Der statistische Test –
Ein Randomisierungstest

Werner Stahel, Seminar für Statistik

Probevorlesung für Schüler, 13. Sept. 2006

1. Statistik

- Volkszählung, Erhebung von volkswirtschaftlich oder gesellschaftlich wichtigen Grössen
  $\longrightarrow$ Öffentliche Statistik
  Betriebswirtsch: Unternehmenserfolg, Arbeitszeiterfass'g
  $\longrightarrow$ Wirtschaftsstatistik

- Modellierung des Zufalls $\longrightarrow$ Wahrscheinlichkeitsmodelle
  „Schliessende" Statistik =
  Zusammenhang zwischen Modellen und Daten

Der wichtigste Begriff des Zusammenhangs (klass. Statistik):
Der statistische Test.

Eine einfach verständl. Klasse von Tests: Randomisierungstests

2. Einführendes Beispiel

a Hagel-Experiment: („Grossversuch IV" im Napfgebiet 1978-1983)
Verringert „Impfen" von potent. Hagelwolken mit Silberiodid die Hagelenergie?

Zielgrösse: Hagelenergie, gemessen für n Wolken
Zwei Gruppen: ca. $n/2$ „geimpft", Rest „Kontrolle".

$$Y_i \quad : \quad \text{Hagelenergie der Wolke } i$$
$$G_i = \begin{cases} 1 & \text{falls Wolke } i \text{ geimpft,} \\ 0 & \text{sonst.} \end{cases}$$

Hoffnung: Y_i mit $G_i = 1$ fallen tendenziell niedriger aus.

2.

b Beobachtet:

$Y_i = y_i^*$	16672	25	855	0	152	0	46	1219
$G_i = g_i^*$	1	1	0	0	0	1	1	0

(In Wirklichkeit 216 Wolken; davon wurden 94 geimpft.)

Hagelenergie „zufällig"!!

Wirkung des Impfens „im Mittel"?

Mittelwert (arithmetischer M.):

ungeimpft	556.50
geimpft	4185.75

Schadet das Impfen?

(Mittelwert „geimpft" ohne 1. Beobachtung: 23.67)

2.

c Verlässliche Antwort muss „Zufälligkeit”
der beobachteten Wolken einbeziehen
$\longrightarrow$ Wahrscheinlichkeits-Modell!!

d Grundidee eines statistischen Tests:
Wahrscheinlichkeitsmodell für die „Hypothese”,
dass das Impfen keine Wirkung hat.
Wie gross können dann die Unterschiede
zwischen geimpften und ungeimpften werden?
Modell für „rein zufällige Unterschiede”.
Kann der beobachtete Unterschied so zustande kommen?
Wenn nein, schliessen wir auf eine Wirkung des Impfens.
($\longrightarrow$ Widerspruchsbeweis!)

2.

e **Ähnliche Situationen**

- Medizin: Wirkung von Medikamenten
- Technik: neues Verfahren
- Landwirtschaft: Düngung, Sorte, ...
- Umwelt: Schutzmassnahme
- ...

$\longrightarrow$ Vergleich von 2 Stichproben

3. Statistische Überlegung

a Nullhypothese = Wahrscheinlichkeitsmodell.
(Üblich in der Statistik: Vert. für Y_i; $G_i = g_i^*$ fest vorgegeben.
Randomisierungstests: G_i zufällig; $Y_i = y_i^*$ als fest betrachtet.)

Grundidee:
Falls das Impfen keinen Einfluss auf die Hagelenergie hat,
hätten wir die genau gleichen Werte y_i^* erhalten,
wenn die Wolken entsprechend $\underline{g}^{(1)} = [0,1,0,0,1,1,0,1]$
(statt $\underline{g}^{(0)} = [1,1,0,0,0,1,1,0]$)
oder entsprechend irgendeiner anderen Auswahl
geimpft worden wären.

3.

b Zufallsauswahl:

Jede Auswahl von $n/2 = 4$ geimpften Wolken aus $n = 8$ hat(te) gleiche Wahrscheinlichkeit

$$p = \binom{8}{4}^{-1} = \frac{1}{70}$$

Damit ist das W.modell der Nullhypothese festgelegt.

c Wir „hoffen" eigentlich, dass die Nullhypothese nicht gilt, sondern eine „Alternative".

In „welcher Richtung" erwarten wir Abweichungen?

Mittlere Hagelenergie für geimpfte kleiner als für ungeimpfte!

Allgemein: Teststatistik soll extreme Werte annehmen, wenn Alternative gilt.

Vorschlag: Differenz der Mittelwerte für ungeimpfte und geimpfte Wolken:

$$T\langle \underline{G}, \underline{y}^* \rangle = \frac{1}{n/2} \sum_{i:G_i=0} y_i^* - \frac{1}{n/2} \sum_{i:G_i=1} y_i^* .$$

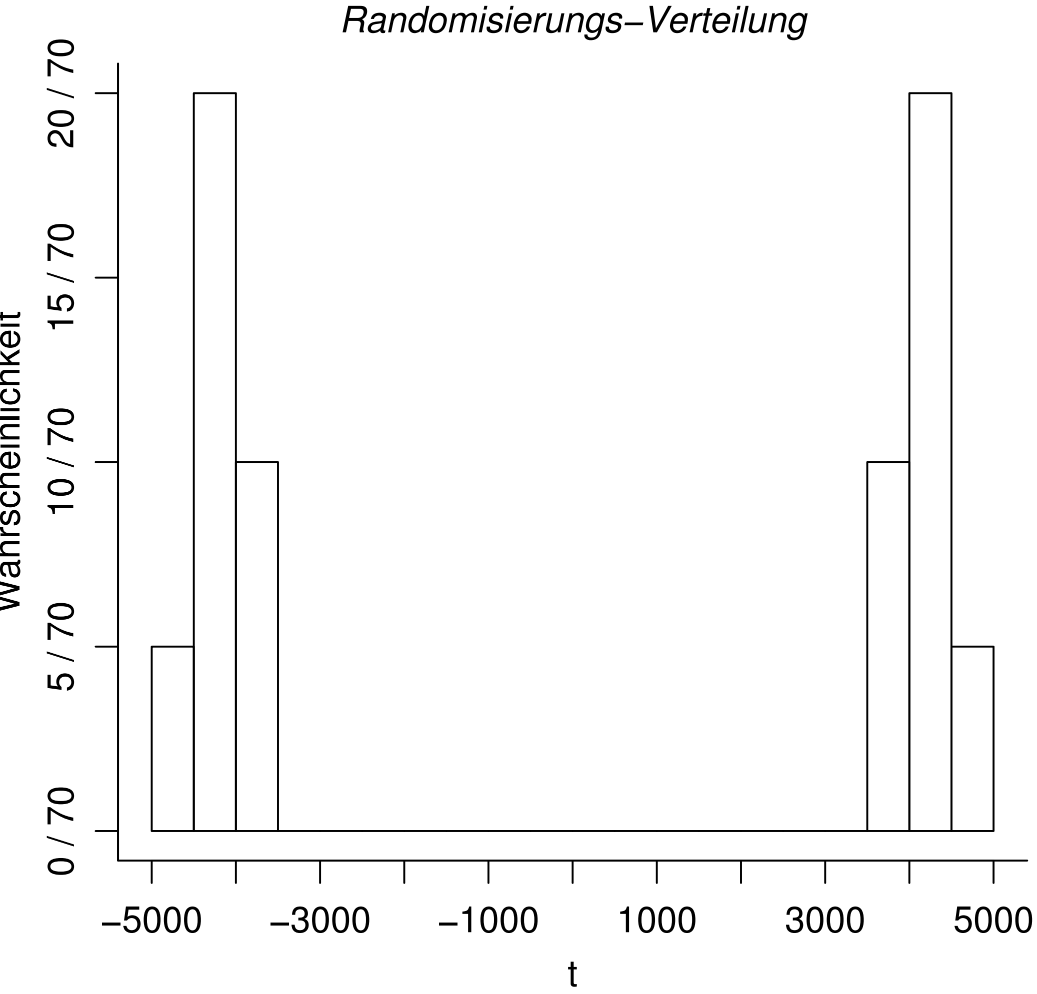
d Wie ist T unter der Nullhypothese „verteilt?”

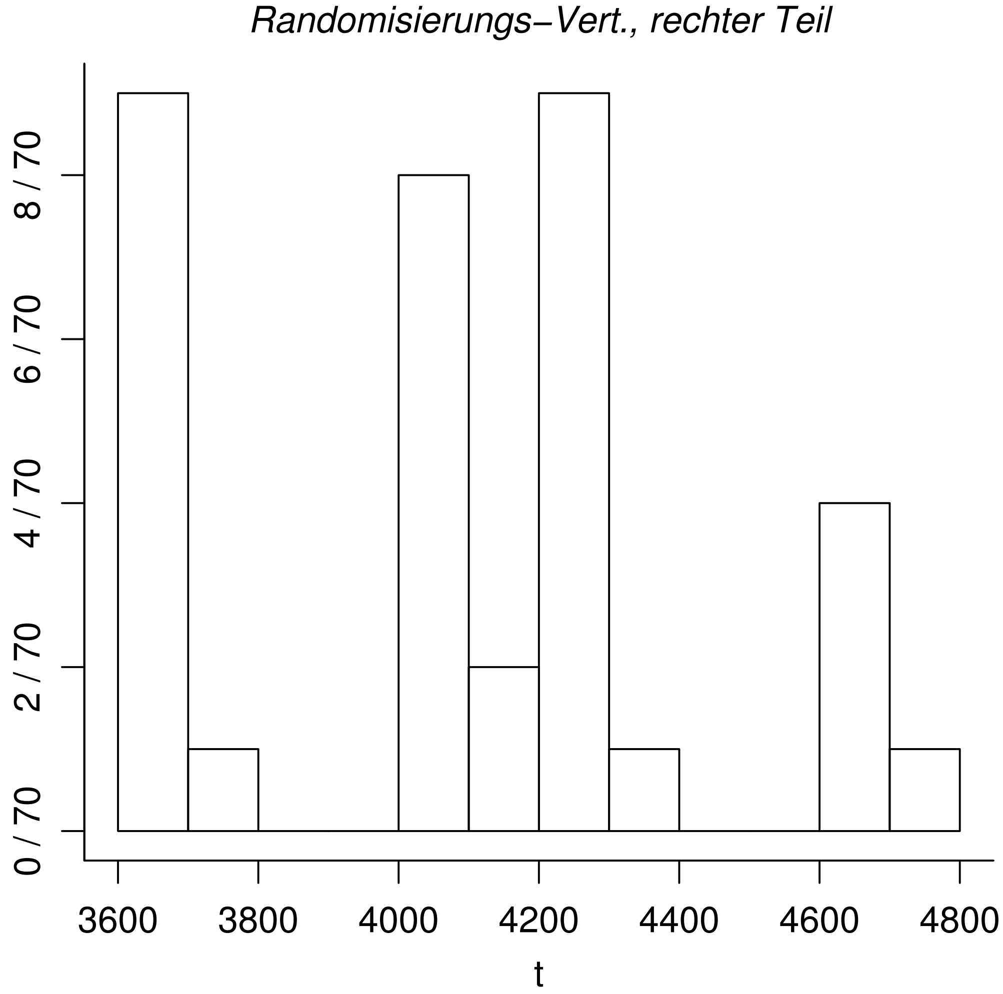
(Welcher Wert für T kommt mit welcher W. heraus?)

$y_1^*, \ldots, y_n^*$ gegeben $\longrightarrow \leq \binom{n}{n/2}$ mögliche Werte für T.

$$P\langle T\langle \underline{G}, \underline{y}^* \rangle = t \rangle = \frac{\#\{\underline{g} \mid T\langle \underline{g}, \underline{y}^* \rangle = t\}}{\binom{n}{n/2}}$$

„Randomisierungs-Verteilung”





Randomisierungs-Verteilung im Beispiel. Links: ganze Verteilung, rechts: Verteilung der positiven Werte (Ausschnitt aus der Figur links)

3.

e $\binom{8}{4} = 70$ Werte sind möglich. Auch der extremste ist möglich $\longrightarrow$ keine sicheren Schlüsse möglich.

Konvention: Der Effekt (des Impfens) gilt als nachgewiesen oder statistisch signifikant, wenn das beobachtete Ergebnis unwahrscheinlich ist, falls die Nullhypothese gilt.

$\longrightarrow$ Bestimme einen Bereich von „extremen" Werten der Teststatistik – so, dass seine Wahrscheinlichkeit unter der Nullhypothese = 5% ist (oder 1% oder ...)
$\longrightarrow$ „Verwerfungsbereich"

Beispiel: $\{t \mid t \geq 4643.25\}$.

3.

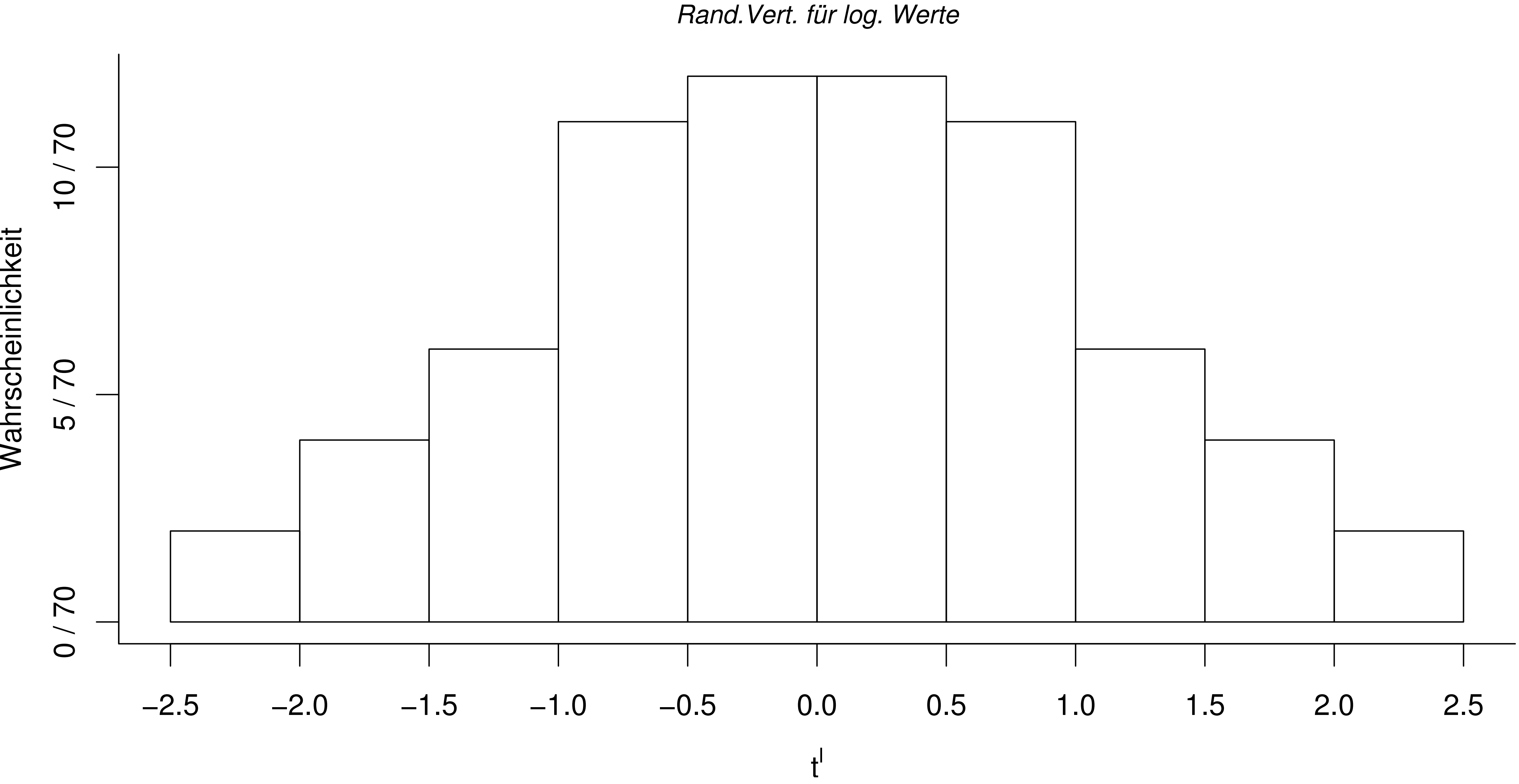
f Im Experiment:

$$\begin{aligned} T\langle \underline{g}^*, \underline{y}^* \rangle &= \frac{1}{4}(855 + 0 + 152 + 1219) \\ &\quad -\frac{1}{4}(16672 + 25 + 0 + 46) \\ &= -3629.25 \end{aligned}$$

Effekt in die unerwartete Richtung!
Nullhypothese nicht verworfen; Effekt nicht nachgewiesen.
(Auch nicht in umgekehrter Richtung.)

g Beliebige Teststatistik.
Differenz der Mittelwerte reagiert vor allem auf die grösste Hagelenergie! (=unrobust)

Andere Möglichkeit für gleiche Problemstellung:
Logarithmus-Transformation, dann Mittelwertsdifferenz.



Randomisierungs-Verteilung für die Mittelwerts-Differenz von logarithmierten Daten im Beispiel

3.

h Resultat des Hagel-Experiments (216 Wolken)

Kein Effekt.

Genauere Analyse: Erste Rakete hat positiven Effekt,
aber nachher kommt es eher schlimmer.

Geschätzter Effekt zeigt in die „falsche" Richtung,
aber (auch) nicht signifikant!

4. Bemerkungen zu Statistik und Mathematik

a Schliessende Statistik allgemein:
Modelle – auch komplexe – für „zufällige" Phänomene
mit Daten in Verbindung bringen.
Modelle auf Grund von Daten entwickeln!

Beispiele:

- Wie wird das Auftreten von Pilzen durch das Wetter beeinflusst?
- Kann man Glukose im Blut bestimmen ohne Einstich?
  Sensoren auf der Haut messen Leitfähigkeit, Temp., ...
  Rückschluss auf Glukose genügend genau & zuverlässig?

4.

b Statistik ist angewandte Mathematik.

- Theorie: Wahrscheinlichkeits-Theorie, math. Statistik
- Anwendung: Datenanalyse, Modell-Entwicklung, Interpretation

Gefragt: Systematisches Denken, Kreativität, Kommunikation

Im Studium: Abstraktion zuerst, Anwendung später.

Merkpunkte Schliessende Statistik: Statistischer Test

- Problem: Daten zeigen Effekt (Unterschied zw. Gruppen)
  Kann der beob. Effekt „rein zufällig zustande kommen?”

- Vorgehen:
  – W.-Modell für die Nullhypothese„kein Effekt”,
  – Wahl der Test-Statistik (soll Effekt messen),
  – Verteilung der Test-Statistik unter der Nullhypothese,
  – „signifikant”, wenn extremer Wert beobachtet wurde.

- Kein exakter Schluss (Beweis) möglich.
  Konvention nötig, was „extrem” heissen soll.

- Schliessende Statistik stellt
  Brücke zwischen Modellen und Daten her.
  Teil der angewandten („unreinen”!) Mathematik.

`www.stat.math.ethz.ch/~stahel/courses/resampling/`

`stat.ethz.ch` ⟶ People ⟶ Werner Stahel ⟶ homepage
⟶ Courses: Resampling-Methoden