„R in Ausbildung, Forschung und Anwendung“

1.10.-3.10.2009, Bad Doberan

Tutorium

Optimale Versuchsplanung

Dieter Rasch, Petr Šimeček und Jürgen Pilz

Einführung

- Das Tutorium basiert auf dem Buch:

Dieter Rasch, Jürgen Pilz and Petr Šimeček

Optimal Design of Experiments with R

Chapman and Hall, 2010

und auf

Rasch, D. Verdooren, R.L. und Gowers

Planung und Auswertung von Versuchen und erhebungen, Oldenbourg, 2007

Einführung

- **Bemerkung:** Zufallsvariablen sind im folgenden fett geschrieben. Nummerierungen entsprechen denen des Buches.

Gesellschaftliche Relevanz - I

Versuchsplanung macht Arbeit, kostet Zeit und ist sehr viel schwieriger, als einfach mit der Untersuchung anzufangen und bei Vorliegen der Daten das Untersuchungsziel zu formulieren und abzuwarten, welche Aussagesicherheit sich ergibt.

Immer weniger ist die Gesellschaft, die die Forschung ja finanziert, bereit, ein solches Vorgehen zu tolerieren.

Gesellschaftliche Relevanz - II

Vor allem in sehr sensiblen Forschungsbereichen wie in der klinischen Forschung, der Reaktorforschung, der toxikologischen Forschung oder der Forschung zur genetischen Modifikation oder Klonung von Organismen oder bei Tierversuchen, wird weltweit gefordert oder danach gestrebt, detaillierte Versuchspläne durch bestimmte Gremien vor dem Versuchsbeginn beurteilen zu lassen. Nur bei Zustimmung darf der Versuch begonnen werden. Was hier ein Problem der Erhaltung der Menschheit und der Natur ist, sollte anderswo analog als Problem der Kostendämpfung gesehen werden.

Gesellschaftliche Relevanz - III

Negative Example:

The following text in italics was taken from an information of the Ministry of Health in Austria entitled:

SCIENTIFIC ARGUMENTS FOR AN IMPORT BAN OF GENETICALLY

MODIFIED MAIZE MON 863 (Zea mays L., line MON 863) OF MONSANTO

(NOTIFICATION C/DE/02/9)

On 8th August 2005 the Decision (2005/608/EC) concerning the placing on the market, in accordance with Directive 2001/18/EC of genetically modified maize MON 863 was adopted by the Commission.

Gesellschaftliche Relevanz - IV

The product may be placed on the market and put to the same uses as any other maize, with the exception of

cultivation and uses as or in food.

On 13th January 2006 the placing on the market of foods and food ingredients derived from genetically modified maize line MON 863 as novel foods or novel food ingredients under Regulation (EC) No 258/97 was authorised.

Gesellschaftliche Relevanz - V

With regard to the ***studies on nutritional equivalence assessment in farm animals****, which are quoted in HAMMOND et al. (2006) as scientific proof for the safety of maize MON 863, a lot of shortcomings have been detected:*

In this document, the scientific arguments, which are justifying the Austrian import ban of this GMO, are described. They focus particularly on the toxicological safety assessment and the antibiotic resistance marker (ARM) gene nptII, which is contained in maize MON 863, but also on the given risk management measures to prevent accidental spillage.

Gesellschaftliche Relevanz - VI

Summarizing *the evaluation of the* ***toxicological safety assessment*** *of the dossier, it can be stated that a lot of deficiencies are obvious:*

Concerning the experimental design it has to be criticised that reference groups are often contributing 60-80% of the sample size. Statistically significant differences between test and control groups are therefore often masked because group differences between iso- and transgenic diets fall into the broad range of reference groups.

Gesellschaftliche Relevanz - VII

An important factor is also the sensitivity of the animal model: HAMMOND et al. (2006) described the use of an outbred rat model. The study compared a high number of different lines of maize, among them MON 863. The data vary considerably in and between the groups. That would allow the assumption that only effects with great deviations from the control would have been detectable with the chosen trial setup.

Gesellschaftliche Relevanz - VIII

Soweit der offizielle Text des Ministeriums. Wir betrachten das Experiment von Hammond et al. (2006) ohne Referenzgruppe. Dann haben wir 2 Gruppen von Ratten gefüttert mit MON 863-freiem Getreide und eine gefüttert mit 11% Anteil von MON 863 und eine gefüttert mit 33% Anteil von MON 863. Jede Gruppe bestand aus n = 20 Tieren.

Gesellschaftliche Relevanz - IX

Was bedeutet das ? Angenommen.

α = 0,05 ; β = 0,01 falls $\delta = 0{,}02$ Falls die Fruchtbarkeit 80% beträgt. Dann müsste man in jeder der 4 Gruppen

10893 Ratten füttern..

Mit 20 Tieren, α= 0.1 (einseitig) und 100*β = 20% wird die Nullhypothese fehlender Schädung angenommen, so lange die Monsantogruppen eine Fruchtbarkeit von mindestens 0.4375 besitzen.

Wie man zu diesen Zahlen kommt, soll nun gezeigt werden.

Untersuchungen = Erhebungen und Versuche

Erhebungen und Versuche unterscheiden sich hinsichtlich der Rolle, die der Forscher spielt. Bei Erhebungen greift er kaum in das Geschehen ein, er beobachtet.

Das ist typisch für die ökonomische und die soziologische Forschung aber auch in Teilen der Forstwissenschaft, in der Ökologie und in der Populationsbiologie.

Versuche

In anderen Bereichen experimentiert man, düngt mit verschiedenen Mitteln, füttert unterschiedlich, vergleicht Therapien. In der industriellen Forschung muss man häufig den Einfluss verschiedener Faktoren auf die Qualität oder/und Quantität eines Endprodukts untersuchen oder Produktionsbedingungen optimal einstellen. Dies kann eine systematische Variation der Faktoren erfordern.

Wie beginnt man in der empirischen Forschung?

Die Hauptsache ist, dass die Aufgabenstellung klar formuliert wird. Das klingt selbstverständlich, in den Beispielen wird verdeutlicht, dass dies oft nicht einfach ist und Tage und Wochen in Anspruch nehmen kann. Da ein Versuch oft viel länger dauert als seine Auswertung, kann man eine falsche Auswertung richtig erhobener Daten sehr schnell korrigieren, einen unzureichend durchgeführten Versuch kann aber oft die raffinierteste Auswertung nicht retten, er muss dann mit hohem Kostenaufwand wiederholt werden.

Stufen der Erkenntnisgewinnung

(i) Formulierung des Problems,
- (ii) Angabe der Genauigkeitsforderungen,
- (iii) Auswahl des statistischen Modells
- (iv) (optimale) Planung der Untersuchung
- (v) Durchführung der Untersuchung,
- (vi) Auswertung der Beobachtungsergebnisse,
- (vii) Interpretation der Ergebnisse.

Beispiel: Vergleich zweier Therapien

Zwei Therapien *A* und *B* für erkrankte Tiere sind hinsichtlich ihrer Wirksamkeit in einem Versuch zu vergleichen. Man bestimmt den minimalen Versuchsumfang (Schritt iv), der für bestimmte Genauigkeitsforderungen für jede Therapie erforderlich ist. Dann wird der Versuch durchgeführt (Schritt v) und die beste der beiden Therapien ausgewählt (Schritt vi).

Alternatives Vorgehen

Alternatives Vorgehen sind **sequentielle Versuche.**

Die Schritte (v) und (vi) werden mehrfach nacheinander abgearbeitet, bis die gewünschte Genauigkeit erreicht wurde.

Beispiel:Vergleiche zweier Therapien *A* und *B*. Nach Anwendung beider Therapien auf mindestens je zwei Tiere kann die Auswertung zu einem der folgenden Ergebnisse führen:

Alternatives Vorgehen

Sequentielle Versuche.

- *A* ist besser als *B*
- *B* ist besser als *A*
- *A* und *B* sind äquivalent
- die Untersuchung wird mit *A* fortgesetzt
- die Untersuchung wird mit *B* fortgesetzt.

Minimale Wahl des Versuchsumfanges

Punktschätzung von μ

$$\frac{\sigma^2}{n} \le C$$

$$n = \left\lceil \frac{\sigma^2}{C} \right\rceil$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ, σ^2 bekannt

$$[\,\bar{y} - u(1-\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}};\ \bar{y} + u(1-\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}]$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ^2 bekannt

$$\mathrm{E}(\boldsymbol{H}) = u(1-\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}.$$

$$u(1-\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}} \le \delta$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ^2 bekannt

$$n = \left\lceil u^2\left(1-\frac{\alpha}{2}\right)\frac{\sigma^2}{\delta^2} \right\rceil$$

R-Programm

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ^2 unbekannt

$$[\,\bar{\boldsymbol{y}} - t(n-1;1-\frac{\alpha}{2})\frac{\boldsymbol{s}}{\sqrt{n}};\ \bar{\boldsymbol{y}} + u(n-1;1-\frac{\alpha}{2})\frac{\boldsymbol{s}}{\sqrt{n}}]$$

$$E(\boldsymbol{H}) = t(n-1;1-\frac{\alpha}{2})\frac{E(\boldsymbol{s})}{\sqrt{n}} = \frac{t(n-1;1-\frac{\alpha}{2})}{\sqrt{n}}\,\frac{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}\sigma$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt

$$n = h(n) = \left\lceil t^2(n-1;1-\frac{\alpha}{2}) \frac{2\cdot\Gamma^2\left(\frac{n}{2}\right)\cdot}{\Gamma^2\left(\frac{n-1}{2}\right)(n-1)} \frac{\sigma^2}{\delta^2} \right\rceil$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt
Näherung für Handrechnungen

$$\frac{2\cdot\Gamma^2\left(\frac{n}{2}\right)\cdot}{\Gamma^2\left(\frac{n-1}{2}\right)(n-1)} \approx 1$$

$$n = h(n) = \left\lceil t^2(n-1;1-\frac{\alpha}{2}) \frac{\sigma^2}{\delta^2} \right\rceil$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt
Näherung für Handrechnungen

$$n = h(n) = \left\lceil t^2(n-1;1-\frac{\alpha}{2}) \frac{\sigma^2}{\delta^2} \right\rceil$$

Iterative Lösung; Beispiel

δ=0,5σ; α=0,05.

$$t(\infty;0{,}975) = 1{,}96$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt
Näherung für Handrechnungen

$$n = \left| 1{,}96^2 \cdot 4 \right| = \lceil 15{,}36 \rceil = 16$$

$$t(15;0{,}975) = 2{,}131$$

$$n = \left| 2{,}131^2 \cdot 4 \right| = \lceil 18{,}16 \rceil = 19$$

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt

$$t(18;0{,}975) = 2{,}101$$

$$n = \left| 2{,}101^2 \cdot 4 \right| = \lceil 17{,}66 \rceil = 18$$

$$t(17;0{,}975) = 2{,}11$$

$$n = \left| 2{,}11^2 \cdot 4 \right| = \lceil 17{,}8 \rceil = 18$$

also n=18!

Minimale Wahl des Versuchsumfanges

Intervallschätzung von μ σ unbekannt

R-Programm

```
[1] 18
```

Hypothesenprüfung

Hypothesen über Parameter statistischer Modelle werden mit Hilfe statistischer Tests geprüft. Ein statistischer Test ist eine Regel, mit der man auf Grund einer oder mehrerer Zufallsstichproben eine Wahl zwischen zwei sich ausschließenden (aber zusammen alle Möglichkeiten umfassenden) Hypothesen über den Parameter θ der Verteilung der Elemente der Zufallsstichprobe trifft.

Hypothesenprüfung

Einen statistischen Test, oder kurz einen Test kann man auch als die Lösung eines Zweientscheidungsproblems ansehen. In der Mathematischen Statistik und speziell in ihren Anwendungen werden oft auch Drei- und Mehrentscheidungsprobleme als Tests bezeichnet.

Hypothesenprüfung

Wir weisen hier ausdrücklich daraufhin, dass das Ergebnis eines statistischen Tests keine Aussage über die Wahrheit einer der zwei Hypothesen ist.

Hypothesenprüfung

Eine Hypothese wird durch einen Test entweder angenommen oder abgelehnt - es wird damit eine Handlungsempfehlung ausgesprochen und nichts darüber ausgesagt, ob die Hypothese, die angenommen wurde, wirklich gilt oder nicht.

Hypothesenprüfung

Das kommt daher, dass wir Wahrheitsaussagen nicht aus empirisch erhaltenen Realisationen einer Zufallsstichprobe herleiten können. Die Annahme einer Hypothese kann also zu Recht oder zu Unrecht erfolgen.

Hypothesenprüfung

Daraus folgt aber, dass zwei Arten von Fehlern möglich sind, in Abhängigkeit davon, welche der zwei Hypothesen zu Unrecht angenommen wurde. Wir bezeichnen diese Fehler als Fehler erster und zweiter Art.

Hypothesenprüfung

Dies spiegelt ein gewisses Ungleichgewicht zwischen den zwei Hypothesen wider. Wir nennen eine von ihnen die

Nullhypothese H_0 und die andere die **Alternativhypothese H_A**.

Hypothesenprüfung

Der Name Nullhypothese ist darauf zurückzuführen, dass man häufig die Hypothese fehlender Unterschiede zwischen Parametern, z.B. Mittelwerten (also, dass die Differenz zweier Mittelwerte gleich "Null" ist) zur Nullhypothese erklärt.

Hypothesenprüfung

Die Wahrscheinlichkeit dafür, die Null-hypothese zu Unrecht zu verwerfen, kann durch einen Test sehr leicht kontrolliert werden.

Die Tatsache, dass die Nullhypothese verworfen wird, obwohl sie richtig ist, bezeichnen wir als Fehler erster Art.

Hypothesenprüfung

Nehmen wir dagegen die Nullhypothese an oder, was dasselbe bedeutet, lehnen wir die Alternativhypothese ab, wenn die Nullhypothese falsch ist, so bezeichnen wir diesen Fehler als Fehler zweiter Art.

Hypothesenprüfung

Beide Fehler haben unterschiedliche Auswirkungen. Manchmal ist der Fehler erster Art und manchmal der Fehler zweiter Art der mit weitreichenderen ökonomischen oder ökologischen Konsequenzen.

GMO-Problematik

Statistische Qualitätskontrolle

Hypothesenprüfung

Beispiel 3.8

Angenommen, eine Zufallsvariable $\boldsymbol{y}$ mit Erwartungswert μ und bekannter Varianz σ^2 ist normalverteilt (kurz nach $N(\mu,\sigma^2)$ verteilt). Mit irgendeinem Wert μ_0 wollen wir die Nullhypothese $H_0 : \mu = \mu_0$

gegen die Alternativhypothese $H_A : \mu \neq \mu_0$

testen.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Der Fehler erster Art tritt auf, wenn $\mu = \mu_0$ ist und die Nullhypothese abgelehnt wird. Der Fehler zweiter Art tritt auf, wenn die Nullhypothese angenommen wird und $\mu \neq \mu_0$ gilt.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Letzterer Fehler ist sicher um so schwerwiegender, je größer der Unterschied zwischen μ und μ_0 ist. Wie wir später sehen werden, ist die Wahrscheinlichkeit dafür, dass dieser Fehler auftritt, umso kleiner, je größer $|\mu - \mu_0|$ ist.

Hypothesenprüfung

Definition 3.2

Eine Hypothese heißt einfach, wenn sie nur aus einem Punkt des Parameterraumes besteht, ansonsten heißt sie zusammengesetzt. Liegt der Alternativhypothesenbereich bei reellen Parametern ausschließlich auf einer Seite des Nullhypothesenbereiches, so heißt die Hypothese und der entsprechende Test einseitig, ansonsten zweiseitig.

Hypothesenprüfung

Definition 3.2 - Fortsetzung

Die (fälschliche) Ablehnung einer gültigen Nullhypothese wird als Fehler erster Art und die (fälschliche) Ablehnung einer gültigen Alternativhypothese wird als Fehler zweiter *Art* bezeichnet.

Hypothesenprüfung

Definition 3.2 - Fortsetzung

Die Wahrscheinlichkeit für einen Fehler erster Art nennen wir das Risiko erster Art $\alpha(\theta)$, das Maximum α von $\alpha(\theta)$ im Nullhypothesenbereich heißt Signifikanzniveau. Die Wahrscheinlichkeit für einen Fehler zweiter Art nennen wir das Risiko zweiter Art $\beta(\theta)$.

Hypothesenprüfung

Definition 3.2 - Fortsetzung

Die Wahrscheinlichkeit $\pi(\theta)$ für die Ablehnung der Nullhypothese (egal, ob sie gültig ist oder nicht) als Funktion des unbekannten Parameter θ einer Verteilung über den die Hypothesen formuliert wurden, heißt Gütefunktion.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Unter den Annahmen von Beispiel 3.8 soll die (einfache) Nullhypothese gegen eine (einseitige aber zusammengesetzte) Alternativhypothese geprüft werden, d.h.:

$H_0 : \mu = \mu_0$ (Nullhypothese)

ist gegen die einseitige Hypothese

$H_A : \mu > \mu_0$ (Alternativhypothese),

mit bekanntem σ^2 zu prüfen.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Ist $\bar{y}$ der Mittelwert der n Elemente y_i der Zufallsstichprobe , wobei letztere nach $N(\mu, \sigma^2)$ verteilt sind, so kann der Test mit der Prüfzahl

$$u = \frac{\bar{y} - \mu_0}{\sigma} \sqrt{n}$$

durchgeführt werden.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Um das für ein bestimmtes Signifikanzniveau, das in diesem Fall gleich dem eindeutigen Wert des Risikos erster Art α ist (da der Bereich der Nullhypothese nur aus dem Punkt μ_0 besteht) tun zu können, müssen wir die Verteilung von u kennen.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Da aber nach obiger Voraussetzung die Elemente der Zufallsstichprobe einer

$N(\mu; \sigma^2)$-Verteilung folgen, ist $\bar{y}$

$N(\mu; \sigma^2/n)$-verteilt und u ist $N(0; 1)$ -verteilt.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Wir nennen

$$\lambda = \lambda(\mu) = \frac{\mu - \mu_0}{\sigma} \sqrt{n} \qquad (3.45)$$

den **Nichtzentralitätsparameter** der Verteilung der Prüfzahl u.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Falls die Nullhypothese gilt, ist der Nichtzentralitätsparameter gleich Null.

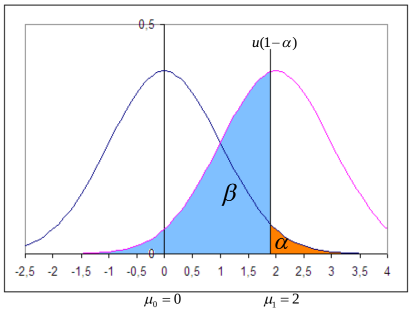
Unter der Nullhypothese haben Prüfzahlen oft eine einfachere Verteilung als unter der Alternativhypothese und Tabellen ihrer Quantile sind einfach verfügbar.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

In unserem Fall ist $\boldsymbol{u}$ $N(0,1)$-verteilt, falls die Nullhypothese gilt. Die Wahrscheinlichkeit dafür, dass eine $N(0,1)$-verteilte Zufalls-variable größer als das $(1-\alpha)$-Quantil $u(1-\alpha)$ der Standardnormalverteilung ist, ist gerade gleich α.

4.2 Testen von Hypothesen



Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Lehnen wir nun H_0 ab, wenn $u > u(1-\alpha)$ ist, so ist die Wahrscheinlichkeit dafür, dass die Nullhypothese abgelehnt wird, wenn sie gilt, gleich dem vorgegebenem Risiko erster Art α..

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Wir wollen nun die Gütefunktion dieses Tests berechnen und setzen dafür vereinfachend voraus, dass σ^2 ganzzahlig ist und $n = \sigma^2$ gewählt wurde. Ferner sei $\mu_0 = 6$. Dann kann die Wahrscheinlichkeit

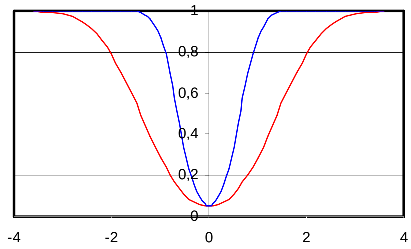
$\pi(\mu) = P(\boldsymbol{u} > u)$ für jedes reelle u einfach als Funktion vom μ dargestellt werden.

Hypothesenprüfung

In Beispiel 3.8 wächst $\pi(\theta)$ von seinem kleinsten Wert $\alpha = 0{,}05$ an der Stelle $\mu = 6$ monoton in μ. Für μ-Werte außerhalb des Nullhypothesenbereiches aber innerhalb des Bereiches der Alternativhypothese ist $\pi(\theta) = 1 - \beta(\theta)$ die Wahrscheinlichkeit für die Ablehnung der Nullhypothese wenn sie gilt.

Hypothesenprüfung

Man beachte, dass $\pi(\theta)$ nur in der Vereinigung der Bereiche beider Hypothesen definiert ist. Je stärker die μ-Werte vom Nullhypothesenwert 6 abweichen, desto kleiner ist $\beta(\theta)$, d.h. die Gefahr, große Abweichungen nicht zu bemerken, ist relativ gering.



Die Gütefunktionen zur Prüfung der Nullhypothese $H_o: \mu = \mu_0$ gegen $H_A: \mu \neq \mu_0$, *bei einem Risiko 1 Art von* $\alpha = 0{,}05$ für $n = 5$ (rot) und $n = 20$ (blau).

Hypothesenprüfung

Die letztere Bemerkung des Beispiels gilt für alle Tests, die wir hier behandeln, derartige Tests heißen unverfälscht. Es gibt aber Tests, bei denen die Gütefunktion ihren kleinsten Wert im Gebiet der Alternativhypothese annimmt, solche Tests heißen dann verfälscht.

Hypothesenprüfung

Tests mit einer monotonen Gütefunktion gestatten es uns folgende Genauigkeitsforderungen als Basis für die Bestimmung von Stichprobenumfängen zu formulieren:

1. Das Risiko erster Art für eine zusammengesetzte Nullhypothese darf eine Obergrenze α nicht überschreiten; im Falle einer einfachen Nullhypothese soll es gleich α sein.

Hypothesenprüfung

2. Das Risiko zweiter Art darf den Wert β_0 nicht überschreiten, solange (für einen vorher definierten Abstand) der Abstand zwischen dem tatsächlichen Wert des Parameters (im Bereich der Alternativhypothese) und dem Wert des Parameters unter der Nullhypothese mindestens gleich einem Wert δ ist, d.h., falls $\delta > |\mu - \mu_0|$.

Hypothesenprüfung

Das Symbol δ wird meist bei Hypothesen über Lageparameter, wie den Erwartungswert benutzt, dann nennt man δ die praktisch relevante Mindestdifferenz. Für Skalenparameter, wie die Varianz, benutzt man oft das Symbol τ.

Hypothesenprüfung

Die Genauigkeitsvorgabe ist damit durch das Tripel (α, β_0, δ) und eventuell durch a-priori-Kenntnisse über Parameter der Verteilung (z.B. σ^2) gegeben.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Mit $\mu_0 = 6$, wie wir das bei der Berechnung von Tabelle Ü4 angenommen hatten, fordern wir neben $\alpha = 0{,}05$ noch $\pi(\theta) \geq 0{,}8 = 1 - \beta_0$, falls $\mu \geq 8$ gilt (und folglich $\beta(\theta) \leq 0{,}2$ falls $\mu \geq 8$).

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Das bedeutet, wir fordern, dass $\beta(\theta) \leq 0{,}2$ falls zwischen μ und μ_0 ein Mindestabstand von $\delta = 2$ besteht. Wegen der Monotonie der Gütefunktion ist diese Forderung erfüllt, falls $\beta(8) \leq 0{,}2$ ist. Es ist damit ausreichend, einen minimalen Stichprobenumfang n zu bestimmen, der garantiert, dass $\beta(8) = \beta_0 = 0{,}2$ ist.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Wir müssen also n so wählen, dass für $\mu = 8$ die Wahrscheinlichkeit einer fälschlichen Annahme der Nullhypothese gleich 0,2 ist und das bedeutet, dass

$\beta(8) = P(\boldsymbol{u} < u(1-\alpha) \mid \mu = 8) = 0{,}2$

sein muss. Für $\alpha = 0{,}05$ ist $u(0{,}95) = 1{,}645$ und wir erhalten

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

$$\beta(8) = P(\boldsymbol{u} < 1{,}645 \mid \mu = 8) = 0{,}2$$

Wir schreiben nun die Prüfzahl

$$\boldsymbol{u} = \frac{\bar{y} - \mu_0}{\sigma}\sqrt{n}$$

um als Summe einer $N(0;1)$-verteilten ZV $\boldsymbol{u}_0$ und des Nichtzentralitätsparameters an der Stelle 8.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Das führt zu der Gleichung

$$\boldsymbol{u} = \frac{\bar{y} - \mu_0}{\sigma}\sqrt{n} = \boldsymbol{u}_0 + \lambda(8)$$

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Unsere Forderung können wir in der Form:

$$\beta(8) = P\left(\boldsymbol{u}_0 < 1{,}645 - \frac{2}{\sigma}\sqrt{n}\right) = 0{,}2$$

schreiben, also muss

$$1{,}645 - \frac{2}{\sigma}\sqrt{n}$$

gleich dem 0,2-Quantil der Standard-normalverteilung , also gleich -0,842 sein.

Hypothesenprüfung

Beispiel 3.8 - Fortsetzung

Daraus folgt aber

$$n = \left\lceil (1{,}645 + 0{,}842)^2 \frac{\sigma^2}{4} \right\rceil$$

Hypothesenprüfung

Allgemein lautet die Umfangsformel

$$n = \left\lceil [u(1-\alpha) - u(\beta)]^2 \frac{\sigma^2}{\delta^2} \right\rceil$$

oder

$$n = \left\lceil [u(1-\alpha) + u(1-\beta)]^2 \frac{\sigma^2}{\delta^2} \right\rceil$$

Hypothesenprüfung

Dieses Prinzip, die Gütefunktion an zwei Stellen festzulegen und daraus den hierfür erforderlichen Versuchsumfang zu berechnen wird im folgenden oft angewendet.

Tests im Einstichprobenproblem

3.4.1Prüfung von Hypothesen über den Mittelwert einer Normalverteilung

Wir wollen eine Zufallsstichprobe vom Umfang n aus einer normalverteilten Grundgesamtheit mit Erwartungswert μ und Varianz σ^2 erheben, um die Nullhypothese

$H_0 : \mu = \mu_0$ (ist eine gegebene Konstante)

Tests im Einstichprobenproblem

gegen eine der folgenden Alternativ-hypothesen:

a) $H_A : \mu > \mu_0$ (einseitige Alternative)

b) $H_A : \mu < \mu_0$ (einseitige Alternative)

c) $H_A : \mu \neq \mu_0$ (zweiseitige Alternative)

zu prüfen.

Tests im Einstichprobenproblem

Für den Fall bekannter Varianz wurde das Vorgehen bereits einleitend beschrieben.

Ist σ **unbekannt**, so ersetzen wir die Prüfzahl $\boldsymbol{u}$ durch die Studentsche Prüfzahl $\boldsymbol{t}$, die wir formal aus $\boldsymbol{u}$ erhalten, indem wir σ durch die Stichprobenstandardabweichung $\boldsymbol{s}$ ersetzen. Diese Prüfzahl ist

Tests im Einstichprobenproblem

$$t = \frac{\bar{y} - \mu_0}{s}\sqrt{n}$$

Für ein Risiko erster Art α wird H_0 abgelehnt falls

in Fall a), $t > t(n\text{-}1;1\text{-}\alpha)$,

in Fall b) $t < t(n\text{-}1;\ \alpha) = -t(n\text{-}1;\ 1\text{-}\alpha)$

in Fall c) $|t| > t(n\text{-}1;\ 1\text{-}\alpha/2)$ gilt und sonst angenommen.

Tests im Einstichprobenproblem

Die Prüfzahl (3.47) ist nichtzentral t-verteilt mit n-1 Freiheitsgraden und dem gleichen Nichtzentralitätsparameter wie bei bekannter Varianz. Unter der Nullhypothese ist wieder λ=0. Mit dem (1-α)-Quantil $t(n\text{-}1;1\text{-}\alpha\text{-})$ der zentralen t-Verteilung mit n-1 Freiheitsgraden und dem β-Quantil der entsprechenden nichtzentralen t-Verteilung $t(n\text{-}1;\lambda\text{-};\beta)$ erhalten wir in Analogie zu Beispiel 3.8

Tests im Einstichprobenproblem

für Fall a) $\quad t(n-1;1-\alpha) = t(n-1;\lambda;\beta)$

Wir verwenden für Handrechnungen eine Approximation, die zur Bestimmung von Stichprobenumfängen hinreichend genau ist und zwar

$$t(n-1;\lambda;\beta) \approx t(n-1;\beta) + \lambda = -t(n-1;1-\beta) + \frac{\mu - \mu_0}{\sigma}\sqrt{n}$$

Tests im Einstichprobenproblem

Daraus ergibt sich als Mindeststichprobenumfang

$$n = \left\lceil \left[t(n-1;1-\alpha) + t(n-1;1-\beta)\right]^2 \frac{\sigma^2}{\delta^2} \right\rceil \text{(3.51)}$$

Tests im Einstichprobenproblem

Bemerkung: Da t(FG, 0,5) = 0 für alle Werte von FG gilt, besteht folgender Zusammenhang zwischen einem zweiseitigen (1-α) - Konfidenzintervall *KI* für μ und einem zweiseitigen Test von H_0: μ = 0 mit ein Risiko erster Art α:

Tests im Einstichprobenproblem

(i) Die Formeln (3.20) und (3.51) sind identisch, falls $P = 1 - \alpha/2$ und $\beta_0 = 0{,}5$.

(ii) H_0 wird genau dann angenommen, wenn das *KI* die Null enthält.

(iii) verwenden wir (ii) als Test, so ist die Güte des Tests 1-β = 0,5 falls $\mu = 2\delta$ oder $\mu = -2\delta$ mit 2δ als die erwartete Länge des *KI*.

Tests im Einstichprobenproblem

Den Stichprobenumfang kann man mit R wie folgt berechnen:

Tests im Einstichprobenproblem

Nun gibt man die Genauigkeitsvorgaben ein und errechnet den Umfang.

Beispiel 3.9

Wir wollen den minimalen Stichgprobenumfang für d = 1/3σ; α = 0,05 und β_0 = 0,2 berechnen. Mit $n^{(0)} = \infty$ erhalten wir $t(\infty$; 0,975) = 1,96 und $t(\infty$, 0,8) = 0,84162 und damit $n^{(1)}$ = [70,64] = 71.

Tests im Einstichprobenproblem

Beispiel 3.9 – Fortsetzung

Da t(70; 0,975) = 1,9944 und t(70; 0,8) = 0,8448 gilt, erhalten wir

$n^{(2)}$ = [72,65] = 73 und dieser Wert ändert sich im nächsten Schritt nicht mehr

((72,60) = 73).

Tests im Einstichprobenproblem

Beispiel 3.1 - Fortsetzung

Wir wählen wieder die Wurfgewichte von Mäusen in der ersten Spalte von Tabelle 3.1 und testen die Hypothese: H_0: $\mu = E(\boldsymbol{x}_i)$ = 10 gegen

H_A: $\mu \neq 10$ mit dem Signifikanzniveau

α = 0,05.

Tests im Einstichprobenproblem

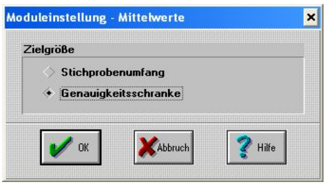
Beispiel 3.1 - Fortsetzung

Der Stichprobenumfang der Daten steht mit n = 13 bereits fest. Wenn wir bei α = 0,05 ein Risiko zweiter Art von 0,2 tolerieren, welches ist die Differenz $\delta = |\mu$- 10|, die dann mit einer Wahrscheinlichkeit 1-β = 0,8 erkannt wird, wobei der Schätzwert 3,947 für die Varianz vorliegt?

Tests im Einstichprobenproblem

Beispiel 3.1 - Fortsetzung

Wir verwenden MIWA ändern aber zunächst in der oberen Menüzeile im Schlter Optionen mit der Befehlsfolge **Optionen – Moduleinstellungen**



Tests im Einstichprobenproblem

Beispiel 3.1 - Fortsetzung

das Zielkriterium von Stichprobenumfang in Genauigkeitsschanke, wie in der vorigen Abbildung zu sehen war. Die Eingabe und das Ergebnis zeigen die folgenden Abbildungen.

R-Programm

Statistische Tests

In welcher Hauptstadt eines EU-Landes steht ein Denkmal eines Statistikers, den fast jeder von Ihnen kennt?

Statistische Tests

a) wie heißt die Stadt?

b) wie heißt der Statistiker?

Statistische Tests

a) Dublin

b) W. Gosset (Student)

Oh my God, oh my Guinness

Statistische Tests

Der Vergleich zweier Erwartungswerte an Hand unabhängiger Stichproben aus zwei Grundgesamtheiten.

Statistische Tests

Gegeben seien zwei Zufallsvariable mit Erwartungswerten

$$E(\boldsymbol{y}_1) = \mu_1; E(\boldsymbol{y}_2) = \mu_2$$

und existierenden vierten Momenten.

Statistische Tests *t* - Test

Zu prüfen ist die Nullhypothese:

$$H_0 : \mu_1 = \mu_2$$

gegen die zweiseitige Alternativ-hypothese:

$$H_0 : \mu_1 \neq \mu_2$$

Statistische Tests *t* - Test

Es seien $\boldsymbol{y}_i^T = \left(\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, ..., \boldsymbol{y}_{in_i}\right), n_i \geq 1; i = 1,2$
zwei unabhängige Zufallstichproben
mit Erwartungswertvektoren $\vec{\mu}_i = \mu_i e_{n_i}; i = 1,2$
, wobei e_n ein Vektor mit *n* Einsen ist.

Statistische Tests *t* - Test

Lemma 1 (Rasch, 1995; Satz 6.17)

Ist die Zufallstichprobe $\boldsymbol{y}^T = (y_1, y_2, ..., y_n), n \geq 1$ nach $N\left(\vec{\mu}, \sigma^2 E_n\right), \sigma^2 > 0$ verteilt, so ist $\boldsymbol{X}^2 = \frac{\boldsymbol{y}^T \boldsymbol{y}}{\sigma^2}$ nach $CQ(n, \lambda)$ verteilt mit Nichtzentralitätsparameter

$$\lambda = \frac{\vec{\mu}^T \vec{\mu}}{\sigma^2}$$

Statistische Tests *t* - Test

Lemma 4 (Beispiel 8.8)

Zu prüfen sei die Nullhypothese:

$H_0 : \mu_1 = \mu_2; \sigma^2 > 0$ beliebig

gegen die zweiseitige Alternativ-hypothese: $H_0 : \mu_1 \neq \mu_2$

Statistische Tests **Lemma 4**

Es mögen dafür zwei unabhängige Zufallsstichproben

$$\boldsymbol{y}_i^T = \left(\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, ..., \boldsymbol{y}_{in_i}\right), n_i \geq 1; i = 1,2$$

vorliegen, die nach $N\left(\vec{\mu}_i, \sigma^2 E_{n_i}\right), \sigma^2 > 0$
verteilt sind. Dann ist die Verteilung von $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

Element einer dreiparametrischen Exponentialfamilie

Statistische Tests **Lemma 4**

von vollem Rang, mit den vollständig suffizienten Maßzahlen:

$$\boldsymbol{S} = \bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2; \boldsymbol{T}_1 = n_1 \bar{\boldsymbol{y}}_1 + n_2 \bar{\boldsymbol{y}}_2; \boldsymbol{T}_2 = \sum_{i=1}^{n_1} \boldsymbol{y}_{1i}^2 + \sum_{i=1}^{n_2} \boldsymbol{y}_{2i}^2$$

Statistische Tests *t* - Test

„Tests mit Neyman-Sruktur“

Def. 10.7 (Rasch, 1995)

Gegeben Zufallsstichproben $y^T = (y_1, \ldots, y_n)$ aus einer Familie $(P^\theta, \theta \in \Omega)$ und es sei $\omega^0 = \omega^A = \Omega / \omega_* = \underline{\omega}^0 \cup \underline{\omega}^A$ heißt gemeinsamer Rand beider Hypothesen.

Statistische Tests *t* - Test

„Tests mit Neyman-Sruktur“

Es sei $(P^\theta, \theta \in \Omega_*) \in (P^\theta, \theta \in \Omega)$ die Teilfamilie auf dem gemeinsamen Rand. Es möge ferner eine nichttriviale suffiziente Maßzahl *T*(***y***) bezüglich *P** existieren, so dass für den Test *k*(***y***) E(*k*(***y***) /T(***y***)] von $\theta \in \Omega_*$ unabhängig ist.

Statistische Tests *t* - Test

„Tests mit Neyman-Sruktur“

Ein Test *k*(***y***) für den E(*k*(***y***) /T(***y***),θεΩ*] = α gilt heißt ein α - Test mit Neyman-Sruktur.

Statistische Tests *t* - Test

Für solche Tests folgt nach Satz 10.14 (Rasch, 1995, S. 333-334) die Existenz eines gleichmäßig besten unverfälschten α -Tests.

Statistische Tests **Lemma 4**

Allgemein ist der Vektor $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

Element einer fünfparametrischen Exponentialfamilie, wenn die Teilvektoren normalverteilte Zufallsstichproben sind. Die vollständig suffizienten Maßzahlen sind:

Statistische Tests **Lemma 4**

$$\sum y_{1i}; \sum y_{2i}; \sum y_{1i}^2; \sum y_{2i}^2; \sum y_{1i} \cdot y_{2i}$$

Statistische Tests **Lemma 4**

Sind beide Zufallsstichproben stochastisch unabhängig, reduziert sich die Familie auf eine vierparametrische ($\rho = 0$) mit den suffizienten Maßzahlen

$$\sum y_{1i};\sum y_{2i};\sum y_{1i}^2;\sum y_{2i}^2$$

03.02.2010 109

Statistische Tests **Lemma 4**

Setzt man darüber hinaus voraus, dass die beiden Varianzen gleich sind, haben wir eine dreiparametrische Exponentialfamilie von vollem Rang mit den suffizienten Maßzahlen

03.02.2010 110

Statistische Tests **Lemma 4**

$$\boldsymbol{S} = \bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2; \boldsymbol{T}_1 = n_1\bar{\boldsymbol{y}}_1 + n_2\bar{\boldsymbol{y}}_2; \boldsymbol{T}_2 = \sum_{i=1}^{n_1} \boldsymbol{y}_{1i}^2 + \sum_{i=1}^{n_2} \boldsymbol{y}_{2i}^2$$

03.02.2010 111

Statistische Tests ***t* - Test**

Daraus folgt schließlich dass

$$\boldsymbol{g} = \frac{\boldsymbol{S}}{\sqrt{\boldsymbol{T}_2 - \frac{1}{n_1+n_2}\boldsymbol{T}_1^2 - \frac{n_1 \cdot n_2}{n_1+n_2}\boldsymbol{S}^2}} =$$

$$\frac{\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2}{\boldsymbol{s}\sqrt{n_1+n_2-2}} \quad \text{mit} \quad \boldsymbol{s}^2 = \frac{\boldsymbol{SQ}_{y1} + \boldsymbol{SQ}_{y2}}{n_1+n_2-2}$$

03.02.2010 112

Statistische Tests ***t* - Test**

unabhängig von den Parametern ist, falls die Nullhypothese gilt. Damit folgt nun aber (Satz 8.5, Rasch, 1995), dass

$$\boldsymbol{t} = \frac{\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2}{\boldsymbol{s}}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}}$$

03.02.2010 113

Statistische Tests ***t* - Test**

nach

$$t\left(n_1+n_2-2; \lambda = \frac{\mu_1-\mu_2}{\sigma}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}}\right)$$

verteilt ist und ein gleichmäßig bester unverfälschter α **-**Test für alle $0<\alpha<1$ die Form

$$k\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{matrix} 1, \text{für} |t| > t\left(n_1+n_2-2; 1-\frac{\alpha}{2}\right) \\ 0, \text{sonst} \end{matrix}$$

hat.

03.02.2010 114

Statistische Tests **Welch - Test**

Für den Fall ungleicher Varianzen bewiesen Welch (1947) und Trickett and Welch (1954) folgenden Satz (Satz 10.18, Rasch, 1995): Es sei mit den bisherigen Bezeichnungen

$$\gamma = \frac{\frac{\sigma_1^2}{n_1}}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}};\; c = \frac{(n_1-1)\frac{s_1^2}{\sigma_1^2}}{(n_1-1)\frac{s_1^2}{\sigma_1^2}+(n_2-1)\frac{s_2^2}{\sigma_2^2}}$$

Statistische Tests **Welch - Test**

und

$$p(c) = \frac{1}{B\left(\frac{n_1-1}{2},\frac{n_2-1}{2}\right)} c^{\frac{n_1-1}{2}} \cdot (1-c)^{\frac{n_2-1}{2}-1}$$

, so ist die Verteilungsfunktion von

$$\boldsymbol{t}^* = \frac{\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\boldsymbol{s}_1^2}{n_1}+\frac{\boldsymbol{s}_2^2}{n_2}}}$$

Statistische Tests **Welch - Test**

(unter der Nullhypothese) gegeben durch

$$F(t^*) = \int_{c=0}^{1} F_{n_1+n_2-2}\left\{\sqrt{n_1+n_2-2}\,\frac{\gamma\cdot c}{n_1-1}+\frac{(1-\gamma)(1-c)}{n_2-1}\right\} p(c)\,dc$$

, wobei $F_{n_1+n_2-2}$ die VF. der zentralen ***t***-Verteilung mit n_1+n_2-2 FG ist. B ist die Betafkt. $B(m,n)=\frac{\Gamma(m)\cdot\Gamma(n)}{\Gamma(m+n)}$

Statistische Tests **Welch - Test**

Trickett and Welch (1954) geben ein Iterationsverfahren zur Bestimmung der Quantile an, Tabellen der Quantile findet man bei Aspin (1949) und Rasch u.a. (2008).

Den erforderlichen Versuchsumfang kann man mit R bestimmen.

Statistische Tests **Approximativer Welch - Test**

In Anlehnung an Satterthwaite verwendet man oft folgenden approximativen Test mit

$$\boldsymbol{t}^{**} = \frac{\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2}{\sqrt{\frac{\boldsymbol{s}_1^2}{n_1}+\frac{\boldsymbol{s}_2^2}{n_2}}}$$

$$k\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{pmatrix} = \begin{matrix}1, \text{für } |t^{**}| > t\left(f; 1-\frac{\alpha}{2}\right)\\ 0, \text{sonst}\end{matrix}$$

Statistische Tests **Approximativer Welch - Test**

mit

$$f = \left\lceil \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2}+\frac{s_2^4}{(n_2-1)n_2^2}} \right\rceil$$

Statistische Tests **Welchen Test ?**

Es besteht eine Unsitte unter Statistikern (vor allem unter Theoretikern, die selten mit Anwendungen zu tun haben), Voraussetzungen für (z.B. Tests) vorher zu prüfen und je nach Ausgang dieser Prüfung entsprechend fortzufahren. Leider oft mit den gleichen Daten.

Statistische Tests **Welchen Test ?**

Dies hat uns zu einer Arbeit angeregt. Zunächst wurde ausgerechnet, wie die Endrisiken unseres Testproblems (*t*-, Welch- oder Wilcoxon-Test) aussehen, wenn für die Tests auf Normalität jeder Grundgesamtheit und auf Varianzhomogenität jeweils unabhängige Stichproben verwendet werden.

Statistische Tests **Welchen Test ?**

Die Endrisiken sind für manche Zweige unerträglich hoch, die Versuchsumfänge für die Vortests viel höher als bei vergleichbarer Genauigkeitsvorgabe für den Endtest. Karl Moder und ich haben uns der realistischen Variante angenommen, das alle Tests am gleichen Material angewendet werden.

Statistische Tests **Welchen Test ?**

Ein umfangreiches Simulationsexperiment wurde durchgeführt.

- Wird das vorgegebene Risiko 1, Art α_{nom} des oben beschriebenen *t* – Tests durch das tatsächliche Risiko α_{akt} „einigermaßen“ eingehalten?
- Einigermaßen soll heißen:

$$\left|\alpha_{akt} - \alpha_{nom}\right| \leq 0{,}2\alpha_{nom}$$

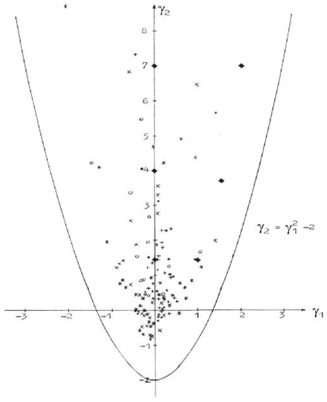
Statistische Tests **Welchen Test ?**

Für das standardisierte 3. (Schiefe) und 4. Moment -3 (Exzess) einer beliebigen Verteilung gilt folgende Ungleichung (eine analoge Ungleichung gilt für die Schätzwerte):

$$\gamma_2 \geq \gamma_1^2 - 2$$

Statistische Tests **Welchen Test ?**

Innerhalb der durch $\gamma_2 = \gamma_1^2 - 2$ definierten Parabel liegen alle theoretischen (und empirischen) Verteilungen, die ein viertes Moment besitzen.



Simulierte Verteilungen

- Es sei $\boldsymbol{u}$ nach $N(0;1)$ verteilt.
- Wir können daraus durch die Transformation

$$\boldsymbol{y} = a + b \cdot \boldsymbol{u} + c \cdot \boldsymbol{u}^2 + d \cdot \boldsymbol{u}^3$$

durch entsprechende Wahl der Koeffizienten *a, b, c* und *d* jede Verteilung innerhalb der Parabel erzeugen.

Fleishman-System, Fleishman, 1978

Fleishman-System

- Wir setzen o.B.d.A. $\mu = E(\boldsymbol{y}) = 0$; var $(\boldsymbol{y}) = 1$
- Dann ergibt sich: $a = -c$ und

$$b^2 + 6bd + 2c^3 + 15d^2 = 1$$

- Nun können wir die verbleibenden freien 2 Parameter benutzen, um jeden Punkt innerhalb der Parabel zu erreichen.

Verteilung	Schiefe γ_1	Exzess γ_2	*c = -a*	*b*	*d*
1	0	0	0	1	0
2	0	1.5	0	0.866993269415	0.042522484238
3	0	3.75	0	0.748020807992	0.077872716101
4	0	7	0	0.630446727840	0.110696742040
5	1	1.5	0.163194276264	0.953076897706	0.006597369744
6	1.5	3.75	0.221027621012	0.865882603523	0.027220699158
7	2	7	0.260022598940	0.761585274860	0.053072273491

Simulationsexperimente

- Jeder Test wurde 100 000 mal mit simulierten Daten (Paaren von ZStpr.) durchgeführt. Die relative Häufigkeit der Ablehnung der Nullhypothese diente zur Schätzung der Gütefunktion.

Statistische Tests **Welchen Test ?**

Gilt die Nullhypothese, erhält man so einen Schätzwert $\hat{\alpha}_{akt}$ für α_{akt} ,der dem wahren Wert α_{akt} recht nahe kommt.

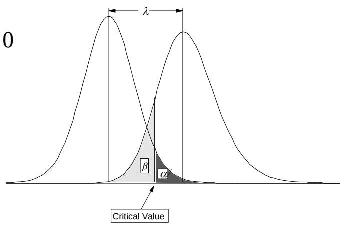
Die Vorgehensweise: Wechsel zu Ra-Ku-Mo

Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

$$H_A : \mu > \mu_0$$

$$t = \frac{\bar{y} - \mu_0}{s}\sqrt{n}$$



Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

$$t = \frac{\bar{y} - \mu_0}{s}\sqrt{n}$$

ist nichtzentral t-verteilt mit n-1 FG und Nichtzentralitätsparameter

$$\lambda = \frac{\delta}{\sigma}\sqrt{n}$$

Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

Nach der Abbildung muss

$t(n\text{-}1;\ 1\text{-}\alpha) = t(n\text{-}1;\lambda;\ 1\text{-}\beta)$ oder im zweiseitigen Fall

$t(n\text{-}1;\ 1\text{-}\alpha/2) = t(n\text{-}1;\lambda;\ 1\text{-}\beta)$ sein.

Dann folgt:

$$t(n-1;11-\frac{\alpha}{2}) = t(n-1;\frac{\delta}{\sigma}\sqrt{n};\beta)$$

Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

Näherung für Handrechnungen

$t(n\text{-}1;\lambda;\beta) = t(n\text{-}1,\beta) + \lambda$ führt zu:

$$n \approx \left\lceil \left[\left\{ t\left(n-1;1-\frac{\alpha}{2} \right) + t\left(n-1;1-\beta\right) \right\} \frac{\sigma}{\delta} \right]^2 \right\rceil$$

Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

Näherung – Beispiel

Example 2.8

$H_0 : \mu = \mu_0$

a) $H_A : \mu > \mu_0$

b) $H_A : \mu \quad \mu_0$

α = 0,01 and β = 0,05, $\delta = 0{,}6\sigma$

$$n \approx \left\lceil \left[\left\{ t(\infty;0{,}995) + t(\infty;0{,}95) \right\} \frac{1}{0{,}6} \right]^2 \right\rceil$$

Minimale Wahl des Versuchsumfanges

Test H_0 $\mu = \mu_0$, σ^2 unbekannt.

Näherung – Beispiel

Example 2.8

$$n \approx \left\lceil \left[\left\{ t(\infty;0{,}995) + t(\infty;0{,}95) \right\} \frac{1}{0{,}6} \right]^2 \right\rceil$$

$$\left\lceil \left[(2{,}576 + 1{,}645) \frac{1}{0{,}6} \right]^2 \right\rceil = \lceil 49{,}49 \rceil = 50$$

Minimale Wahl des Versuchsumfanges

Test H$_0$ $\mu = \mu_0$, σ^2 unbekannt.

Näherung – Beispiel

Example 2.8

$$n \approx \left\lceil \left[\{t(49;0{,}995) + t(49;0{,}95)\} \frac{1}{0{,}6} \right]^2 \right\rceil$$

$$\left\lceil \left[(2{,}682 + 1{,}677) \frac{1}{0{,}6} \right]^2 \right\rceil = \lceil 52{,}78 \rceil = 53$$

und das bleibt stabil, also ***n* = 53**

Minimale Wahl des Versuchsumfanges

Test H$_0$ $\mu = \mu_0$, σ^2 unbekannt.

Beispiel - exakt

Example 2.8

$$n \approx \left\lceil \left[\{t(49;0{,}995) + t(49;0{,}95)\} \frac{1}{0{,}6} \right]^2 \right\rceil$$

$$\left\lceil \left[(2{,}682 + 1{,}677) \frac{1}{0{,}6} \right]^2 \right\rceil = \lceil 52{,}78 \rceil = 53$$

Our R-program gives the solution *n* = 47 in the one-sided case a) and *n* = 53 in the two-sided case b).