ST. PÖLTEN UNIVERSITY
OF APPLIED SCIENCES

Informatik
& Security

fh
st.pölten

Hypothesentests

Dipl.-Ing. Hubert Schölnast, BSc
Stand: 05. Juli 2022

Inhaltsverzeichnis

1 Hypothesentest

1.1 Hypothese und Theorie

In der Wissenschaft ist eine Hypothese eine wohlbegründete, aber noch unbewiesene Annahme über die genaue Art des Zusammenhangs zwischen einer Ursache und der sich daraus ergebenden Wirkung. Sehr oft gibt es mehrere einander widersprechende Hypothesen, die versuchen, denselben Zusammenhang zu erklären. Aus diesen Hypothesen kann man bestenfalls einige als falsch entlarven, es ist aber prinzipiell unmöglich, zu beweisen, dass eine Hypothese den Zusammenhang korrekt erklärt. Das Widerlegen falsche Hypothesen kann gelingen, wenn eine Hypothese vorhersagen macht, die über den ursprünglichen Zusammenhang hinausgehen. Wenn sich diese Vorhersagen als falsch erweisen, ist die Hypothese zu verwerfen. Wenn nicht nachgewiesen werden kann, dass die Vorhersagen falsch sind, bleibt die Hypothese als möglicher Kandidat für die korrekte Beschreibung der Realität im Rennen.

Erst wenn viele Hypothesen formuliert, getestet und verworfen worden sind, kann man davon ausgehen, dass jene Hypothesen, die allen Bemühungen, sie zu widerlegen standgehalten haben, dafür geeignet sind die Wahrheit zu beschreiben. Erst wenn die Gemeinschaft der Wissenschaftler*innen gemeinsam zu dem Schluss kommt, dass eine bestimmte Hypothese diesen Grad an Verlässlichkeit aufweist, erst dann nennt man die betreffende Hypothese eine Theorie.

1.1.1 Beispiel Gravitation:

Dass die meisten Dinge nach unter fallen wenn man sie loslässt, ist eine Erkenntnis, die seit jeher den Menschen so vertraut ist, dass man es in jenen Zeiten, in denen man versuchte, die Welt durch das Wirken von Göttern zu beschrieben, nicht für notwendig hielt, sie durch ein solches göttliches Wirken zu begründen. Weder bei den Ägyptern, noch bei den Mesopotamiern, nicht bei den German, nicht bei den Azteken und auch nicht bei den Ureinwohnern Australiens gibt es einen Gott der Gravitation. Auch in der Bibel wird die Kraft, die Dinge nach unten fallen lässt, nicht erklärt.

Aber die Menschen dachten trotzdem über die Ursache dieser allgegenwärtigen nach unten gerichteten Kraft nach, und kamen zu der Auffassung, dass "unten" der natürliche Ort aller Dinge ist, und dass alle Dinge daher nach unten streben wenn man ihnen die Möglichkeit dazu gibt. Bei Wolken und Vögeln war das anders, weil deren natürlicher Ort "oben" war.

Sogar der Astronom Nikolaus Kopernikus schrieb im Jahr 1543 über die Schwerkraft: »Ich bin wenigstens der Ansicht, dass die Schwere nichts Anderes ist, als ein von der göttlichen Vorsehung des Weltenmeisters den Theilen eingepflanztes, natürliches Streben, vermöge dessen sie dadurch, dass sie sich zur Form einer Kugel zusammenschließen, ihre Einheit und

Ganzheit bilden. Und es ist anzunehmen, dass diese Neigung auch der Sonne, dem Monde und den übrigen Planeten innewohnt.«

Johannes Keppler (um 1600) und Galileo Galilei (um 1640) teilten im Wesentlichen diese Meinung, steuerten aber auch Ideen bei, die dazu betrugen das Ausmaß dieser Kraft abzuschätzen.

Erst im Jahr 1687 begann Isaac Newton von *Massen* zu sprechen die aufeinander eine anziehende Kraft ausüben und er war auch der erste, der die Gravitation mit einer exakten Formel beschrieb. Diese *Hypothese* Newtons wurde zur Newtonschen Gravitations*theorie* und löste die älteren Gravitationshypothesen ab, und zwar in dem Sinn, als dass durch Newtons Theorie die alten Hypothesen als ungenaue Näherungen der neuen Theorie anzusehen waren.

Abgelöst wurde diese Theorie erst 1916 durch Einsteins Allgemeine Relativitätstheorie, welche die Gravitation als eine Wirkung der Verformung des Raumes beschreibt, wobei diese Verformung wiederum eine Folge der Anwesenheit von Masse ist. Die Newtonsche Gravitation erscheint nun selbst als ungenaue Näherung der Einstein'schen Gravitation.

Die Allgemeine Relativitätstheorie und die Quantenmechanik sind heute die Basistheorien der gesamten Physik, und alle Maschinen, die wir heute bauen, beruhen darauf, dass wir diesen beiden Theorien vertrauen können. Aber es ist schon seit fast hundert Jahren bekannt, dass sie einander in bestimmten Bereichen eklatant widersprechen (nämlich bei der Beschreibung schwarzer Löcher und bei der Beschreibung der Zustände während und kurz nach dem Urknall). Mindestens eine der beiden "Theorien" (sehr wahrscheinlich aber beide) sind daher selbst auch nur ungenaue Näherungen einer noch nicht verfügbaren besseren Theorie.

1.2 Nullhypothese und Alternativhypothese

Eine Möglichkeit, um eine Hypothese zu widerlegen, stellt die Statistik in Form von Hypothesentests bereit. Dabei wird eine Hypothese in Form einer Gleichung oder Ungleichung formuliert. Eine andere Hypothese beschreibt genau das Gegenteil. Jene der beiden Hypothesen, die entweder die ursprüngliche ist, oder der man im Zweifel lieber vertraut, bekommt die Rolle der Nullhypothese (H_0) zugewiesen. Ihre Verneinung ist die Alternativhypothese (H_1). Welche der beiden Hypothesen man als Nullhypothese und welche man als Alternativhypothese betrachten will, hängt von der Aufgabenstellung ab. Sehrt oft ist die Aufteilung folgende:

- Nullhypothese: Es bleibt alles wie es ist, es gibt keine Veränderung oder Abweichung
- Alternativhypothese: Es gibt eine Veränderung bzw. Abweichung

1.2.1 Testfunktion

Eine Testfunktion ist eine Stichprobenfunktion, die dazu verwendet wird, eine Entscheidung zwischen zwei Hypothesen zu treffen.

1.2.2 Beibehalten, verwerfen und annehmen

Man sollte sich aber vor der Durchführung eines statistischen Hypothesentests darüber im Klaren sein, dass das Ergebnis des Hypothesentests eines dieser beiden Ereignisse ist:

1. Die Nullhypothese erweist sich mit hoher Wahrscheinlichkeit als falsch, daher sieht man sich gezwungen, davon auszugehen, dass die Alternativhypothese zutrifft.
   H_0 wird *abgelehnt* bzw. verworfen; H_1 wird *angenommen*
2. Die Hinweise, die gegen die Nullhypothese sprechen, reichen nicht aus um sie zu verwerfen. Sie könnte gültig sein.
   H_0 wird *beibehalten*; H_1 wird *nicht angenommen*

- H_0: beibehalten oder ablehnen
- H_1: annehmen oder nicht annehmen

Man geht also vor dem Test immer davon aus, dass die Nullhypothese stimmt, und weicht von dieser Annahme nur dann ab, wenn schwerwiegende Gründe dafür vorliegen.

Beachte, dass es kein Szenario gibt, in dem die Nullhypothese bestätigt wird. Entweder man verwirft sie, weil die Daten nicht zu ihr passen, oder man behält sie bei, weil man keine ausreichenden Gegenargumente hat.

Beachte auch, dass die Alternativhypothese im Fall 1 "notgedrungen" angenommen wird (man würde ja eigentlich lieber bei der Nullhypothese bleiben), dass es aber im Fall 2 keine konkrete Aussage über die Alternativhypothese gibt. Sie gilt im Fall 2 keineswegs als ausgeschlossen. Es gibt nur nicht genügend Argumente um sie auszuschließen.

1.3 Ein- und zweiseitige Tests

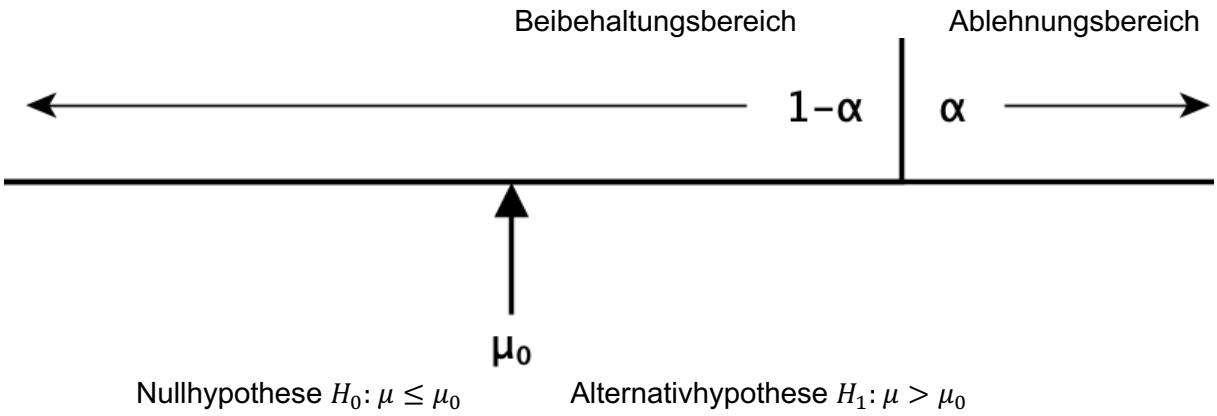
1.3.1 Einseitige Tests (Bereichstests)

Beim einseitigen Test behauptet die Nullhypothese, dass eine bestimmte statistische Kennzahl (meist ein Lagemaß, z.B. das arithmetische Mittel, manchmal aber auch die Varianz oder ein anderes Maß), größer oder kleiner als ein bestimmter Grenzwert ist.

Beispiele:

- Das mittlere Gewicht der Kartoffelsäcke, die in einem Supermarkt zum Verkauf angeboten werden, beträgt mindestens 2,0 kg.
- Der mittlere Anteil fehlerhafter CPUs auf den Silizium-Wafers eines bestimmten Herstellers ist kleiner als 10%.

Die Alternativhypothese behauptet dann, dass die Säcke im Schnitt leichter als 2 kg sind, bzw. dass der Anteil im Schnitt größer als 10% ist. Beide Hypothesen definieren also Bereiche (Intervalle).

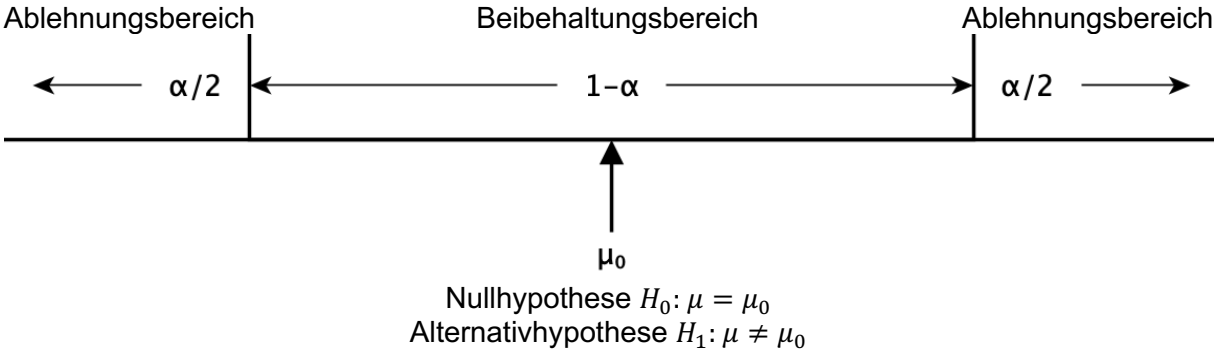


1.3.2 **Zweiseitige Tests (Punkttests)**

Beim zweiseitigen Test behauptet die Nullhypothese, dass die Kennzahl einen genauen Wert hat. Die Alternativhypothese behauptet dann, dass die Kennzahl entweder kleiner oder größer ist.

Beispiele:

- Männliche Österreicher im Alter zwischen 18 und 25 Jahren sind im Schnitt 178 cm groß[1]
- Das Pro-Kopf-Bruttonationaleinkommen Österreichs beträgt 44.145 Euro pro Jahr[2]



Die Alternativhypothese behauptet dann, dass der wahre Wert eine beliebige andere Zahl ist (größer oder kleiner). Die Nullhypothese definiert also einen Punkt, keinen Bereich. Trotzdem gibt es einen Bereich (den Beibehaltungsbereich), innerhalb dessen die Nullhypothese beibehalten wird.

[1] https://www.laenderdaten.info/durchschnittliche-koerpergroessen.php

[2] https://www.laenderdaten.info/durchschnittseinkommen.php

1.4 Signifikanzniveau

Bei einem Hypothesentest definiert man **vor dem Test** ein bestimmtes Vertrauensintervall. Kommt die Schätzgröße der Stichprobe in diesem Intervall zu liegen, vertraut man darauf, dass die Differenz zwischen der Schätzgröße μ und dem wahren Wert der Grundgesamtheit μ_0 auf erwartbare zufällige Schwankungen bei der Stichprobenziehung zurückzuführen ist. Kommt μ außerhalb des Vertrauensintervalls zu liegen, erscheint die Annahme, dass die Abweichung zwischen μ und μ_0 rein zufällig ist, nicht mehr glaubwürdig.

Die Breite dieses Vertrauensniveaus hängt von der Varianz der Mittelwerte vieler Stichproben ab und entspricht genau dem Konfidenzintervall von Intervallschätzungen und wird auch genau wie dort beschrieben berechnet (siehe Skriptum Schätzverfahren, Abschnitt 3).

Was bei der Intervallschätzung noch »Irrtumswahrscheinlichkeit« geheißen hat, heißt jetzt »Signifikanzniveau«. Sie wird in beiden Fällen mit dem Buchstaben α bezeichnet.

Bei Hypothesentests gibt das Signifikanzniveau an, wie groß die Wahrscheinlichkeit dafür ist, die Nullhypothese irrtümlicherweise abzulehnen, obwohl sie in Wahrheit zutrifft.

1.4.1 α-Fehler (Fehler 1. Art)

Der eben beschriebene Fehler (H_0 trifft in Wahrheit zu, wird aber irrtümlich abgelehnt) heißt »α-Fehler« oder »Fehler 1. Art«. Manchmal, vor allem in einem medizinischen Kontext, bezeichnet man Ereignisse, die diesem Fehler entsprechen, auch als »falsch positiv«. Als »positiv« gilt dabei ein Testergebnis, das eine Abweichung von der Norm (also meist eine Erkrankung) anzeigt.

Den Fehler 1. Art kann man in einem Hypothesentest kontrollieren. Man definiert nämlich den Wert von α **vor** dem Test fest und legt somit eine Obergrenze für das Ausmaß dieses Fehlers fest.

1.4.2 β-Fehler (Fehler 2. Art)

Natürlich kann es auch passieren, dass die Nullhypothese in Wahrheit gar nicht zutrifft, und man sie trotzdem beibehält weil die Daten der Stichprobe für eine Ablehnung nicht ausreichen. Dann liegt ein »Fehler 2. Art« vor. Das Ausmaß dieses Fehlers gibt man mit der Variablen β an, daher heißt dieser Fehler auch »β-Fehler«. Die Bezeichnung »falsch negativ« kommt daher, dass ein Test keine Normabweichung anzeigt, obwohl in Wahrheit eine vorliegt.

Ohne Annahmen über den wahren Wert der Grundgesamtheit ist es nicht möglich, den genauen Wert von β zu berechnen. Ironischerweise ist der Wert von β nämlich umso größer, je näher der wahre Wert und die Schätzung beieinander liegen.

		In Wahrheit	
		H_0 trifft zu	H_0 trifft nicht zu
Ergebnis des Hypothesentests	H_0 wird beibehalten	richtig negativ P = 1-α	falsch negativ β-Fehler P = β
	H_0 wird abgelehnt	falsch positiv α-Fehler P = α	richtig positiv P = 1-β

Fehler 1. und 2. Art

1.5 Gütefunktion

Die Güte eines Tests ist die Wahrscheinlichkeit dafür, dass die Nullhypothese vom Test abgelehnt wird wenn die Nullhypothese tatsächlich nicht zutrifft. (Wahrscheinlichkeit für »richtig positiv« unter der Voraussetzung, dass H_0 nicht zutrifft). Die Güte ist also der Wert $1 - \beta$ in der obigen Darstellung.

Das Problem dabei ist, dass die Alternativhypothese üblicherweise in folgender Form formuliert wird: »Der wahre Wert der Grundgesamtheit hat **irgendeinen** anderen Wert als μ_0.«

Um β (und damit $1 - \beta$) ausrechnen zu können, bräuchten wir also eine spezifische Alternativhypothese folgender Form: »Der wahre Wert der Grundgesamtheit hat **genau** den Wert μ_1.« (Wobei natürlich gilt: $\mu_0 \neq \mu_1$).

(Mehr über die Gütefunktion steht weiter unten, in 2.1.1, wo die Gütefunktion bei einem konkreten Beispiel besprochen wird.)

1.6 p-Wert

Der p-Wert ist die Wahrscheinlichkeit dafür, dass die Stichprobe, die man erhalten hat, beim Zutreffen der Nullhypothese durch reinen Zufall zustande gekommen ist.

Andere, aber gleichwertige Interpretation: Der p-Wert ist das kleinste hypothetische Signifikanzniveau bei dem H_0 bereits zu verwerfen wäre. (Der p-Wert markiert genau die Grenze zwischen beibehalten und verwerfen.)

Viele Testverfahren sind so konstruiert, dass man für sie gar kein Signifikanzniveau angeben muss. Sie liefern daher auch nicht die Entscheidung zwischen »H_0 beibehalten« und »H_0 ablehnen«, sondern sie geben den p-Wert p aus, den man **nach** dem Test mit dem Signifikanzniveau α vergleichen muss, für das man sich **vor** dem Test entschieden hat.

- $p \leq \alpha$: H_0 ablehnen, H_1 annehmen
- $p > \alpha$: H_0 beibehalten, H_1 nicht annehmen

2 Testverfahren

2.1 Binomialtest

Sie nehmen einem Trickbetrüger eine Münze ab, von der Sie glauben, dass die beiden Werte Kopf und Zahl nicht gleich häufig erscheinen wenn die Münze geworfen wird. Der Eigentümer der Münze behauptet, die Münze wäre fair, würde also beide Seiten gleich häufig anzeigen.

Das wiederholte Werfen der Münze ist ein Bernoulli-Experiment mit dem Parameter $p = 0{,}5$ (das ist die Wahrscheinlichkeit dafür, dass *Kopf* geworfen wird). Der Erwartungswert nach n Würfen ist $n \cdot p$ (Siehe Skriptum »diskrete Verteilungen«)

Bei einem einseitigen Binomialtest lautet die Nullhypothese, dass nach n Würfen höchstens (oder mindestens) $n \cdot p$-mal *Kopf* erscheint.

Beim zweiseitigen Binomialtest lautet die Nullhypothese, dass genau $n \cdot p$-mal *Kopf* erscheint.

Die Alternativhypothese behauptet jeweils das Gegenteil.

Wir führen im Folgenden einen zweiseitigen Binomialtest durch.

2.1.1 Güte, Gütefunktion

Nehmen wir an, der Falschspieler hätte 2 Münzen. Eine Münze ist so beschaffen, dass sie immer so landet, dass *Kopf* erschein. Die Kopf-Wahrscheinlichkeit w hat also den Wert $1{,}0$. Der Erwartungswert μ_1 dieser Münze ist also n (Wirft man die Münze n-Mal, erscheint n-Mal Kopf). Die zweite Münze zeigt Kopf nur geringfügig öfter als Zahl. Nehmen wir an, dass bei dieser Münze die Wahrscheinlichkeit für Kopf bei $w = 0{,}51$ liegt. Dann ist $\mu_2 = 0{,}51 \cdot n$ Das heißt, dass bei 100 Würfen der Erwartungswert 51 ist, bei 1000 Würfen 510.

Es ist offensichtlich, dass die völlig einseitige Münze leichter zu enttarnen ist als die mit der kleinen Tendenz zu einer Seite. Bei einer Stichprobe von 20 Würfen wird man die einseitige Münze klar erkennen. Der Test mit dieser Stichprobengröße, kann also klar zwischen »faire Münze« und »völlig einseitige Münze« unterscheiden. Das ist so, weil die Wahrscheinlichkeit, bei einer fairen Münze 20-Mal Kopf zu erhalten, bei ungefähr einem Millionstel liegt. Diese Wahrscheinlichkeit ist der p-Wert.

Die Güte eines Tests ist $1 - p$. Das ist die Wahrscheinlichkeit dafür, bei der gewählten Stichprobengröße die gezinkte Münze zu erkennen. Sie ist bei einer Stichprobe der Größe 20 nahe bei 1.

Bei der fast-fairen Münze werden wir keine Chance haben, sie mit nur 20 Würfen von einer tatsächlich fairen Münze zu unterscheiden. Dazu ist die Güte des Tests nicht ausreichend. Der Gütewert für eine Münze mit $w = 0{,}51$ ist nahe bei 0.

Die Gütefunktion gibt an, wie die Güte eines Test mit einer vorgegebenen Stichprobengröße von der Wahrscheinlichkeit der Münze, Kopf zu zeigen, abhängt.

2.2 t-Test

Es gibt drei Arten des t-Tests:

2.2.1 Ein-Stichproben-t-Test

Sie erwarten, dass eine Stichprobe einen bestimmten fest vorgegebenen Mittelwert hat, wissen aber nichts über die Varianz. Der tatsächliche Mittelwert der Stichprobe liegt aber knapp neben dem erwarteten Wert. Da diese Mittelwerte gemäß einer t-Verteilung verteilt sind, kann man einen t-Test verwenden, um einen p-Wert zu berechnen. Liegt der p-Wert unter einem vorgegebenen Signifikanzniveau, ist die Nullhypothese »Der Mittelwert der Stichprobe ist eine zufällige Abweichung vom theoretisch erwarteten Wert« zu verwerfen.

2.2.2 Zwei-Stichproben-t-Test

An die Stelle des fest vorgegebenen zu erwartenden Mittelwerts tritt der Mittelwert einer zweiten Verteilung.

Beispiel: Eine Portionierungsmaschine erzeugt vor der Reparatur Portionen mit einem Mittelwert von 352 g. Nach der Reparatur werden wieder Stichproben gezogen, der Mittelwert dieser Stichproben beträgt 346 g. Stellt die Maschine nach wie vor ungefähr gleich große Portionen her? Auch diese Frage kann mit einem t-Test beantwortet werden. Dazu dürfen die beiden Stichproben auch unterschiedlich groß sein.

2.2.3 Paarweiser t-Test

Anstatt die Mittwerte zweier Stichproben zu vergleichen, schaut man sich den Mittelwert der Differenzen zweier abhängiger Stichproben an. Was damit gemeint ist, soll ein Beispiel zeigen:

Bei allen Personen, die an einem Medikamententest teilnehmen, wird vor und nach Verabreichung des Medikaments der Blutdruck gemessen. Es gibt also von jeder Person 1 Messungen. Die Messungen vor der Medikamentengabe bilden eine Stichprobe, die Messungen danach bilden die zweite Stichprobe. Beide Stichproben sind exakt gleich groß. Dann bildet man von jeder Person die Differenz der beiden Werte. Die Null-Hypothese bei Medikamententests lautet immer »Das Medikament ist wirkungslos«. Das heißt, dass der Mittelwert der Differenzen genau 0 betragen sollte. Wenn das Medikament den Blutdruck beeinflusst, wird sich der Mittelwert der Differenzen vom Wert 0 signifikant unterscheiden. Auch das kann mit einem t-Test getestet werden.

2.3 Chi-Quadrat-Test und G-Test

Beide Tests können für dieselben Aufgaben verwendet werden. Der schwierigste Rechenschritt in einem Chi-Quadrat-Test ist das Quadrieren von Zahlen. Im G-Test müssen Logarithmen berechnet werden. Daher wurde in den Zeiten, als man Statistik hauptsächlich mit Papier und Bleistift unter Zuhilfenahme von Tabellenbüchern gemacht wurde, praktisch nur der Chi-Quadrat-Test gemacht. Seit es aber Computer und ausgereifte Statistik-Programme gibt, gehört der Chi-Quadrat-Test eigentlich zum alten Eisen, weil der G-Test dieselbe Aufgabe viel genauer löst, aber "intern" eben aufwändige Rechenschritte erfordert. Trotzdem wird der G-Test in Lehrbüchern über Statistik kaum erwähnt. Auch hier wird auf die genaue mathematische Beschreibung des G-Tests verzichtet. Bitte konsultieren Sie die einschlägige Fachliteratur wenn Sie mehr darüber wissen wollen.

Beide Tests (Chi-Quadrat- und G-Test) gibt es jeweils in zwei Ausprägungen. In beiden Fällen ist die Grundgesamtheit in Kategorien unterteilt (z.B. Gummibärchen in verschiedenen Geschmacksrichtungen). Der Test ist ungeeignet für stetige Verteilungen.

2.3.1 Chi-Quadrat-Test auf Anpassungsgüte

Es wird behauptet, dass in jeder Packung Gummibärchen alle fünf Geschmacksrichtungen zu je 20% vertreten sind.

Der Chi-Quadrat-Test ist eine Näherung des mathematisch viel komplexeren G-Tests. Der G-Test kann unabhängig von den zu erwartenden Häufigkeiten in den Stichproben gemacht werden, den einfacheren Chi-Quadrat-Test sollte man nur durchführen, wenn in der Stichprobe pro Kategorie jeweils mindestens 5 Exemplare erwartet werden. Wenn eine Stichprobe eine Tüte Gummibärchen ist, sollten also pro Geschmacksrichtung 5 Bärchen erwarten werden, sonst liefert der Chi-Quadrat-Test möglicherweise eine falsche Antwort.

Der Chi-Quadrat-Test wird wie folgt durchgeführt:

Sie ermitteln pro Geschmacksrichtung die Differenz zwischen der erwarteten Anzahl und der Anzahl aus der Stichprobe. Diese Differenzen quadrieren Sie und teilen Sie durch die erwartete Anzahl. Die so erhaltenen Werte addieren sie um die Prüfgröße zu erhalten. Wenn diese Prüfgröße klein genug ist, kann man davon ausgehen, dass die Verteilung in der Stichprobe der erwarteten Verteilung entspricht.

Beispiel:

Geschmack	Anzahl in der Stichprobe	erwartete Anzahl	Differenz	Quadrat	Quotient
Apfel	18	20	$18-20=-2$	$(-2)^2=4$	$4/20=0{,}2$
Kirsche	25	20	$25-20=5$	$5^2=25$	$25/20=1{,}25$
Orange	12	20	$12-20=-8$	$(-8)^2=64$	$64/20=3{,}2$
Zitrone	22	20	$22-20=2$	$2^2=4$	$4/20=0{,}2$
Traube	23	20	$23-20=3$	$3^2=9$	$9/20=0{,}45$

Die Summe der Werte in der letzten Spalte ergibt 5,3, und dieser Wert muss nun mit einem kritischen Wert verglichen werden, der in einem separaten Verfahren berechnet wird. Darin fließen das zuvor festgelegte Signifikanzniveau (z.B. $\alpha = 5\%$) und eine Anzahl von Freiheitsgraden ein. Diese Anzahl der Freiheitsgrade ist die Anzahl der Kategorien minus 1. Da wir im Beispiel 5 Kategorien hatten, muss man also in die Berechnung als Anzahl der Freiheitsgrade der Wert 4 einsetzen.

2.3.2 **Chi-Quadrat-Test auf Unabhängigkeit**

Hier analysiert man zwei Verteilungen, die aus zwei Stichproben stammen, und setzt in der oben beschriebenen Berechnung anstelle der theoretisch erwarteten Anteile die Anteile der zweiten Stichprobe ein.

Beispiel:

Sie stellen bei einem Kindergeburtstag Gummibärchen bereit, weisen die Kinder aber an, dass die Mädchen nur Bärchen aus der runden Schüssel auf dem Küchentisch essen dürfen, während die Buben nur Gummibärchen aus der eckigen Schüssel auf dem Couchtisch essen dürfen. Wenn das Fest fertig ist, zählen Sie bei den Mädchen und bei den Buben, wie viele Gummibärchen pro Sorte jeweils gegessen wurden, und führen mit diesen Messergebnissen die ohne beschriebene Rechnung durch. Sie bekommen dann die Antwort auf die Frage, ob Buben und Mädchen dieselben Gummibärchen im selben Ausmaß bevorzugen, oder ob es Unterschieden zwischen den Geschlechtern gibt.