

Statistische Tests Version 1.2

Uwe Ziegenhagen
ziegenhagen@wiwi.hu-berlin.de

7. Dezember 2006

1 Einführung

Ein statistischer Test dient der Überprüfung einer statistischen Hypothese. Mithilfe des Tests läßt sich überprüfen, ob ein bestimmtes Verhältnis (Anteil, Mittel) in einer Stichprobe auf den Zufall zurückzuführen ist oder nicht.

Generell geht dabei man in folgenden Schritten vor:

1. Aussage über die Verteilung der Grundgesamtheit, diese Information muß bekannt sein.
2. Formulierung von Nullhypothese H_0 und Alternativhypothese H_1 , Festlegung von des Signifikanzniveaus α
3. Bestimmung des Annahme- und Ablehnungsbereichs der Nullhypothese
4. Ziehung der Stichprobe und Berechnung der Prüfgröße
5. Treffen der Testentscheidung und Interpretation: Liegt die Prüfgröße innerhalb des Annahmebereichs, wird H_0 angenommen, anderenfalls abgelehnt.

2 Hypothesenwahl

Für das richtige Aufstellen der Hypothesen gibt es ein paar Regeln:

1. Was ich zeigen/beweisen will, gehört in die Alternativhypothese.
2. Hypothesenaufstellung der Art, dass die irrtümliche Ablehnung der Nullhypothese die schlimmeren Auswirkungen hätte.
3. Das Gleichheitszeichen gehört immer in die Nullhypothese.

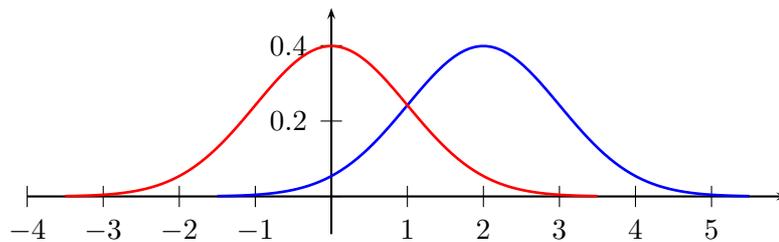
4. Bei Verteilungstests: Daß die Verteilung der Stichprobenverteilung einer bestimmten Verteilung folgt, läßt sich nicht beweisen. Wohl aber läßt sich statistisch beweisen, wenn die Stichprobenverteilung einer theoretischen Verteilung *nicht* folgt, d.h. die Abweichungen von den theoretisch zu erwartenden Werten zu groß ist.

[Zum Begriff 'Beweis' ein Auszug aus Becker/Genschel]

Beim Testen wird nur die Wahrscheinlichkeit α kontrolliert, d.h. die Wahrscheinlichkeit der irrtümlichen Ablehnung der Nullhypothese $P(H_1|H_0)$. Wenn H_0 tatsächlich wahr ist, wird nur in $\alpha \cdot 100\%$ der Fälle für H_1 entschieden.

Die Entscheidung für H_1 ist in diesem Sinn statistisch abgesichert. Bei einer Entscheidung für die Alternativhypothese und gegen die Nullhypothese spricht man von einem signifikanten Ergebnis.

Die Wahrscheinlichkeit β für den Fehler 2. Art wird nicht kontrolliert, daher ist eine Entscheidung für H_0 nicht statistisch abgesichert. Wenn die Nullhypothese H_0 nicht verworfen wird, bedeutet das nicht, sich aktiv für H_0 zu entscheiden. Es spricht nur nichts gegen die Nullhypothese.



Beispiel Ein Reifenfabrikant¹ fertigt Fahrradreifen und hat ein neues Modell entwickelt. Die Frage ist nun, ob es besser als das bisherige Modell ist, das eine mittlere Haltbarkeit von 8000 Km aufwies. Die Laufleistung eines Reifens sei normalverteilt, die Standardabweichung sei bekannt mit 1800 Km.

Nach Angaben der Forschungsabteilung bringt das neue Modell eine Leistung von 10000 Km. Bevor aber mit der teuren Werbekampagne begonnen wird, soll intern mit einem statistischen Test mit der Stichprobengröße $n = 36$ und dem Signifikanzniveau $\alpha = 5\%$ geprüft werden, ob der neue Reifen besser ist als der alte.

Welche Fehler kann die Firma in unserem Fall machen? Bei dem Test könnten die in Wahrheit besseren Reifen als fehlerhaft (nicht besser als die jetzigen) bezeichnet werden und müßten zurück ins Labor. Falls die Reifen hingegen nicht besser sind, aber als 'besser' klassifiziert werden, müßte die Firma einen Imageverlust hinnehmen.

¹Beispiel aus Bohley, ISBN: 3-486-21835-2

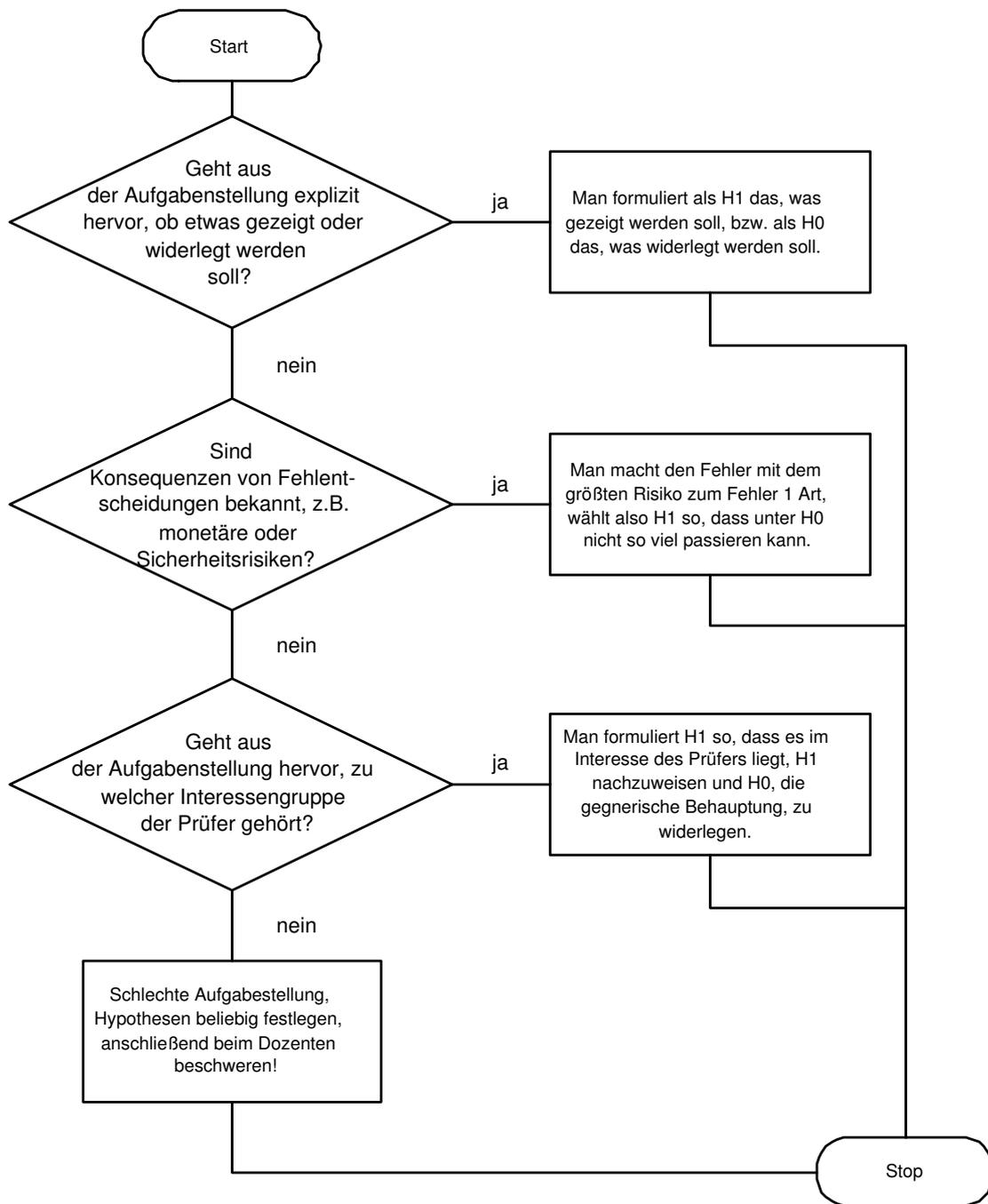


Abbildung 1: Hypothesenwahl, aus Rinne, 'Taschenbuch der Statistik'

Wenn man diese beiden Fälle miteinander vergleicht, wäre der Schaden im zweiten Fall vermutlich größer. Als Konsequenz möchte die Firma die Wahrscheinlichkeit α der *irrtümlichen* Behauptung 'Die neuen Reifen sind einfach besser!' nach oben begrenzen.

	H_0 ist richtig	H_1 ist richtig
H_0 annehmen ' H_0 '	ok	β , Fehler 2. Art
H_1 annehmen ' H_1 '	α , Fehler 1. Art	ok

Für die Erstellung der Hypothesen gibt es einen einfachen Trick:

Die Wahrscheinlichkeit, dass der Test sagt, die Reifen wären super, obwohl sie es nicht sind, soll maximal α betragen.

- $P(\text{'Reifen sind besser'} | \text{'Reifen sind nicht besser'}) = \alpha$

α ist definiert als die 'Wahrscheinlichkeit für die irrtümliche Ablehnung der Nullhypothese'²

- $P(H_1 | H_0) = \alpha$

Also korrespondiert H_1 mit 'Reifen sind besser' und H_0 mit 'Reifen sind nicht besser'.

- $H_0: \mu \leq \mu_0$
- $H_1: \mu > \mu_0$

Wir unterscheiden normalerweise drei Arten von Tests: linksseitig, rechtsseitig und zweiseitig. Bei einer $>$ -Alternativhypothese spricht man von einem rechtsseitigen Test, bei einer $<$ -Alternativhypothese von einem linksseitigen Test und bei \neq von einem zweiseitigen Test.

3 Aufstellung der Entscheidungsregeln

Um bei unserem Test zu einem quantifizierbaren Ergebnis zu kommen, benötigen wir eine Entscheidungsregel, einen Schätzer, die aus unserer Stichprobe eine konkrete Zahl formt. Diese Stichprobenfunktion nennt man *Teststatistik*, bezeichnet mit $V = V(X_1, X_2, \dots, X_n)$. Die Teststatistik V ist eine Funktion der zufälligen Stichprobenvariablen und damit auch eine Zufallsvariable mit einer dazugehörigen Verteilung. Bei der Stichprobenziehung erhalten wir dann eine Ausprägung der Teststatistik, $v = v(x_1, x_2, \dots, x_n)$.

Wenn die Ausprägung v 'sehr nahe' am hypothetischen Parameterwert (z.B. μ_0) liegt, kann man den Unterschied auf zufällige Einflüsse schieben, in diesem Fall wird man die Nullhypothese nicht ablehnen. Große Abweichungen hingegen machen die Nullhypothese unplausibel. In diesem Fall ist es plausibler, wenn man annimmt, daß die Stichprobe aus einer anderen Grundgesamtheit (mit einem anderen Parameter μ zum Beispiel) stammt.

Obwohl wir dann die Nullhypothese verwerfen, bedeutet das noch nicht, dass diese falsch ist. Wir können nur sagen, dass es nicht sehr vermutlich ist, dass die hypothetische Verteilung unsere Stichprobe generiert hat. Um jetzt festzulegen was 'sehr nahe' bedeutet, bestimmen wir den Ablehnungs- und Annahmehereich.

²Auswendig lernen!

4 Ablehnungs- und Annahmehbereich

Zur Bestimmung des Ablehnungs- und Annahmehbereichs nutzen wir das Wissen aus der Wahrscheinlichkeitsrechnung. Die Wahrscheinlichkeit, dass V in den Ablehnungsbereich der Nullhypothese fällt, obwohl diese richtig ist, soll ja nicht größer sein als α

$$P(V \text{ liegt im Ablehnungsbereich} | \text{Nullhypothese gilt}) \leq \alpha$$

Daraus folgt:

$$P(V \text{ liegt im Annahmehbereich} | \text{Nullhypothese gilt}) \geq 1 - \alpha$$

Mit der Wahrscheinlichkeit α für die irrtümliche Ablehnung der Nullhypothese, die historisch bedingt meist auf 0.01, 0.05 oder 0.1 gesetzt wird und mit der Verteilungsannahme über die Grundgesamtheit³ können wir jetzt für den ein- und zweiseitigen Test die Grenzen der Intervalle bestimmen.

In unserem Beispiel, dem rechtsseitigen Test, waren die Hypothesen so gewählt:

- $H_0: \mu \leq \mu_0 = 8000 \text{ Km}$
- $H_1: \mu > \mu_0 = 8000 \text{ Km}$

Der Ablehnungsbereich der Teststatistik besteht also aus den Realisationen v der Zufallsvariablen V , die größer sind als der kritische Wert c .

$$\{v | v > c\}$$

(Lies: Die Menge aller Realisationen v , für die gilt: $v > c$) Die Wahrscheinlichkeit, dass V in den Ablehnungsbereich fällt, obwohl die Nullhypothese gilt, ist dann

$$P(V > c | H_0) \leq \alpha$$

Der Annahmehbereich für H_0 ist dann natürlich die Menge v der Realisationen von V , für die gilt:

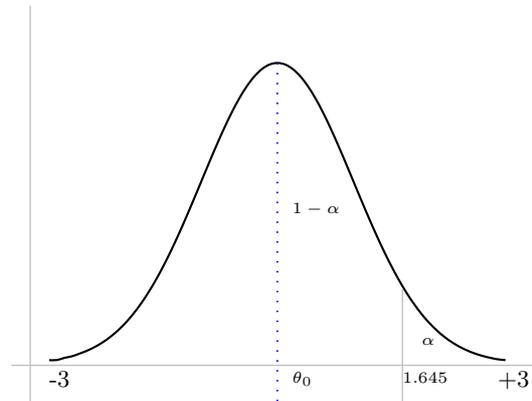
$$\{v | v \leq c\}$$

Die Wahrscheinlichkeit, dass V in den Annahmehbereich fällt, ist dann

³In unserem Fall ist die Lauflistung eines Reifens normalverteilt mit μ und σ , die durchschnittliche Lauflistung dann ebenfalls normalverteilt mit μ und σ/\sqrt{n}

$$P(V \leq c | H_0 \text{ gilt}) \geq 1 - \alpha$$

Grafisch sieht das dann so aus:



Das weitere Vorgehen ist jetzt einfach. Wir suchen in der Tabelle der Standardnormalverteilung den kritischen Wert c , der einer Wahrscheinlichkeit von 95% entspricht⁴ und lesen den Wert 1.645 ab. Wenn also der *standardisierte* Wert v größer 1.645 ist, lehnen wir die Nullhypothese ab, ansonsten nehmen wir sie an.

Zur Standardisierung: Jede Normalverteilung mit μ und σ kann zur Standardnormalverteilung transformiert werden, für die wir tabellierte Werte haben. Die Formel zur Standardisierung lautet:

$$Z = \frac{\bar{X} - \mu_0}{\sigma} \quad \text{mit} \quad \sigma = \sigma / \sqrt{n}$$

Mittels dieser Formel können wir jetzt auch Grenzen für die nicht-standardisierten Werte angeben. Dann müssen wir den Mittelwert der Stichprobe nicht transformieren, um eine Entscheidung über Annahme oder Ablehnung zu treffen.

$$\bar{X} = \frac{Z \cdot \sigma}{\sqrt{n}} + \mu_0$$

In unserem Beispiel, mit $\mu = 8000$ Km und $\sigma = 1800$ Km:

⁴Es ist ein einseitiger Test, daher suchen wir nicht nach 97.5%

$$\bar{x}_c = 8000 + 1.645 \cdot \frac{1800}{6} = 8493.5$$

und damit der Annahmereich der Nullhypothese:

$$\bar{x} \leq 8493.5$$

und der Ablehnungsbereich der Nullhypothese:

$$\bar{x} > 8493.5$$

Für die standardisierte Prüfgröße gilt also $V \leq 1.645$ und $V > 1.645$, für die nicht-standardisierte Größe $\bar{X} \leq 8493.5$ und $\bar{X} > 8493.5$ als Annahme- bzw. Ablehnungsbereich der Nullhypothese.

Nachdem (!) wir unsere Hypothesen formuliert und die Prüfgrößen festgelegt haben, ziehen wir die Stichprobe und berechnen ggf. die Stichprobenkennzahl.

Ergibt in unserem Fall die Stichprobe ein $\bar{x} = 9784.3$ Km (standardisiert: $V = 5.94767$), so lehnen wir die Nullhypothese ab und können sagen, dass sich die Reifen signifikant von den alten unterscheiden.

Linksseitiger Test

Unser Reifenherstellen muß das Material wechseln, eine Verschlechterung der Qualität ist nicht auszuschließen. In diesem Fall möchte sich der Hersteller versichern, dass er mit dem neuen Material die gleiche Qualität liefern kann, d.h. die neuen Reifen sollen nicht schlechter sein. Daher soll die Wahrscheinlichkeit, dass man gute Reifen als fehlerhaft bezeichnet, klein sein.

Als Hypothesen stellen wir auf

- $H_0: \mu \geq \mu_0 = 8000$ Km
- $H_1: \mu < \mu_0 = 8000$ Km

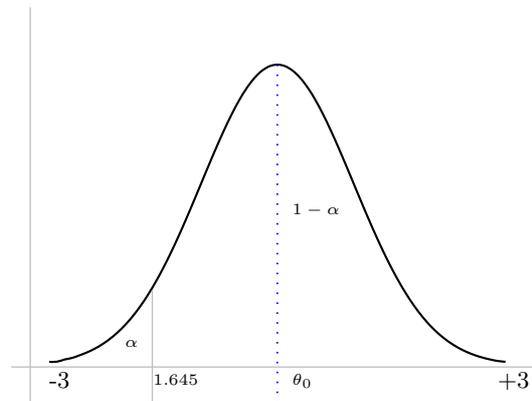
es gelte weiterhin $\alpha = 0.05$, $n = 36$ und $\sigma = 1800$.

Annahmereich der Nullhypothese:

$$\bar{x} \geq \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \quad \text{bzw.} \quad V > 1.645$$

und der Ablehnungsbereich als

$$\bar{x} < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \quad \text{bzw.} \quad V \leq 1.645$$



Als kritischen Wert ermitteln wir:

$$\bar{x}^* = 8000 - 1.645 \cdot \frac{1800}{\sqrt{36}} = 7506.5$$

Bei einem Wert einer Stichprobenziehung von $\bar{x} = 7895$ würde wir die Nullhypothese der gleichbleibenden Qualität annehmen und die Differenz zu 8000 als zufällige Differenz ansehen.

Zweiseitiger Test

Die Stiftung Warentest testet Reifen verschiedener Hersteller, die durchschnittliche Qualität liege bei 8000 Km, es gelte weiterhin $\alpha = 0.05$, $n = 36$ und $\sigma = 1800$. Testen möchte die Stiftung Warentest, ob der Reifen unseres Herstellers dieser durchschnittlichen Qualität entspricht oder nicht.

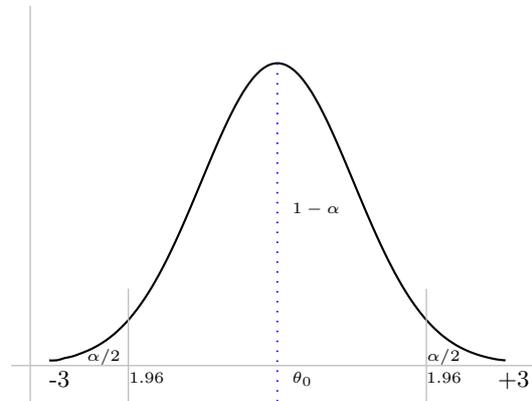
- $H_0: \mu = \mu_0 = 8000 \text{ Km}$
- $H_1: \mu \neq \mu_0 = 8000 \text{ Km}$

Als kritische Werte erhalten wir

$$\bar{x}_u = \mu_0 - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{bzw.} \quad -V$$

und

$$\bar{x}_o = \mu_0 + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{bzw. } V$$



Wir nehmen die Nullhypothese an, falls

$$\mu_0 - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

bzw.

$$-V \leq Z \leq V$$

und lehnen ab, falls

$$\bar{x} < \mu_0 - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

oder

$$\bar{x} > \mu_0 + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

In Zahlen ergibt sich der Annahmebereich der Nullhypothese als⁵

$$8000 - 1.96 \cdot \frac{1800}{6} \leq \bar{x} \leq 8000 + 1.96 \cdot \frac{1800}{6}$$

$$[7412 \leq \bar{x} \leq 8588]$$

Erhält die Stiftung Warentest in der Stichprobe einen Wert von 8213 Km fest, wird die Nullhypothese angenommen, d.h. es konnte nicht gezeigt werden, dass sich die Lebensdauer des Reifens signifikant von 8000 Km unterscheidet.

Revisionen

1.2

- Hinweise auf Verteilung der Grundgesamtheit eingefügt
- \bar{x}^* gegen \bar{x}_c für den kritischen Wert getauscht
- Fehler in Notation behoben: Mit V wird jetzt ausschließlich die standardisierte Prüfgröße bezeichnet, für die nicht-standardisierten Werte wird \bar{x} geschrieben.
- Grafik aus Rinne hinzugefügt

⁵Wir suchen in der Tabelle jetzt 0.975!