Vorlesung 12b

Kann das Zufall sein?

Beispiele von statistischen Tests

1. Fishers exakter Test

“Passen die Verhältnisse in den Rahmen?”

(vgl. Buch S. 130/131

Sir Ronald Fisher

1890-1962

Aus einer Urne mit 80 roten und 87 blauen Kugeln

wurden 113 Kugeln entnommen.

40 davon waren rot, und 73 waren blau.

Passt das zur Hypothese, dass die Kugeln

rein zufällig gezogen wurden?

Stimmen die Verhältnisse einigermaßen,

oder fallen sie aus dem Rahmen?

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Unter den 113 gezogenen Kugeln erwartet man

ähnliche Verhältnisse wie in der gesamten Urne:

80: 167 für rot, 87 : 167 für blau.

Tatsächlich ergab sich in der Stichprobe

für rot ein **sehr** unterdurchschnittliches Ergebnis !

Wie lässt sich das quantifizieren?

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Unter des Hypothese des rein zufälligen Ziehens

ist die Anzahl X der gezogenen roten Kugeln

hypergeometrisch verteilt

mit Parametern $n = 113$, $g = 167$, $r = 80$.

Dafür ergibt sich:

$$\mathbf{E}[X] = n \cdot \frac{r}{g} = 54.1 \ .$$

$$g = 167,\ r = 80,\ n = 113$$

X ist Hyp(n, g, r)-verteilt

$$\mathbf{E}[X] = n \cdot \frac{r}{g} = 54.1\ .$$

Zur Erinnerung:

$$\mathbf{P}(X = k) = \frac{\binom{r}{k}\binom{g-r}{n-k}}{\binom{g}{n}}\,, \quad k = 0, 1, \ldots, n\ .$$

	gezogen	nicht gezogen	Summe
rot	**40**	40	80
blau	73	14	87
Summe	113	54	167

Die Wahrscheinlichkeit, ein Ergebnis zu erhalten,

das mindestens so weit von 54 weg ist

wie der beobachtete Wert 40, ist

$$\mathbf{P}\big(|X - 54| \geq |40 - 54|\big)$$

$$= \mathbf{P}(X \leq 40) + \mathbf{P}(X \geq 68)$$

$$= 5.57 \cdot 10^{-6} \, .$$

$$\mathbf{P}\Big(|X - 54| \geq |40 - 54|\Big) = 5.6 \cdot 10^{-6}$$

Was bedeutet das?

Fazit: Angenommen die Hypothese trifft zu.
Dann tritt ein Ergebnis, das so extrem ist wie das beobachtete, gerade 6 mal in einer Million auf.
Damit wird die Hypothese mehr als fragwürdig.

Man nennt die berechnete Wahrscheinlichkeit

den zu den Daten gehörigen p-Wert

oder auch das *beobachtete Signifikanzniveau*,

zu dem die Hypothese abgelehnt wird.

Wie passt unser Urnen-Beispiel in die Welt?

Es geht um die Fragestellung

“Passen die Proportionen

– oder sollte man an der Hypothese der reinen Zufälligkeit zweifeln?”

Zwei Verpackungen einer Botschaft – und eine Frage dazu:

A. Die sanfte Therapiemethode T1 brachte
in nicht weniger als 30% der Fälle keinen Heilungserfolg,
wohingegen die harte Therapiemethode T2
in immerhin 80 % der Fälle erfolgreich war.

B. Sogar die harte Therapiemethode T2 brachte
in nicht weniger als 20% der Fälle keinen Heilungserfolg,
wohingegen die sanfte Therapiemethode T1
in immerhin 70 % der Fälle erfolgreich war.

Welche Therapiemethode würden Sie (als Arzt) bevorzugen?

Von insgesamt 167 Ärzten

wurden rein zufällig 80 ausgewählt,

denen die Botschaft in der Form A vermittelt wurde,

die restlichen 87 bekamen die Botschaft in der Form B.

Jeder der Ärzte hatte sich daraufhin für die Bevorzugung

einer der beiden Therapiemethoden zu entscheiden.

Das Ergebnis war:

	für Methode T1	für Methode T2	Summe
A	40	40	80
B	73	14	87
Summe	113	54	167

Für das Testen der Hypothese

“Die Verpackung der Botschaft

hat keinen Einfluss auf die Entscheidung”

eignet sich das eingangs besprochene Urnenmodell.

Unter dieser Hypothese

kommt die Aufteilung der $80 + 87$ Formulare

auf die 113 Befürworter von T1

und die 54 Befürworter von T2

rein zufällig zustande.

So gesehen kann das Ergebnis “wohl kaum Zufall sein”:

unter unserer Hypothese tritt ein Ausgang,

der so extrem ist wie der beobachtete,

gerade mal 6 mal in einer Million auf.

Eine weitere Möglichkeit zum Testen der Hypothese

“Zwei Verhältnisse sind gleich”

bietet die *Normalapproximation*.

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

Anteilsschätzung über die Normalapproximation:

X := Anzahl roter Kugeln bei $n = 113$ Zügen

$$\mathbf{E}[X] = np$$

mit

$$p = \frac{80}{167}$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

$H = \frac{X}{n}$ ist approximativ normalverteilt

(trotz der schwachen Abhängigkeiten beim Ziehen ohne Zurücklegen,
vgl. die beiden letzten Folien von Vorlesung 7b)

$$\mu_H = p = \frac{80}{167},$$
$$\sigma_H^2 = \frac{p(1-p)}{n} \frac{g-n}{g-1},$$
$$\sigma_H = \sqrt{0.022 \cdot 0.326} = 0.0268$$

	gezogen	nicht gezogen	Summe
rot	40	40	80
blau	73	14	87
Summe	113	54	167

$$Z := \frac{H - \mu_H}{\sigma_H}$$ ist approximativ N$(0,1)$-verteilt

Der beobachtete Wert von Z war $z = -4.67$.

$$\mathbf{P}(|Z| > 4.67) = 3 \cdot 10^{-6}$$

ist hier der p-Wert, zu dem die Hypothese abgeleht wird.

2. z-Test und t-Test

“Kann *diese* Verschiebung des Mittelwertes Zufall sein?”

n reelle Messwerte $x_1, \ldots, x_n$ haben den Mittelwert m

(alles gemessen auf einer bestimmten Skala.)

Unterscheidet sich der beobachtete Mittelwert m signifikant von einem hypothetischen “Populationsmittelwert” μ_0?

Eine erste Auskunft gibt ein Vergleich

des Unterschiedes $|m - \mu_0|$

mit dem *Standardfehler*

$$f := s/\sqrt{n}.$$

Wir können fragen:

Um weches Vielfache des Standardfehlers

unterscheidet sich m von μ_0?

Dies erhält seinen theoretischen Unterbau

durch die goldene Idee der Statistik

(*man fasse die* x_i *auf als Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen* X_i *mit Erwartungswert* μ *und Varianz* σ^2)

und den Zentralen Grenzwertsatz:

Für große n *ist der Stichprobenmittelwert* $\bar{X}$ *approximativ* N$\left(\mu, \frac{\sigma^2}{n}\right)$*-verteilt*.

Für große n ist

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{approximativ} \quad \mathrm{N}(0, 1)\text{-verteilt.}$$

Bei bekanntem σ sei $z := \dfrac{m - \mu_0}{\sigma/\sqrt{n}}$.

Unter der Hypothese $\mu = \mu_0$ ist

$$\mathbf{P}(|\bar{X} - \mu_0| \geq |m - \mu_0|) \approx \mathbf{P}(|Z| \geq |z|),$$

mit $\mathrm{N}(0, 1)$-verteiltem Z.

In der Praxis ist σ meist unbekannt.

Aber:

Für große n ist s mit großer W'keit nahe bei σ.

Also ist für große n

$$T := \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \text{approximativ} \quad \mathrm{N}(0, 1)\text{-verteilt.}$$

Sei $t := \dfrac{m - \mu_0}{s/\sqrt{n}}$.

Unter der Hypothese $\mu = \mu_0$ ist

$$\mathbf{P}(|T| \geq |t|) \approx \mathbf{P}(|Z| \geq |t|),$$

mit $\mathrm{N}(0, 1)$-verteiltem Z.

Und was kann man bei kleinem n sagen?

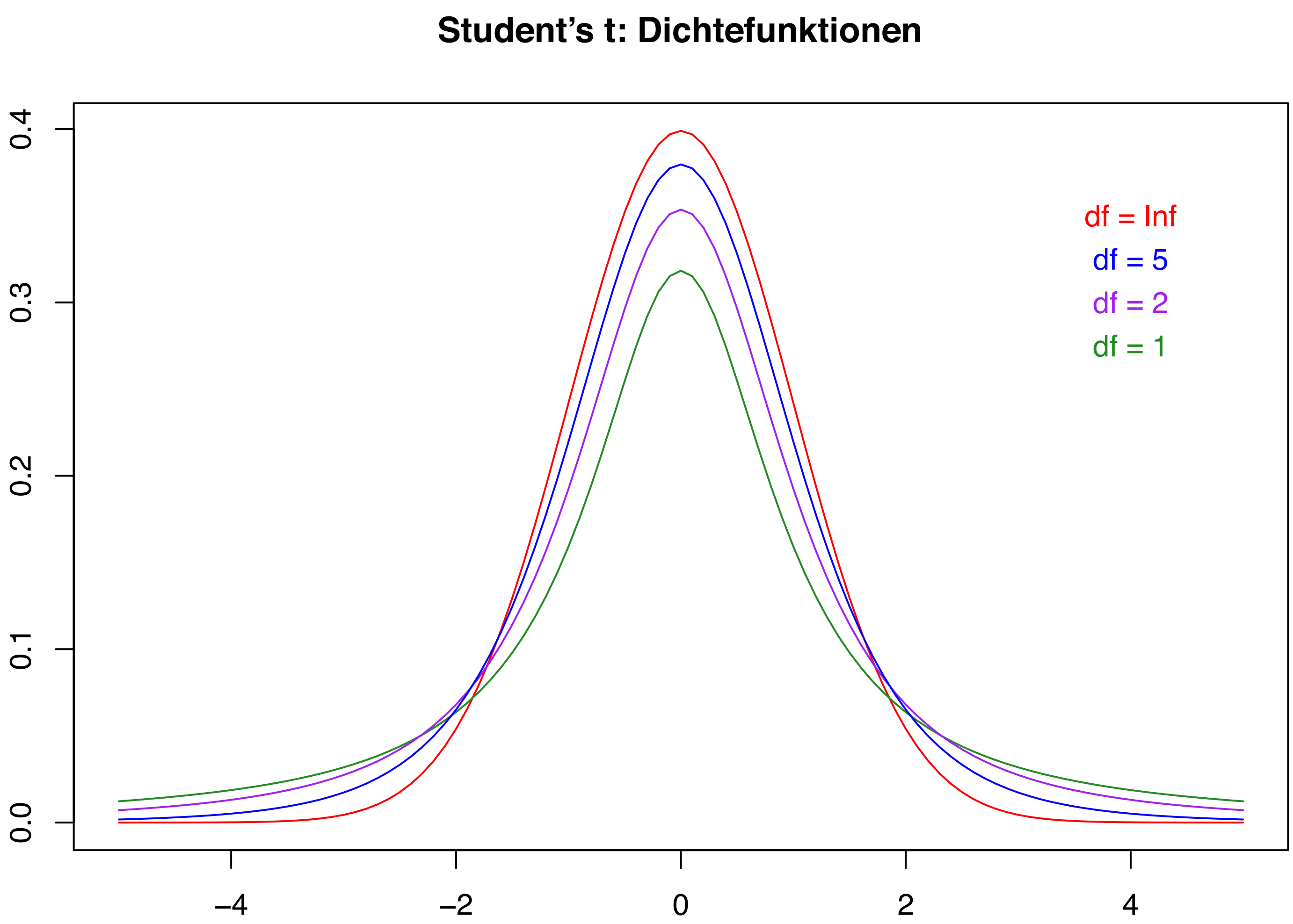
Unter der zusätzlichen Modellannahme

$X_1, \ldots, X_n$ sind (unabhängig und) $N(\mu, \sigma^2)$-verteilt

ist $T = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$ so verteilt wie

$T_{n-1} :=$ eine student-verteilte Zufallsvariable
mit $n - 1$ Freiheitsgraden (degrees of freedom, df))

Diese Verteilung kennt R gut (Befehl für Verteilungsfunktion: `pt(q,df)`);
sie lässt sich aus der Rotationssymmetrie der n-dimensionalen
Standard-Normalverteilung auch gut verstehen (VL 7a , Buch Seite 132).



Dichten von T_{df}

Beispiel:

Ist für $n = 16$ der "Wert der t-Statistik"

$$t = \frac{m - \mu_0}{s/\sqrt{n}}$$

glech 2.5, dann ergibt sich

$$\mathbf{P}(|T_{15}| \geq 2.5) = 2(1 - \mathrm{pt}(2.5, 15)) = 0.025.$$

Man spricht vom p-Wert für die Ablehnung

der Hypothese $\mu = \mu_0$ zugunsten der Alternative $\mu \neq \mu_0$.

Oft gibt man sich ein *Signifikanzniveau* α vor.

Wenn der p-Wert kleiner als α ist, sagt man: Die Hypothese $\mu = \mu_0$ kann zugunsten der Alternative $\mu \neq \mu_0$ zum Niveau α abgelehnt werden.

Populär ist die Wahl $\alpha = 0.05$:

Wenn der p-Wert kleiner als 0.05 ist, sagt man auch kurz: m ist (nach dem t-Test) signifikant von μ_0 verschieden.