung entschieden werden.

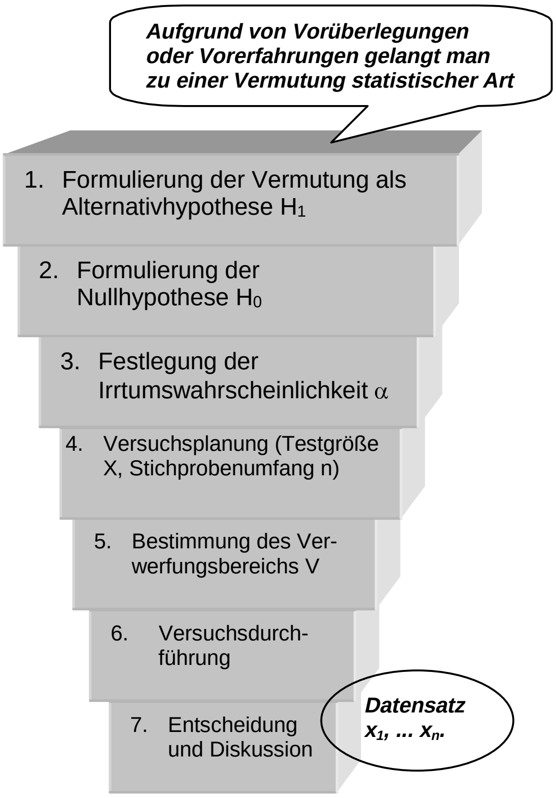
- Die Hypothese H_1 (Das neue Medikament ist besser; $p > 60$ %) ist zu akzeptieren oder zu verwerfen.
  Da p nicht bekannt ist, wird die Gegenhypothese H_0 (genannt Nullhypothese: Das neue Medikament ist nicht besser als die bisherigen Mittel; $p = 60$ %) geprüft. Es sollte durch Vorteste ausgeschlossen sein, dass durch das neue Medikament sogar ein schlechteres Ergebnis herauskommen könnte als mit den bisherigen Mitteln.
- Weichen die Ergebnisse des Testes deutlich nach oben ab, so ist es plausibel anzunehmen, dass für das neue Medikament nicht $p = 0{,}6$ zutrifft, sondern $p > 0{,}6$. Dann würde H_0 verworfen.
- Da $p > 0{,}6$ getestet wird, nennt man den Test rechtsseitig.
- Liegen die Testergebnisse im "normal zu erwartenden" Bereich für $p = 60$ %, so handelt es sich wohl nur um Zufallsschwankungen und H_0 ist akzeptabel. Dann wird H_0 nicht verworfen und H_1 nicht akzeptiert.
- Für die Abweichung "deutlich nach oben" bzw. für den "normal zu erwartenden" Bereich wird bei Medikamententests i. d. R. eine Sicherheitswahrscheinlichkeit von 95 % gefordert, bzw. ein "Restrisiko" von $\alpha = 5$ % in Kauf genommen.
- Die Testplanung für das Beispiel:
  H_1: $p > 0{,}6$ (rechtsseitig); H_0: $p = 0{,}6$; $\alpha = 5$ %; X: Zahl der Patienten, bei denen Verbesserung eintritt; X ist binomialverteilt mit $p = 0{,}6$ und $n = 80$:
  $P(X \leq 54) = 93{,}3$ %; $P(X \leq 55) \approx 95{,}8$ %.
  $P(X = 55) = 2{,}5$ % "überdeckt" gerade die geforderte 95 %-Sicherheitswahrscheinlichkeit. Ergebnisse ab $k = 56$ führen zu einem Verwerfen von H_0. Deshalb lautet der sogenannte Verwerfungsbereich $V = \{56, 57, ..., 80\}$. Liegt das Testergebnis in V, so wird H_0 verworfen und H_1 akzeptiert. Liegt es nicht in V, so wird H_0 nicht verworfen und H_1 nicht akzeptiert.
- Mit diesen Vorgaben kann der Test durchgeführt und das Ergebnis bewertet werden. Die Testdaten führen zur Entscheidung für oder gegen die Markteinführung des Medikaments.

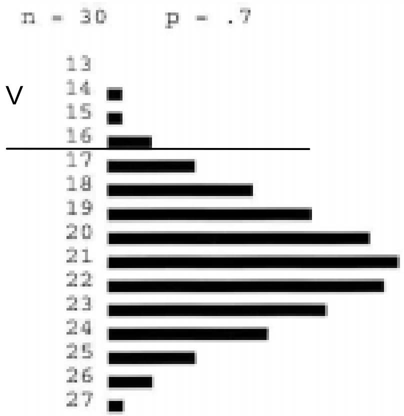
Testturm

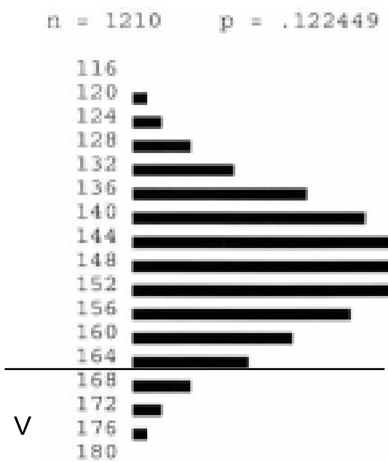
Linksseitiger Test:

Ein Medikament soll auf den Markt kommen, von dem behauptet wird, dass es den Ausbruch einer Krankheit besser verhindere als bisherige Mittel, bei denen es in 70 % der Fälle doch zu einem Ausbruch der Krankheit kommt.

1. H_1: Das neue Medikament senkt die Zahl der Erkrankungen; $p < 0{,}7$ (linksseitig).
2. H_0: Das neue Medikament wirkt wie die bisherigen; $p = 0{,}7$.
3. $\alpha = 5$ %.
4. X: Anzahl der Patienten, bei denen nach der Medikamentengabe die Krankheit nicht ausbricht
$n = 30$
Die Zufallsgröße X ist binomialverteilt.
5. $P(X \leq 16) \approx 4{,}0$ %
$P(X \leq 17) \approx 8{,}5$ %
$V = \{0, 1, ..., 15, 16\}$.
6. Beim Test ergeben sich z. B. a) 17 Erkrankte,
b) 14 Erkrankte.
7. a) Da $17 \notin V$, wird H_0 nicht verworfen und H_1 nicht akzeptiert. Das neue Medikament wird nicht als besser akzeptiert (bei einer Irrtumswahrscheinlichkeit von 5 %).
b) Da $14 \in V$, wird H_0 verworfen und H_1 akzeptiert. Das Medikament wird als besser akzeptiert und für den Markt zugelassen ($\alpha = 5$ %).







Rechtsseitiger Test:

Ist 49 eine Glückszahl? Unter 996 Ziehungen vor 1974 war die 49 die Zahl mit der höchsten Ziehungsanzahl.

1. H_1: 49 kommt häufiger vor als üblich; $p > \frac{6}{49}$ (rechtsseitig).
2. H_0: 49 ist eine normale Zahl; $p = \frac{6}{49}$.
3. $\alpha = 5$ %.
4. X: Zahl der Ziehungen mit 49.
$n = 1210$; Anzahl der Ziehungen von November 1974 bis Februar 1998.
X ist binomialverteilt.
5. $P(X \leq 166) \approx 94{,}4$ %
$P(X \leq 167) \approx 95{,}3$ %
$V = \{168, 169, ..., 1209, 1210\}$.
6. Die 49 kam in den Ziehungen 163 mal vor.
7. Da $163 \notin V$, wird H_0 nicht verworfen und H_1 nicht akzeptiert. 49 ist eine normale Zahl (bei $\alpha = 5$ %).

6. b) Angenommen, 49 kam 172 mal.

7. b) Da dann $172 \in V$, wird H_0 verworfen und H_1 akzeptiert. 49 ist eine Glückszahl ($\alpha = 5$ %).

Arbeitsblatt: Kindersicherheit und Schnupfen

a) Berechne die Wahrscheinlichkeit, dass Fingerhut und Plecher bei 100 kontrollierten Autos genau (mindestens) 40 Zitronen verteilen konnten. Nimm an, dass die Ergebnisse der zitierten Studie allgemein gelten.
b) Ist mit dem Habenhausener Ergebnis die vom ADAC zitierte Studie überholt ?
c) Der Zeitungsbericht lässt offen, ob sich die Eltern tatsächlich verantwortungsbewusster verhalten. Entwickle daher einen einseitigen Hypothesentest auf der Basis einer 300er Stichprobe. (Irrtumswahrscheinlichkeit 5%).
d) Wie würdest du dich entscheiden, wenn bei dem anschließenden Test 120 Kinder nicht ordnungsgemäß gesichert wären?
e) Welcher Fehler kann dabei auftreten? Bitte beschreibe ihn auch inhaltlich.
f) Wie groß ist die Fehlerwahrscheinlichkeit?

Zitronen statt Strafzettel für fehlende Kindersitze im Auto

ADAC und Polizei wiesen gemeinsam auf Sicherheitsmängel hin

"Das Kissen? Das hat meine Frau im Auto." Um Ausreden waren die Väter und Mütter nicht verlegen, die ihre Kinder ohne vorschriftsmäßige Sicherung zur Schule brachten und dabei von ADAC und Polizei auf den Missstand hingewiesen wurden. Wer ohne Kindersitz, Kissen oder speziellen Gurt an der Schule vorbeifuhr, bekam diesmal keinen Bon – dafür aber eine saure Zitrone überreicht.
(...) So viele Früchte wie erwartet konnten Wolfgang Fingerhut und Bernd Plecher vom ADAC aber nicht verteilen. (Sie) beziehen sich auf eine Erhebung, nach der die Hälfte aller Kinder im Alter zwischen 6 und 12 Jahren im Auto nicht den Bestimmungen entsprechend geschützt sind. In Habenhausen war an diesem Morgen um 8 Uhr die Ausbeute deutlich geringer, die meisten Fahrzeuge waren vorschriftsmäßig ausgerüstet ...

aus dem Weserkurier vom 28.09.1995

In Großbritannien gab es langjährige Untersuchungen von 1957 bis 1990 zum Schnupfen, bei denen Freiwillige ein Virus verabreicht bekamen. Es ergab sich immer mit und ohne Behandlung ein Drittel erkrankter Probanden. Das deutsche Gesundheitsministerium möchte wissen, ob das in der BRD genauso ist.

Salzwasser spült die Erkältung weg

(mm) Richard Ravizza von der Pennsylvania State University machte mit 294 College-Studenten den Schnupfentest. Eine Gruppe musste täglich mit Salzwasser spülen, während die anderen ein Scheinmedikament erhielten oder unbehandelt blieben. Die Spülungen führten dazu, dass die Studenten wesentlich seltener erkältet waren. Die Vorgehensweise stammt aus der indischen Yoga-Lehre: Kopf zurücklegen und warmes Salzwasser in ein Nasenloch füllen, anschließend wieder herauslaufen lassen.

Westfälische Rundschau, 1998

a) Entwickle einen Hypothesentest mit n = 300 und einer Irrtumswahrscheinlichkeit von 5%. Formuliere eine Entscheidungsregel.
b) Bei der Durchführung werden 111 Schnupfenkranke gezählt. Was wirst du dem Gesundheitsministerium mitteilen?
c) Bei der Abgabe der Ergebnisse wirst du von der Ministerin gefragt: "Und das ist jetzt sicher?" Formuliere eine Antwort.
d) Gleichzeitig schlägst du mit Hinweis auf den nebenstehenden Zeitungsbericht als Kostendämpfungsmaßnahme die morgendliche Salzwasserspülung vor. Das Verfahren könnte als Standardmethode in der 5. Klassenstufe im Fach Biologie eingeübt werden. "Ich glaube nicht, dass das hilft ...", meint die Ministerin. Aber du bekommst wieder einen Auftrag für einen Hypothesentest. Aus Kostengründen dürfen diesmal leider nur 100 Leute getestet werden. Formuliere die Testlogik.

"Kinder im Auto"

a) X gibt die Anzahl der verteilten Zitronen an und ist binomialverteilt ($n = 100$ und $p = 0{,}5$).
   $P(X = 40) = 1{,}08\ \%$ und $P(X \geq 40) = 1 - P(X \leq 39) = 1 - 1{,}76\ \% = 98{,}24\ \%$
   Die Wahrscheinlichkeit, dass genau 40 Zitronen verteilt werden, ist (natürlich) mit rund 1% sehr niedrig. Dagegen ist die Wahrscheinlichkeit, mehr als 40 Zitronen zu verteilen, mit mehr als 98% sehr hoch. Schließlich sind 50 Stück zu erwarten.
b) Nein. Es kann Zufall sein, dass an diesem Morgen weniger vorschriftsmäßig angeschnallte Kinder entdeckt wurden ... Oder einer der erwischten Eltern hat einen Rundruf gestartet ...
c) Es wird ein linksseitiger Test durchgeführt:
   1. H_1: $p < 0{,}5$ (Weniger Kinder sind nicht richtig angeschnallt.)
   2. H_0: $p = 0{,}5$ (alles beim Alten – siehe Kontrolle in Habenhausen)
   3. Irrtumswahrscheinlichkeit $\alpha = 5\ \%$
   4. X sei die Anzahl der Kinder, die nicht richtig gesichert waren. X ist binomialverteilt. Es werden 300 Eltern-Pkw kontrolliert.
   5. $P(X \leq 135) \approx 4{,}7\ \%$
      $P(X \leq 136) \approx 5{,}9\ \%$
      $V = \{0, ..., 135\}$.

   Entscheidungsregel: Findet man also bei dieser Kontrolle höchstens 135 Kinder, die nicht richtig gesichert sind, so wird Ho verworfen und H_1 (die Eltern sind in ihrer Gesamtheit vernünftiger geworden) als richtig angenommen.
d) Entsprechend der getroffenen Festlegung wird H_0 verworfen und H_1 akzeptiert.
e) Wenn Ho doch richtig ist, liegt ein α-Fehler vor. Man nähme dann zu unrecht an, dass weniger als die Hälfte der Kinder nicht hinreichend geschützt sind.
f) Nach Testkonstruktion ist $\alpha \leq 5\%$, nachgerechnet ergibt sich: $P(X \leq 135) = 0{,}047$.

"Schnupfen"

a) 1. H_1: Die Schnupfenempfindlichkeit der Deutschen unterscheidet sich von der der Briten; $p \neq \frac{1}{3}$ (zweiseitiger Test)
   2. H_0: $p = 1/3$
   3. Irrtumswahrscheinlichkeit $\alpha = 5\ \%$. Sie ist zur Hälfte (2,5 %) nach "links" und zur Hälfte (97,5 %) nach "rechts" zu verteilen.
   4. X sei die Anzahl der Schnupfenkranken nach Virengabe, X ist binomialverteilt. Es werden 300 Leute gesucht, die sich freiwillig mit dem Virus impfen lassen.
   5. $P(X \leq 83) \approx 2{,}1\ \%$; $P(X \leq 84) \approx 2{,}8\ \%$
      $P(X \leq 115) \approx 97{,}0\ \%$; $P(X \leq 116) \approx 97{,}7\ \%$
      $V = \{0, 1, ..., 83, 117, 118, ..., 300\}$.
      Verwerfe H_0, wenn bis zu 83 oder mindestens 117 Erkrankte auftreten.
b) Auf Grund des Testergebnisses wird (weiter) davon ausgegangen, dass die Menschen in Deutschland genauso schnupfenempfindlich sind wie in Großbritannien.
c) Tests liefern niemals sichere Ergebnisse. Es könnte passieren, dass die Schnupfenempfindlichkeit in Deutschland doch von der in Britannien unterschiedlich ist, wir das auf Grund einer ungewöhnlichen Zufallsschwankung im Test aber nicht merken, d.h. wir haben irrtümlich die falsche Nullhypothese beibehalten. Andererseits wäre es möglich, dass wir irrtümlich die richtige Nullhypothese verwerfen (α-Fehler).
d) Diesmal braucht nur ein einseitiger Hypothesentest durchgeführt werden, denn nur wenn die Schnupfenanfälligkeit dabei sinkt, macht das Verfahren Sinn:
   1. H_1: $p < 1/3$ (Die Salzwassermethode hilft.)
   2. H_0: $p = 1/3$ (Es bleibt alles beim Alten.)
   3. Irrtumswahrscheinlichkeit $\alpha = 5\ \%$
   4. X sei die Anzahl der Menschen, die nach der Salzwassermethode Schnupfen bekommen. X ist binomialverteilt. Gesucht werden 100 Testpersonen, die sich dem Verfahren unterziehen und sich regelmäßig auf Schnupfen untersuchen lassen.
   5. $P(X \leq 25) \approx 4{,}6\ \%$
      $P(X \leq 26) \approx 7{,}2\ \%$
      $V = \{0, 1, ..., 25\}$.
      Entscheidungsregel: Treten höchstens 25 Schnupfen-Kranke auf, so wird H_0 verworfen und H_1 (die Salzwassermethode ist erfolgreich) als richtig angenommen.

Arbeitsblatt: Fehler beim Testen von Hypothesen

Beim Testen sind zwei Fehlerarten schon von der Testkonstruktion "mit eingebaut".

α-Fehler

Wenn man sich bei Testen dazu entschließt, H_0 zu verwerfen und H_1 als richtig zu akzeptieren, so geschieht das auf Grund eines Versuchs oder der Erhebung von Daten. Der Wert der Testgröße liegt im Verwerfungsbereich. Dies kann daran liegen, dass H_1 tatsächlich richtig ist – oder daran, dass durch Zufall ein ungewöhnlich seltenes (man sagt auch signifikantes) Ergebnis herausgekommen ist. Wie groß die Wahrscheinlichkeit dafür ist, dass so etwas passiert, wird vor Testbeginn durch Vorgabe einer Irrtumswahrscheinlichkeit (man sagt auch Signifikanzniveau) festgelegt. In einem solchen Fall macht man einen Fehler 1. Art (α-Fehler), weil man H_0 verwirft, obwohl diese Hypothese eigentlich richtig ist ...

β-Fehler

Wenn man sich bei Testen dazu entschließt, H_0 beizubehalten und H_1 nicht zu akzeptieren, so geschieht das auf Grund eines Versuchs oder der Erhebung von Daten. Der Wert der Testgröße liegt in diesem Fall nicht im Verwerfungsbereich. Dies kann daran liegen, dass H_0 tatsächlich richtig ist oder per Zufall ein solches Ergebnis zustande gekommen ist.
Im letzteren Fall machen wir einen Fehler 2. Art (β-Fehler), weil wir H_0 beibehalten, obwohl eigentlich H_1 richtig ist.
Insbesondere kann man aus einem Testergebnis, das zur Beibehaltung der Nullhypothese führt, nicht ableiten, die Nullhypothese sei nun richtig.

Keine Fische da!
aus: Beck-Bornholdt/Dubben: Der Hund, der Eier legt, Rowohlt Taschenbuch Verlag, Hamburg 1997

Wie man sich auch entscheidet, in jedem Fall kann man einen Fehler machen:

Wenn H_0 verworfen wird, obwohl sie richtig ist, spricht man von einem "α-Fehler" (Fehler 1. Art). Wenn H_1 nicht akzeptiert wird, obwohl sie richtig ist, macht man einen "β-Fehler" (Fehler 2. Art).

	Entscheidung für H_0	Entscheidung für H_1
H_0 ist richtig	alles klar	α-Fehler
H_1 ist richtig	β-Fehler	alles klar

Die Größe des α-Fehlers

Der α-Fehler ist in der Testlogik vorgegeben: Nur mit maximal α = 5 % (oder α = 1 % oder ...) will man H_0 irrtümlich verwerfen und H_1 irrtümlich akzeptieren. α kann dann tatsächlich etwas kleiner sein, da X diskret verteilt ist; d. h. dass nur natürliche Zahlen vorkommen, die i. d. R. die 5 %-Schranke nicht genau treffen.

Die Größe des β-Fehlers

a) Im Einführungsbeispiel sei die Heilungsquote des neuen Medikaments tatsächlich 75 %. Kamen im Test zufällig trotzdem nur 53 Gesundete heraus, so wird das Medikament irrtümlich nicht akzeptiert (da $53 \notin V$), obwohl es besser ist. β gibt die Wahrscheinlichkeit an, mit der dieser Irrtum geschehen kann. Er tritt immer dann auf, wenn das Testergebnis nicht in V liegt: $\beta = P_{80;\,0{,}75}(X \le 55) \approx 12{,}4$ %.
Der β-Fehler kann nur berechnet werden, wenn die Heilungsquote des neuen Medikaments tatsächlich bekannt ist. Er beträgt hier rund 1/8.
b) Im linksseitigen Test auf der Testturmseite sei p_{neu} = 45 %.
$\beta = 1 - P_{30;\,0{,}45}(X \le 16) = \quad 1 - 86{,}4$ % = 13,6 %.
c) Im Test zur Kinder-Sicherheit sei p_{neu} = 40 %. $\beta = 1 - P_{300;\,0{,}4}(X \le 135) = 1 - 96{,}6$ % = 3,4 %.
d) Im zweiseitigen Schupfentest sei p_{neu} = 25 %.
$\beta = P_{300;\,0{,}25}(84 \le X \le 116) = P(X \le 116) - P(X \le 83) = 1 - 87{,}1$ % = 12,9 %.
e) Im Yoga-Technik-Schupfentest sei p_{neu} = 25 %. $\beta = 1 - P_{100;\,0{,}25}(X \le 25) = 1 - 55{,}4$ % = 44,6 %.

3d) Aufgaben zum Hypothesentesten

Übung 1: Mischungen

Von einer Maschine werden zwei gleichartige grobkörnige chemische Substanzen gemischt. Zur Kontrolle der Mischung werden mit einem kleinen Gefäß stichprobenweise Körner entnommen. Ungewöhnliche Abweichungen vom Erwartungswert weisen auf eine schlechte Mischung hin. Die Mischung soll 20% der einen Substanz enthalten.

a) Notiere den Testturm. Bestimme V (n = 400; α = 5%).
b) Entscheide für den Fall, dass 62 Körner der 1. Substanz gefunden werden.
c) Was wäre zu 72 gefundenen Körnern der Sorte 1 zu sagen?

Übung 2: Geburtenhäufigkeit

Gibt es Monate im Jahr, in denen bei uns besonders viele oder besonders wenige Kinder geboren werden? (α = 5%)

Prüfe: a) 4/82, b) 2/76.

Tipp 1: Je mehr Tage ein Monat hat, desto mehr Geburten sind zu erwarten.

Tipp 2: Achte auf Schaltjahre! Tipp 3: Σ steht für Summe.

Anzahl der Lebendgeborenen in der Bundesrepublik Deutschland (einschl. West-Berlin)				
Monat	1982	1979	1976	1973
Januar	48150	45670	50150	52930
Februar	47870	43800	48200	50850
März	53820	47330	53230	55420
April	50270	48660	49340	53280
Mai	49230	50880	50890	55210
Juni	53550	47590	50930	54300
Juli	54310	51270	51990	56120
August	54140	50590	51660	53780
September	54780	48990	52430	50260
Oktober	51540	50470	48100	52170
November	50780	47400	47660	49120
Dezember	52940	49340	49720	52190
n =	Σ	Σ	Σ	Σ

Übung 3: Zufallszahlen

Wir würfeln mit dem Zufallszahlengenerator eines Computers und zählen das Auftreten der Augenzahl 6. Bei 1485 Würfen trat sie 234mal auf.

Ist der Zufallszahlgenerator akzeptabel (α = 5%)?

Übung 4: Linkshändigkeit

Aus umfangreichen Untersuchungen weiß man, dass 11% der 6- bis 10jährigen Mädchen manuelle Tätigkeiten eher mit der linken als mit der rechten Hand ausführen. Obwohl es heutzutage kaum noch vorkommt, dass linkshändige Kinder zur Rechtshändigkeit gezwungen werden, hat man die Vermutung, dass Kinder diese Linkshändigkeit verlernen, d.h. dass der Anteil der Linkshänder mit den Lebensjahren abnimmt. Zur Untersuchung dieser Frage will man eine Stichprobe vom Umfang 1000 unter 13jährigen Mädchen durchführen.

a) Welche Hypothese muss man testen, wenn man die Vermutung überprüfen will?
b) Bei welchen Stichprobenergebnissen würde man davon ausgehen können, dass der Anteil der Linkshänder sich verkleinert hat? (Irrtumswahrscheinlichkeit α = 5 %)

Übung 5: Jungen holen auf

Ein ungewöhnlicher Knabenüberschuss bei den 1982 in der damaligen Bundeshauptstadt Bonn geborenen Kindern beschäftigt Statistiker und Mediziner. Wie aus einer Mitteilung des Standesamtes der Stadt hervorgeht, erblickten in Bonns Entbindungskrankenhäusern im letzten Jahr 2 762 Knaben und nur 2 494 Mädchen das Licht der Welt.
Liegt hier wirklich eine signifikante Abweichung zu p = 0,514 vor (Anteil der Jungengeburten 1982 in der Bundesrepublik) für $\alpha = 5\%$?

Übung 6: Große Lottozahlen

Jemand behauptet, dass die Ziehungsmethode beim Lottospiel *6 aus 49* die größeren Zahlen bevorzuge: Die zuerst in die Lostrommel fallenden Kugeln werden häufiger gezogen als die anderen. Lässt sich diese Hypothese halten? ($\alpha = 5\%$)
Die großen Zahlen (43 bis 49) der unteren Reihe wurden in 1473 Ziehungen als *erste* Zahl einer Wochenziehung 226mal gezogen.

Übung 7: Mendel-Gesetz

Gregor Mendel untersucht bei seinen Kreuzungsversuchen mit Erbsen (um 1860), wie oft verschiedene Merkmalsausprägungen nach der Kreuzung auftraten.
Sein Hypothese lautete: Bei dominanter Vererbung haben in der zweiten Tochtergeneration 75% der Merkmalsträger die eine und 25% die andere Ausprägung. Prüfe sie mit $\alpha = 7\%$.
Die ersten Versuchsreihen lieferten folgende Ergebnisse:
a) Gestalt der Samen: Von 7324 Samen waren 5474 rund oder rundlich, 1850 kantig.
b) Färbung der Samen: von 8023 Samen waren 6022 gelb und 2001 grün.

Übung 8: Würfelmanipulation

Von einem Würfel wird vermutet, er sei manipuliert. – Daraufhin wird das Erscheinen der Augenzahl 6 in einem Versuch mit 400 Würfen getestet.
a) Gib alle Schritte des Test-Turmes an. Wie lautet die Entscheidungsregel bei einer Irrtumswahrscheinlichkeit von 5%?
b) Der Versuch liefert 80mal die Augenzahl 6.
c) Wie verläuft derselbe Test, wenn die Ausgangsvermutung lautete: die Augenzahl 6 erscheint zu häufig?
d) Erläutere den Unterschied der Teste und der Konsequenzen aus a und c.

Übung 9: Erkältung und Vitamin C

Die erste umfangreiche Studie zur Wirkung von Vitamin C stammt von dem Schweizer Arzt G. Ritzel (1961). Es war eine doppelblinde Studie an 279 französischen Skifahrern an einem Ski-Ort.
Das Ergebnis:

	Erkältet	Nicht erkältet	Summe
Vitamin C	17	122	139
Placebo	31	109	140
Summe	48	231	279

Prüfe, ob auf dem 99%-Sicherheitsniveau behauptet werden kann, Vitamin C schütze vor Erkältung.
a) Führe einen einseitigen und einen zweiseitigen Parametertest durch. Verwende als p den Vergleichswert der Placebo-Gruppe.
b) Welcher der beiden Tests in a passt zur Fragestellung?

Übung 10: Naserümpfen

9 % der BundesbürgerInnen leiden unter dem unwiderstehlichen Drang des Naserümpfens. Wohlmeinende LehrerInnen am Ri-Gri behaupten steif und fest, an unserer Schule sei diese "Volksseuche" weit weniger verbreitet.

a) Der Mathe-Kurs 13 plant einen statistischen Test: es sollen 350 SchülerInnen getestet werden. Das Ergebnis soll zu 99 % sicher sein. Es werden tatsächlich 19 naserümpfende SchülerInnen festgestellt.
b) Wie lautet a), wenn die LehrerInnen vermuten, an der Schule sei auch beim Naserümpfen alles anders.
c) Wie lautet der einseitige Test, wenn nur die Jahrgangsstufe 13 (100 SchülerInnen) befragt wird?
d) Begründe oder widerlege: Lässt sich H_0 nicht verwerfen, dann ist H_0 bestätigt.

Übung 11: Trasylol

Aus der Abteilung für Biometrie (Leiter: Prof. Dr. B. Schneider) der Medizinischen Hochschule Hannover, Hannover

Ergebnisse einer Feldstudie über den therapeutischen Wert von Aprotinin beim traumatischen Schock: **Unterschiede zwischen den verschiedenen Behandlungsgruppen**
Die Zuteilung der Patienten zu den beiden Behandlungsgruppen (Behandlung ohne Trasylol, Behandlung zusätzlich mit Trasylol) sollte in jedem Krankenhaus so vorgenommen werden, dass die an einem ungeraden Tag aufgenommenen Patienten ohne Trasylol und die Patienten, die an einem geraden Tag aufgenommen wurden, mit Trasylol behandelt werden. Bei der Auswertung zeigt sich aber, dass diese Zuteilung von den Krankenhäusern nicht immer strikt eingehalten wurde. Es wurden sowohl Patienten, die an einem geraden Tag aufgenommen wurden, ohne Trasylol als auch insbesondere Patienten mit einem ungeraden Aufnahmetag mit Trasylol behandelt. Somit sind für die Auswertung insgesamt 4 verschiedene Behandlungsgruppen zu unterscheiden:
Gruppe I: Behandlung ohne Trasylol, ungerader Aufnahmetag
Gruppe II: Behandlung ohne Trasylol, gerader Aufnahmetag
Gruppe III: Behandlung mit Trasylol, ungerader Aufnahmetag
Gruppe IV: Behandlung mit Trasylol, gerader Aufnahmetag
Die Tabelle zeigt die Patientenzahlen. Bemerkenswert ist die starke Besetzung der Gruppe III, also der Patienten, die entsprechend der Zuteilungsordnung ohne Trasylol hätten behandelt werden sollen, aber trotzdem mit Trasylol behandelt wurde. Diese große Zahl deutet darauf hin, dass ein wesentlicher Teil dieser Patienten vom behandelnden Arzt als so schwer verletzt angesehen wurde, dass er diesen Patienten die Behandlung mit Trasylol nicht vorenthalten wollte.

	Ungerader Tag			Gerader Tag		
Ohne Trasylol	I	insgesamt:	1909	II	insgesamt:	225
		verstorben:	223		verstorben:	19
mit Trasylol	III	insgesamt:	588	IV	insgesamt:	1962
		verstorben:	125		verstorben:	227

Aus: Langbein u.a. Gesunde Geschäfte(1981), S. 229

Testansatz 1: Alle Patientenzahlen mit und alle Patientenzahlen ohne Verabreichung von Trasylol werden verwendet, ohne Beachtung der richtigen oder falschen Tageswahl.
Testansatz 2: Es werden nur Patientenzahlen verwendet, die am "richtigen" Tag behandelt bzw. nicht behandelt wurden.
Testansatz 3: Der Leiter Prof. Schneider schlägt vor: Gruppe I, II und III gelten als "Nicht-Trasylol-Behandelte", Gruppe IV als "Trasylol-Behandelte" – aus welchen Gründen auch immer.

Führe die Tests auf dem 99%-Sicherheitsniveau durch. Teste jeweils den Parameter p aus der Kontrollgruppe.

Lösungen zu den Übungen

Lösung zu 1:

a) H_1: $p \neq 0{,}2$ (beidseitiger Test) – Die Maschine produziert eine schlechte Mischung.
H_0: $p = 0{,}2$ – Die Mischung der Maschine ist okay.
$\alpha = 5\%$; $n = 400$
X: Zahl der Körner aus der 1. Substanz. X ist binomialverteilt.
Gesucht ist k_1, so dass $P(X \leq k_1)$ den Wert 2,5 % letztmalig unterschreitet, und k_2, so dass $P(X \leq k_2)$ erstmals 97,5 % überschreitet.
$P(X \leq 64) \approx 2{,}4\,\%$; $P(X \leq 65) \approx 3{,}3\,\%$
$P(X \leq 95) \approx 97{,}2\,\%$; $P(X \leq 96) \approx 97{,}9\,\%$
$V = \{0,1,2,...,64,97,...,400\}$

b) Da 62 im Verwerfungsbereich V liegt, ist H_0 zu verwerfen und H_1 zu akzeptieren: Die Mischung ist nicht in Ordnung.

c) Wären 72 Körner der Substanz 1 gefunden worden, ließe sich H_0 nicht verwerfen. Eine weitere Aussage wäre <u>nicht</u> möglich: Weder wäre H_0 bestätigt (72 passt auch zu anderen p-Werten!), noch wäre H_1 als falsch gesichert. Es ist lediglich keine Aussage möglich, nicht zu H_0 und nicht zu H_1.
Konsequenz: H_0 wird nicht verworfen, H_1 nicht akzeptiert. Die Mischung wird weiterhin als normal unterstellt.

Lösung zu 2:

a) Der Test ist zweiseitig. Für April erwartet man $\frac{30}{365}$ als Anteil.
$H_1: p \neq \frac{30}{365}$ – Die Geburtenanzahl im April war ungewöhnlich.
$H_0: p = \frac{30}{365} = 8{,}22\%$ – Es gab im April 1982 normal viele Geburten.
$\alpha = 5\%$; $n = 621\,380$ (Gesamtzahl der Geburten 1982)
X: Zahl der Geburten im April 82. X ist binomialverteilt.
Gesucht k_1, so dass $P(X \leq k_1)$ letztmalig 2,5 % unterschreitet, und k_2, so dass $P(X \leq k_2)$ erstmals 97,5 % überschreitet.
$P(X \leq \quad) \approx$
$P(X \leq \quad) \approx$
$P(X \leq \quad) \approx$
$P(X \leq \quad) \approx$
Da $50270 \in V$, wird H_0 verworfen und H_1 akzeptiert: Der Anteil der Geburten im April 1982 entspricht nicht ihrem Tagesanteil im Jahr.

b) Zweiseitiger Test wie a.
$H_1: p \neq \frac{29}{366}$ – Es gab ungewöhnlich viele Geburten im Februar 1976.
$H_0: p = \frac{29}{366}$ (Februar im Schaltjahr 1976) – Die Zahl der Geburten lag im normalen Rahmen.
$\alpha = 5\%$; $n = 604\,300$ (Geburtenzahl 1976)
X: Zahl der Geburten im Februar 1976. X ist binomialverteilt.
Gesucht k1, so dass $P(X \leq k_1)$ letztmalig 2,5 % unterschreitet, und k_2, so dass $P(X \leq k_2)$ erstmals 97,5 % überschreitet.
$P(X \leq \quad) \approx \quad$; $P(X \leq \quad) \approx$
$P(X \leq \quad) \approx \quad$; $P(X \leq \quad) \approx$
V = {0, 1, 2 ... 47 469, 47 470, 48 294, 48 295, ... 604 300}
Da $48200 \notin V$, wird H_0 nicht verworfen. Es ist keine weitere Aussage zu H_0 und H_1 möglich. Die Zahl der Geburten im Februar 1976 gilt als normal.

Lösung zu 3:

(Zweiseitiger Test, da keine Tendenz vermutet ist)

H_1: $p \neq \frac{1}{6}$ – Der Zufallsgenerator des PC taugt nicht.

H_0: $p = \frac{1}{6}$ – Der Zufallsgenerator ist okay.

$\alpha = 5\%$; $n = 1485$

X: Zahl der "gewürfelten" 6en; X ist binomialverteilt.

Gesucht k_1, so dass $P(X \leq k_1)$ letztmalig 2,5 % unterschreitet, und k_2, so dass $P(X \leq k_2)$ erstmals 97,5 % überschreitet.

$P(X \leq 219) \approx 2{,}4\,\%$; $\quad P(X \leq 220) \approx 2{,}9\,\%$

$P(X \leq 275) \approx 96{,}9\,\%$ $\quad P(X \leq 276) \approx 97{,}6\,\%$

$V = \{ 0,1 \ldots 219, 277, 278, \ldots 1485\}$

Da $234 \notin V$, kann man H_0 nicht verwerfen. Weitere Aussagen zu H_0 und H_1 sind nicht im Widerspruch zu H_0. Man wird den Zufallsgenerator als akzeptabel hinnehmen.

Lösung zu 4:

H_1 : $p < 0{,}11$ (linkseitiger Test) – Die Zahl der Linkshänder nimmt ab.

H_0 : $p = 0{,}11$; $\alpha = 5\,\%$; $n = 1000$

X: Zahl der Linkshänder; X ist binomialverteilt.

Gesucht ist k, so dass $P(X \leq k)$ letztmalig 5 % unterschreitet.

$P(X \leq 93) \approx 4{,}5\,\%$; $P(X \leq 94) \approx 5{,}6\,\%$

$V = \{0, 1, \ldots, 93\}$. Wenn bis zu 93 Linkshänder auftreten, wird man wegen der auffällig kleinen Zahl eine Abnahme der Linkshändigkeit annehmen.

Lösung zu 5:

Der Test ist in Ordnung, wenn vorher (1980, 1981) die Vermutung H_1 geäußert wurde, dann 1982 als Testjahr ausgerufen wird. Problematisch ist es, aus Daten eine Vermutung zu gewinnen und diese mit genau denselben Daten zu "beweisen"!

Es handelt sich um eine einseitige Hypothese ("Knabenüberschuss"); "Jungen holen auf", "Kluft zugunsten der männl. Babys").

H_1: $p > 0{,}514$ – Die Geburtenzahl der Jungen war 1982 in Bonn ungewöhnlich hoch.

H_0: $p = 0{,}514$ – Die Zahl bewegte sich im normalen Rahmen.

$\alpha = 5\%$; $n = 5256$ (Zahl der Geburten 1982)

X: Zahl der Jungengeburten; X ist binomialverteilt.

Gesucht ist k, so dass $P(X \leq k)$ erstmalig 95 % überschreitet.

$P(X \leq 2760) \approx 94{,}8\,\%$; $P(X \leq 2761) \approx 95{,}1\,\%$

$V = \{2762, \ldots 5256\}$

Da $2762 \in K$, ist H_0 zu verwerfen und H_1 zu akzeptieren: In Bonn liegt der Anteil der Knabengeburten 1982 höher als im Bundesgebiet ($\alpha = 5\%$).

Lösung zu 6:

Rechtsseitiger Test, da eine Bevorzugung behauptet wird.

Zu p: Die 7 Zahlen stellen $\frac{1}{7}$ der 49 Zahlen dar.

H_1 : $p > \frac{1}{7}$ – Die 7 großen Zahlen werden beim Lotto bevorzugt gezogen.

H_0 : $p = \frac{1}{7}$ – Sie werden ganz normal gezogen.

$\alpha = 5\%$; $n = 1473$

X: Zahl der Ziehungen (als erste Zahl) einer Zahl aus $\{43, 44, \ldots, 49\}$. X ist binomialverteilt.

Gesucht ist k, so dass P (X $\leq$ k) erstmalig 95 % überschreitet.
P (X $\leq$ 232) $\approx$ 94,8 %; P (X $\leq$ 233) $\approx$ 95,6 %
V = {234, 235 .... 1473}
Da 226 $\notin$K, wird H_0 nicht verworfen und H_1 nicht akzeptiert. Die größeren Zahlen werden beim Lotto nicht ungewöhnlich häufig gezogen.

Lösung zu 7:

Zweiseitiger Test, da keine Tendenz nach oben oder unten vermutbar ist.

a) H_1: p $\neq$ 0,75 – Die $\frac{3}{4}$ / $\frac{1}{4}$-Regel stimmt nicht.
   H_0: p = 0,75 – Die Versuchsreihen passen zu der Regel.
   α = 7% (ungewöhnlich, aber mal so zur Übung ...); n = 7324.
   X: Zahl der Samen, die rundlich sind; X ist binomialverteilt.
   Gesucht ist k_1, so dass P (X $\leq$ k_1) letztmalig 3,5 % unterschreitet, und k_2, so dass P (X $\leq$ k_2) erstmalig 96,5 % überschreitet.
   P (X $\leq$ 5425) $\approx$ 3,47 %; P (X $\leq$ 5426) $\approx$ 3,68 %
   P (X $\leq$ 5559) $\approx$ 96,40 %; P (X $\leq$ 5560) $\approx$ 96,61 %
   V = {0,1, ... 5425, 5561, 5562, ...5493}
   Da 5474 $\notin$ V, widerspricht das Testergebnis nicht der Hypothese H_0. Die Ergebnisse sind mit der $\frac{3}{4}$ / $\frac{1}{4}$- Regel verträglich.

b) n = 8023; sonst s. a.
   P (X $\leq$ 5946) $\approx$ 3,44 %;   P (X $\leq$ 5947) $\approx$ 3,64 %
   P (X $\leq$ 6086) $\approx$ 96,33 %; P (X $\leq$ 6087) $\approx$ 96,53 %
   V = {0,1, ... 5946, 6088, 6089, ... 8023}
   Da 6022 $\notin$ V, wird H_0 nicht verworfen und H_1 nicht akzeptiert. Das Ergebnis widerspricht nicht der $\frac{3}{4}$ / $\frac{1}{4}$- Regel.

Lösung zu 8:

a) Zweiseitiger Test
   H_1 : p $\neq$ $\frac{1}{6}$ – Der Würfel ist manipuliert.
   H_0 : p = $\frac{1}{6}$ – Der Würfel ist okay.
   α = 5 %; n = 400
   X: Zahl der 6en im Test; X ist binomialverteilt.
   Gesucht ist k_1, so dass P (X $\leq$ k_1) letztmalig unter 2,5 % liegt, und k_2, so dass P (X $\leq$ k_2) erstmalig über 97,5 % liegt.
   P (X $\leq$ 51) $\approx$ 1,8 %; P (X $\leq$ 52) $\approx$ 2,6 %
   P (X $\leq$ 81) $\approx$ 97,4 %; P (X $\leq$ 82) $\approx$ 98,1 %
   V = {0,1, ... 51, 83, 84, ... 400}

b) Da 80 $\notin$ V, lässt sich H_0 nicht verwerfen, H_1 nicht akzeptieren. Der Würfel ist okay.

c) Einseitiger Test
   H_1 : p > $\frac{1}{6}$ – Der Würfel ist so manipuliert, dass er zu viele 6en liefert.
   H_0, α, n, X wie oben
   P (X $\leq$ 78) $\approx$ 94,2 %
   P (X $\leq$ 79) $\approx$ 95,5 %
   V = {80, 81, ..., 400}
   Da 80 $\in$ V, wird H_0 verworfen zugunsten von H_1: Der Würfel liefert zu oft die 6.

d) Es liegt ein zweiseitiger bzw. einseitiger Test vor. In a) werden 2 Prüfungen (p > $\frac{1}{6}$ und p < $\frac{1}{6}$) mit jeweils der Irrtumswahrscheinlichkeit $\frac{\alpha}{2}$ = 2,5 % gemacht. In c) wird nur p > $\frac{1}{6}$ mit α = 5 % getestet. Entsprechend sind die rechtseitigen Verwerfungsbereiche verschieden groß. Einmal liegt 80 im Verwerfungsbereich, einmal nicht – mit entsprechender Wertung.

Lösung zu 9:

a) X: Zahl der Nicht-Erkälteten in der Vitamin-C-Gruppe; n = 139; α = 1 %
Aus der Placebo-Gruppe: p = 109/140 ≈ 77,86 %.
P (X ≤ 94) ≈ 0,35 %; P (X ≤ 95) ≈ 0,6 %
P (X ≤ 118) ≈ 98,5 %; P (X ≤ 119) ≈ 99,2 %; P (X ≤ 120) ≈ 99,5 %
H_1: p ≠ 77,86 % (zweiseitig) führt auf V = {0,1, ..., 94, 121, 122, ..., 139}
H_1: p ≠ 77,86 % (einseitig) führt auf V = {120, 121, ..., 139}
In beiden Fällen ist H_0 (Vitamin C beeinflusst die Erkältung nicht; p = 77,86 %) zu verwerfen zugunsten von H_1 (Vitamin C erhöht die Zahl der Nicht-Erkälteten;

b) p > 77,86 % bzw. p ≠ 77,86 %).
Es steht schon vor dem Test fest, dass geprüft werden soll, ob Vitamin C die Zahl der Nicht-Erkälteten erhöht. Alles andere wäre gegen die medizinische Ethik! Der einseitige Test passt.
((a und b lassen sich analog für die Zahl der Erkälteten formulieren.))

Lösung zu 10:

H_0: Am Ri-Gri wie überall beträgt der Anteil der Naserümpfer p = 9 %.
α = 1 %; n = 350
X: Zahl der Naserümpfer unter den Befragten; X ist binomialverteilt
P (X ≤ 18) ≈ 0,49 %; P (X ≤ 19) ≈ 0,9 %; P (X ≤ 20) ≈ 1,6 %
P (X ≤ 45) ≈ 99,4 %; P (X ≤ 46) ≈ 99,6 %
(linksseitiger Test) führt auf

a) H_1: am Ri-Gri gibt es weniger Naserümpfer als üblich; p < 9 % (linksseitiger Test) führt auf den Verwerfungsbereich = {0, 1, 2, ..., 19}
Da 19 ∈ K, wird H_0 verworfen und H_1 akzeptiert: Am Ri-Gri gibt es weniger Naserümpfer als üblich (Irrtumswahrscheinlichkeit 1 %).

b) H_1: Am Ri-Gri ist alles anders; p ≠ 9 % (zweiseitiger Test)
führt auf den Verwerfungsbereich V = {0, 1, 2, ..., 18, 47, ..., 350}.
Da das Testergebnis 19 nicht im Verwerfungsbereich liegt, lässt sich H_0 nicht verwerfen und über H_0 und H_1 sind keine weiteren Aussagen möglich. Insbesondere ist - anders als in a - H_1 nicht bestätigt. Das Nasenrümpfen ist am Ri-Gri wie sonst auch üblich.

c) Jahrgangsstufe 13, n = 100; sonst wie a, b.
P (X ≤ 2) ≈ 0,5 %; P (X ≤ 3) ≈ 1,7 %
V = {0, 1, 2}
Nur wenn 2 oder weniger Naserümpfer in der Jahrgangsstufe 13 gefunden werden, ist Nasenrümpfernormalität für das Ri-Gri zu verwerfen zu Gunsten geringerer Verbreitung am Ri-Gri.

d) Die Aussage ist falsch.
Richtig ist eine Aussage wie in der Lösung zu b).
Der Satz kann so nicht stimmen, da ein Testergebnis <u>nie</u> ein bestimmtes p bestätigen kann. Das Testergebnis passt immer zu vielen Werten für p (innerhalb des Sicherheitsniveaus). Es wird nur geprüft, ob ein außergewöhnliches Ergebnis vorliegt unter der Annahme, p gelte. Ist das Ergebnis nicht außergewöhnlich, so ist das kein Argument gegen p, aber sowieso keines für p.

Lösung zu 11:

1. q sei die Sterblichkeitsquote: $q_{mit} = 352/2550 \approx 13,8$ %; $q_{ohne} = 242/2134 \approx 11,3$ % disqualifiziert Trasylol als patientengefährdend.
2. $q_{mit} = 227/1962 \approx 11,6$ %; $q_{ohne} = 223/1909 \approx 11,7$ %. Ein kaum merkbarer Vorteil für Trasylol; nicht testenswert.
3. X: Zahl der Gestorbenen, die Trasylol erhielten (Gruppe IV); X ist binomialverteilt.
   H_0: $p_{ohne} = 367/2722 \approx 13,48$ % – Trasylol ist ohne Wirkung.
   H_1: $p < 13,5$ % – Trasylol senkt die Sterblichkeit (linksseitig!).
   $n = 1962$; $\alpha = 1$ %
   $P(X \leq 229) \approx 0,9$ %; $P(X \leq 230) \approx 1,1$ %
   Verwerfungsbereich für H_0: $V = \{0, 1, 2, ..., 229\}$.
   Da 227 im Verwerfungsbereich liegt, ist H_0 zu verwerfen und H_1 zu akzeptieren. Trasylol hilft ($\alpha = 1$ %), sofern die Datenmanipulationen akzeptiert werden.

Viele weitere interessante, witzige, relevante, ansprechende Beispiele zu Wahrscheinlichkeiten, Deutungen und Binomialverteilungen siehe in:
A. Warmeling (Hg.) PROST – Problemorientierte Stochastik, Sammlung 2, Von Verteilungen, vom Schätzen und vom Testen im Alltag, Appelhülsen, 2002; zu beziehen über: MUED e. V, Bahnhofstraße 72, 48301 Appelhülsen, Tel. 02509/606; Fax. 02509/996516, eMail: mued.ev@t-online.de , Internet: www.mued.de

3e) Medikamenten- und Placebowirkungen

Warum werden Menschen in Stresssituationen krankheitsanfällig?

Zusammenspiel von Gemütszustand und Immunsystem soll erforscht werden

Zusammenhänge zwischen psychischem, nervlichem und körperlichem Befinden erahnt jeder Mensch, aber erklärt und bewiesen sind sie bisher nicht. In diesen Tagen trafen sich in Hannover Experten vieler verschiedener Wissenschaftszweige aus West- und Osteuropa, den USA und Ostasien, um sich darüber auszutauschen, wie die hormonellen Wechselwirkungen zwischen Gemütsverfassung und Immunsystem näher erforscht werden können.

Der Bayreuther Zoologe Dietrich von Holst berichtete über Tierexperimente, die gezeigt haben, wie massiv Stress, Schockerlebnisse, Angst oder aufgebaute Aggressivität das Immunsystem beeinflussen und zu schnellem Tod führen können. Dies gelte offenkundig auch für Menschen. Holst verwies auf eine schwedische Untersuchung, die nach einer Betriebsschließung die verminderte Abwehrkraft der Entlassenen nachgewiesen habe. Er stellte die These auf, emotionale Verarbeitung sei entscheidend für die Widerstandsfähigkeit des Körpers.

An Beispielen für Situationen, in denen Menschen krankheitsanfällig werden, fehlte es nicht: Examensstress, chronischer Ehestreit oder Partnerverlust.

Der Onkologe Hans-Joachim Schmoll erwähnte die Tatsache, dass das "Besprechen" von Warzen in vielen Fällen Warzen verschwinden lasse. Hypnose aktiviere die Lymphozyten, die eine Hauptrolle im Immunsystem spielen. Die Placebo-Forschung habe nachgewiesen, dass die Wirkung von Medikamenten bis zu 70 Prozent auf dem bloßen Glauben des Patienten an deren Wirkung beruhe.

Als Forschungsziel bezeichnete es der Immunologe Reinhold-Ernst Schmidt, die Lernfähigkeit des Immunsystems nutzbar zu machen, die mindestens ebenso groß sei wie die des Nervensystems.

nach: Frankfurter Rundschau, 18.09.1989

Konsequenzen für die Durchführung eines Medikamententests:

a) 1961 hat der Schweizer Arzt G. Ritzel 279 französische Ski-Fahrer an einem Ski-Ort in zwei Gruppen eingeteilt. Die erste Gruppe diente als *Versuchsgruppe V*, die andere als *Kontrollgruppe K* (*kontrolliertes Experiment*).

b) Die Einteilung erfolgte mit einem Zufallsgerät (*randomisiertes kontrolliertes Experiment*).

c) Jede Person aus V erhielt täglich eine Vitamin-C-Tablette. Da eine Person nicht wissen durfte, ob sie zu V oder K gehörte, erhielt jede Person aus K eine Placebo-Tablette, die wirkungslos war, aber in Aussehen und Geschmack von einer Vitamin-C-Tablette nicht unterscheidbar war. Dadurch wurde sichergestellt, dass die beobachtete Wirkung nicht durch Einbildung zustande kam. Der Glaube kann heilen! (*Blindes randomisiertes kontrolliertes Experiment.*)

d) Auch der Arzt, der die Wirkung der Behandlung feststellt, wusste nicht, ob eine Versuchsperson zu V oder K gehörte. Dadurch wurde die unbewusste Beeinflussung der Diagnose ausgeschaltet (*doppelblindes randomisiertes kontrolliertes Experiment*).

Ein medizinische Experiment mit Menschen muss unter Beachtung von a) bis d) ausgewertet werden. Wird dies nicht beachtet, so betrügt man sich selbst, und man beträgt andere.

(Engel, Stochastik, 1987; siehe auch Übung 9)

I) Grippe-Impfung:
Jeden Herbst wird allen Bediensteten der Universität Frankfurt eine freiwillige Grippeimpfung angeboten. Von 25 Mitgliedern eines Instituts ließen sich 20 impfen, und 6 der Geimpften sind im Winter an Grippe erkrankt. Von den übrigen 5 sind 3 an Grippe erkrankt. Ist die Wirksamkeit der Impfung auf dem 5 %-Niveau gesichert?
Sind die experimentellen Bedingungen einwandfrei?

II) Was war beim Trasylol-Test (Übung 11) nicht in Ordnung?

Zu Übung I

Alle 4 Forderungen sind nicht erfüllt. Es gibt keine Kontrollgruppe. Die Probanden sind nicht zufällig ausgewählt. Die Patienten sind informiert und der Arzt auch.

Zu Übung II

Das Experiment war randomisiert und kontrolliert geplant. Da die Patienten bewusstlos sind, gibt es keinen Placebo-Effekt (blindes Experiment). Da aber der behandelnde Arzt Bescheid wusste (kein doppelblindes Experiment), wurde die randomisierte Einteilung in Versuchs- und Kontrollgruppe zunichte gemacht.
Die nachträgliche Zusammenfügung von Gruppen durch Dr. Schneider ist unzulässig; der Versuch insgesamt unbrauchbar.