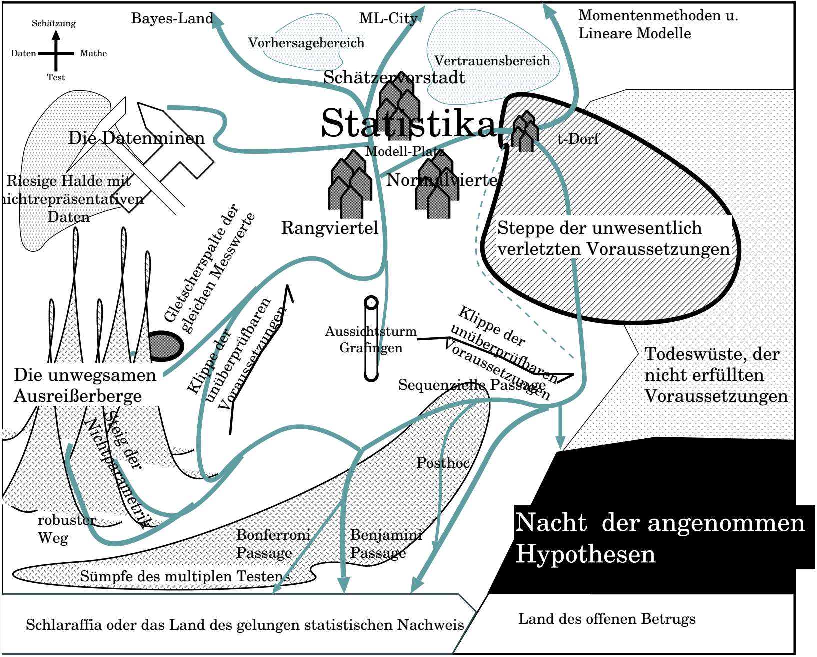
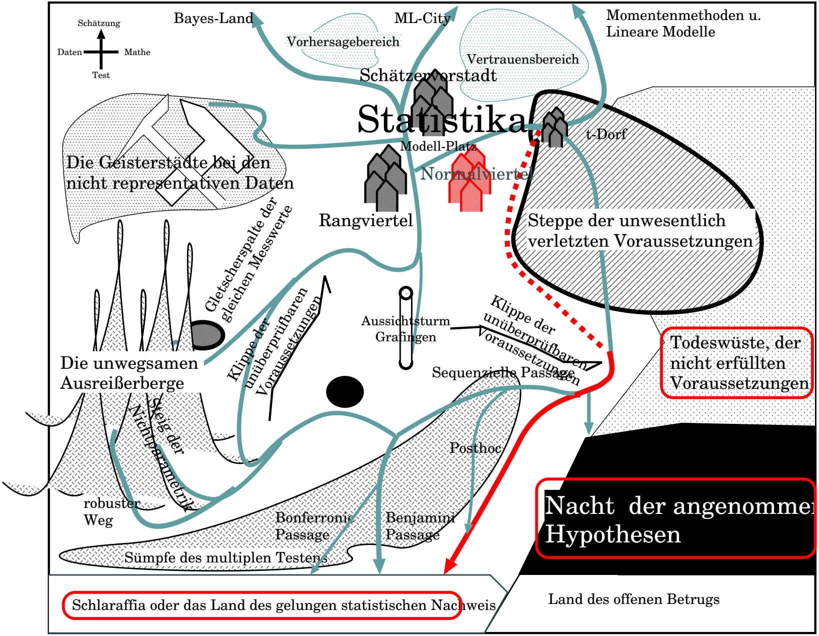
Statistik

Vorlesung 5 (Tests I)

K.Gerald van den Boogaart
http://www.stat.boogaart.de

11. November 2019


Schätzung
Daten
Mathe
Test
Bayes-Land
ML-City
Momentenmethoden u.
Lineare Modelle
Vorhersagebereich
Vertrauensbereich
Schätzervorstadt
Statistika
Die Datenminen
t-Dorf
Modell-Platz
Riesige Halde mit
nichtrepräsentativen
Daten
Normalviertel
Rangviertel
Gletscherspalte der
gleichen Messwerte
Steppe der unwesentlich
verletzten Voraussetzungen
Klippe der
unüberprüfbaren
Voraussetzungen
Aussichtsturm
Grafingen
Klippe der
unüberprüfbaren
Voraussetzungen
Todeswüste, der
nicht erfüllten
Voraussetzungen
Die unwegsamen
Ausreißerberge
Sequenzielle Passage
Steig der
Nichtparametrik
Posthoc
robuster
Weg
Bonferroni
Passage
Benjamini
Passage
Nacht der angenommen
Hypothesen
Sümpfe des multiplen Testens
Schlaraffia oder das Land des gelungen statistischen Nachweis
Land des offenen Betrugs


Schätzung
Daten
Mathe
Test
Bayes-Land
ML-City
Momentenmethoden u.
Lineare Modelle
Vorhersagebereich
Vertrauensbereich
Schätzervorstadt
Statistika
t-Dorf
Modell-Platz
Normalviertel
Die Geisterstädte bei den
nicht representativen Daten
Rangviertel
Steppe der unwesentlich
verletzten Voraussetzungen
Gletscherspalte der
gleichen Messwerte
Klippe der
unüberprüfbaren
Voraussetzungen
Aussichtsturm
Grafingen
Klippe der
unüberprüfbaren
Voraussetzungen
Todeswüste, der
nicht erfüllten
Voraussetzungen
Die unwegsamen
Ausreißerberge
Sequenzielle Passage
Steig der
Nichtparametrik
Posthoc
robuster
Weg
Bonferronie
Passage
Benjamini
Passage
Nacht der angenommen
Hypothesen
Sümpfe des multiplen Testens
Schlaraffia oder das Land des gelungen statistischen Nachweis
Land des offenen Betrugs

Vereinfachtes Beispiel

Die Situation

Eine Firma leitet mit Rimit verunreinigte Abwässer in ein Fluß ein. Gemäß der gesetzlichen Bestimmungen dürfen höchsten $7,5 ppm$ Rimit in den eingeleiteten Wässern sein.

Die Beobachtung

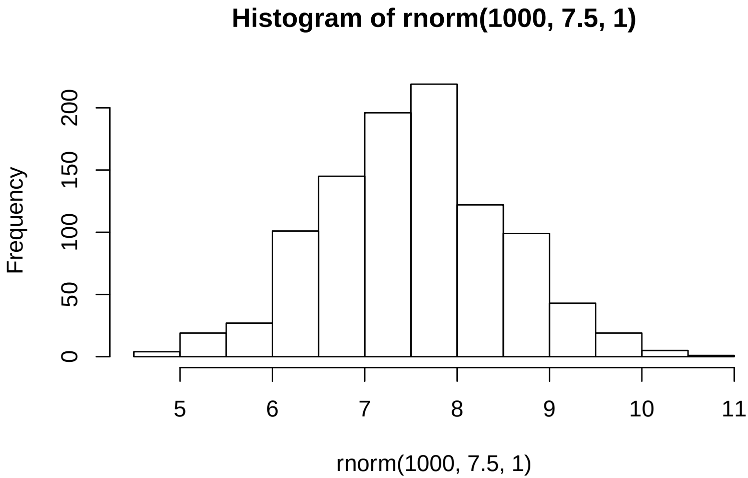
Ihre Umweltschutzorganisation entnimmt eine Probe. Das Labor bestimmt den Rimit Gehalt mit $9.25ppm$ mit einem standardisierten Verfahren mit einer Standardabweichung von $1ppm$ und einem normalverteilten Messfehler.

Das Probleme

Um das Unternehmen anzuzeigen, benötigt Ihre Umweltschutzorganisation einen wissenschaftlichen Nachweis, dass tatsächlich zu viel Rimit eingeleitet wurde.

Probemessungen im Labor

Das Labor hat 1000 Proben mit genau 7,5ppm Rimit gemessen:



Messungen streuen

Es kann also auch wenn die Firma sich korrekt verhalten hat, durch Zufall ein Wert über $9ppm$ gemessen werden.

Hypothese

Es gibt zwei Möglichkeiten:
Hypothese H_0: Die Firma leitet korrekt ein (d.h. $7,5ppm$)

vs. (gegen)

Alternative H_1: Die Firma überschreitet den Grenzwert.

Test

Wir benötigen ein wissenschaftliches Verfahren (genannt Test), welches in der Lage ist zu beweisen, dass die Firma zu viel einleitet.

$$T(Daten) = \begin{cases} 0 & \text{Hypothese möglich} \\ 1 & \text{Hypothese widerlegt} \end{cases}$$

Die Fehlerarten

	Einleitung korrekt	zu viel Rimit
$T(X) = 0$	keine Aktion richtige Entscheidung	fortgesetzte Umweltverschmutzung falsche Entscheidung Fehler 2.Art/β-Fehler
$T(X) = 1$	Unschuldige beschuldigt falsche Entscheidung Fehler 1.Art/α-Fehler	Berechtige Klage richtige Entscheidung

Nachweis auf einem α-Niveau

Vorgehensweise:
Man bestimmt eine Entscheidungsregel so, dass die Wahrscheinlichkeit für einen Fehler 1.Art, wenn die Hypothese stimmt unter α liegt.
z.B. in den Naturwissenschaften normalerweise

$$\alpha = 0.05 = 5\%$$

Leistungsfähige Tests

Ein guter Test sollte, wenn die Alternative stimmt, so wahrscheinlich wie möglich tatsächlich die Hypothese ablehnen.

Die Wahrscheinlichkeit für einen Fehler 2.Art sollt also so gering wie möglich sein.

Konstruktion eines Tests

Je größer der Messwert, deso eher wurde der Grenzwert überschritten:

$$T(X) = \begin{cases} 0, & \text{falls } X < k \\ 1, & \text{falls } X \geq k \end{cases}$$

Konstruktion eines Tests

Je größer der Messwert, deso eher wurde der Grenzwert überschritten:

$$T(X) = \begin{cases} 0, & \text{falls } X < k \\ 1, & \text{falls } X \geq k \end{cases}$$

Wähle k so, dass (wenn die Hypothese stimmt)

$$P(X > k) = \alpha = 0.05$$

Konstruktion eines Tests

Je größer der Messwert, deso eher wurde der Grenzwert überschritten:

$$T(X) = \begin{cases} 0, & \text{falls } X < k \\ 1, & \text{falls } X \geq k \end{cases}$$

Wähle k so, dass (wenn die Hypothese stimmt)

$$P(X > k) = \alpha = 0.05$$

k heißt kritischer Wert.

Konstruktion eines Tests

Je größer der Messwert, deso eher wurde der Grenzwert überschritten:

$$T(X) = \begin{cases} 0, & \text{falls } X < k \\ 1, & \text{falls } X \geq k \end{cases}$$

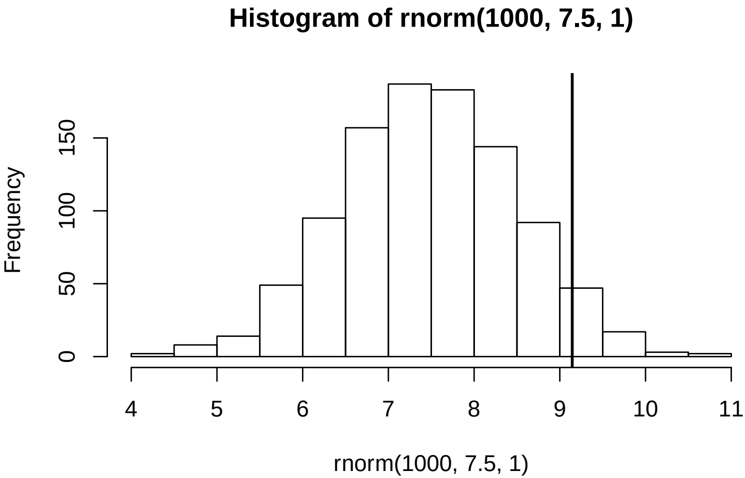
Wähle k so, dass (wenn die Hypothese stimmt)

$$P(X > k) = \alpha = 0.05$$

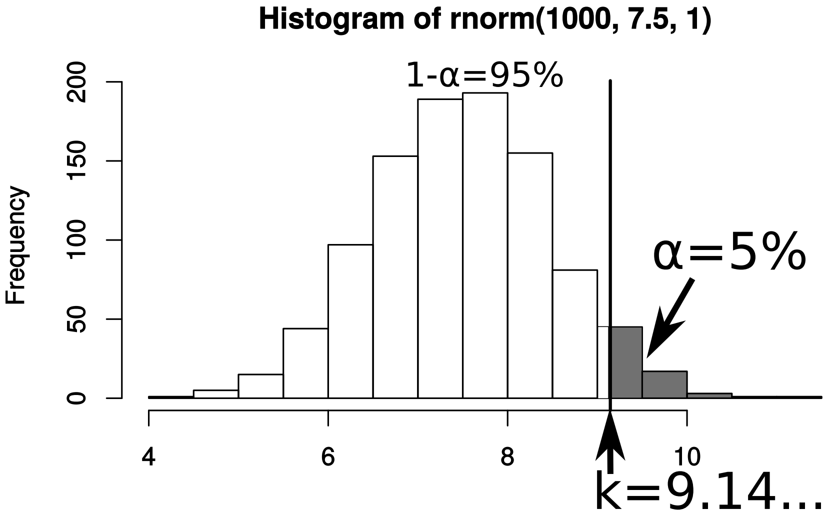
k heißt kritischer Wert.
Falls X normalverteilt mit Mittelwert 7,5 und Standardabweichung 1 so ist

$$k = F^{-1}_{N(7.5,1)}(1 - \alpha) = 9.14485362695147$$

Kritischer Wert



Kritischer Wert

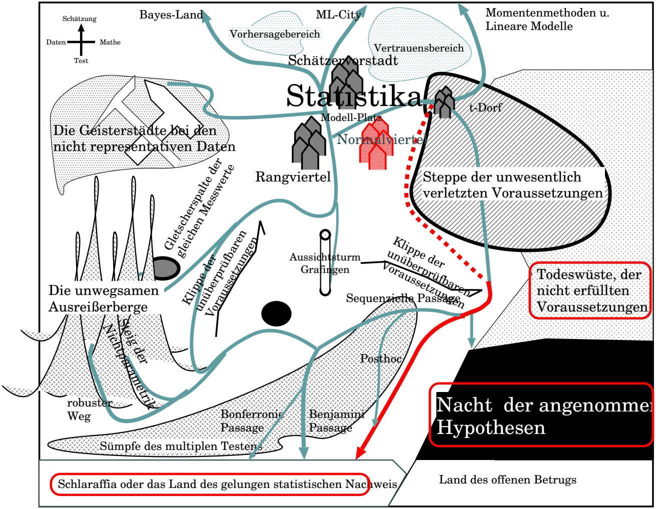


Die Umweltschutzorganisation

Es wurde $9,25 ppm$ gemessen:

$$T(9,25ppm) = \begin{cases} 0, & \text{falls } 9,25ppm < 9.14485362695147 \\ 1, & \text{falls } 9,25ppm \geq 9.14485362695147 \end{cases} = 1$$

Also ist auf dem 5%-Niveau der Nachweis erbracht, dass die Firma zu viel Rimit einleitet.


Schätzung
Daten
Mathe
Test
Bayes-Land
Vorhersagebereich
ML-City
Vertrauensbereich
Momentenmethoden u.
Lineare Modelle
Schätzervorstadt
Statistika
Modell-Platz
t-Dorf
Normalviertel
Die Geisterstädte bei den
nicht representativen Daten
Rangviertel
Steppe der unwesentlich
verletzten Voraussetzungen
Gletscherspalte der
gleichen Messwerte
Klippe der
unüberprüfbaren
Voraussetzungen
Aussichtsturm
Grafingen
Klippe der
unüberprüfbaren
Voraussetzungen
Todeswüste, der
nicht erfüllten
Voraussetzungen
Die unwegsamen
Ausreißerberge
Sequenzielle Passage
Steig der
Nichtparametrik
Posthoc
robuster
Weg
Bonferroni
Passage
Benjamini
Passage
Sümpfe des multiplen Testens
Nacht der angenommenen
Hypothesen
Schlaraffia oder das Land des gelungen statistischen Nachweis
Land des offenen Betrugs

Ein Test hat

- Namen (hier: einfacher einseitiger Gauss-Test)

Ein Test hat

- Namen
- Anwendungssituation
  Überprüfen ob der wahre Erwartungswert einen festen Wert (hier 7,5) übersteigt, wenn eine normalverteilte Messung mit bekannter Varianz σ^2 vorliegt.

Ein Test hat

- Namen
- Anwendungssituation
- Hypothese und Alternative
  $H_0 : \mu = 7,5$ vs. $H_1 : \mu > 7,5$

Ein Test hat

- Namen
- Anwendungssituation
- Hypothese und Alternative
- Voraussetzungen
  $X \sim N(\mu, \sigma^2)$

Ein Test hat

- Namen
- Anwendungssituation
- Hypothese und Alternative
- Voraussetzungen
- Ein Entscheidungsverfahren
  Meist im Computer implementiert.

α-Niveau Test

Notation:
$P(X|B)$ =Wahrscheinlichkeit von X unter der Bedingung B.

α-Niveau Test

Notation:
$P(X|B)$ =Wahrscheinlichkeit von X unter der Bedingung B.

Definition Eine Entscheidungsregel $T(X)$ mit Werten 0 oder 1 zu einer Testsituation, die besteht aus aus Voraussetzungen V, einer Hypothese H_O und einer Alternative H_1, für die unter der Voraussetzung V mit einem $\alpha \in [0, 1]$ gilt, dass

$$P(T(X) = 1|V, H_0) \leq \alpha$$

heißt ein α-**Niveau Test** für die Testsituation.

α-Niveau Test

Notation:
$P(X|B)$ =Wahrscheinlichkeit von X unter der Bedingung B.

Definition Eine Entscheidungsregel $T(X)$ mit Werten 0 oder 1 zu einer Testsituation, die besteht aus aus Voraussetzungen V, einer Hypothese H_O und einer Alternative H_1, für die unter der Voraussetzung V mit einem $\alpha \in [0, 1]$ gilt, dass

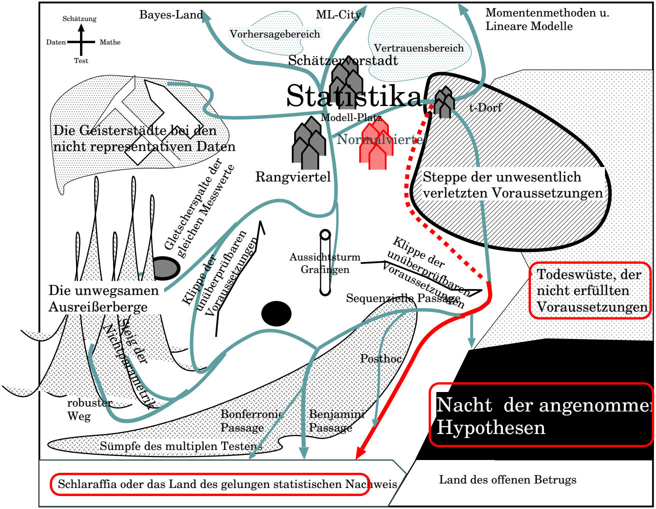
$$P(T(X) = 1|V, H_0) \leq \alpha$$

heißt ein α-**Niveau Test** für die Testsituation.

Gilt weiter:

$$P(T(X) = 1|V, H_1) \gneq \alpha$$

so heißt der α-Niveau Test **unverzerrt**.



Schätzung
Daten
Mathe
Test
Bayes-Land
Vorhersagebereich
ML-City
Vertrauensbereich
Momentenmethoden u.
Lineare Modelle
Schätzervorstadt
Statistika
t-Dorf
Modell-Platz
Normalviertel
Die Geisterstädte bei den
nicht representativen Daten
Rangviertel
Steppe der unwesentlich
verletzten Voraussetzungen
Gletscherspalte der
gleichen Messwerte
Klippe der
unüberprüfbaren
Voraussetzungen
Aussichtsturm
Grafingen
Klippe der
unüberprüfbaren
Voraussetzungen
Todeswüste, der
nicht erfüllten
Voraussetzungen
Die unwegsamen
Ausreißerberge
Sequenzielle Passage
Steig der
Nichtparametrik
Posthoc
robuster
Weg
Bonferroni
Passage
Benjamini
Passage
Nacht der angenommen
Hypothesen
Sümpfe des multiplen Testens
Schlaraffia oder das Land des gelungen statistischen Nachweis
Land des offenen Betrugs

Multiples Testen

Werden mehrer Tests durchgeführt, so entsteht eine hohe Wahrscheinlichkeit dafür, dass auch richtige Hypothesen abgelehnt werden.
Erläuterung an der Tafel

Bonferroni

Lösungsmöglichkeit von Bonferroni:
Will man insgesamt ein α-Niveau einhalten, so wählt man bei der Durchführung von n Tests für jeden einzelnen Test ein α-Niveau von $\tilde{\alpha} := \frac{\alpha}{n}$
Erläuterung an der Tafel

p-Werte

Werden statistische Tests am Computer durchgeführt, so wird statt einer Entscheidung ein p-Wert ausgegeben:
Der p-Wert ist das kleinste α-Niveau zu dem der Test gerade noch ableht:
Wichtig!!!!, Wichtig!!!!, Wichtig!!!!

Test lehnt die Hypothese ab, genau dann wenn $p \leq \alpha$

Wichtig!!!!, Wichtig!!!!, Wichtig!!!!