
Was bedeutet $p < 0.05$?

Oliver Kuß

Institut für Medizinische Epidemiologie, Biometrie und Informatik,
Universität Halle-Wittenberg,
Magdeburger Str. 27, 06097 Halle (Saale)
Oliver.Kuss@medizin.uni-halle.de

Programm:

1. Einleitung
2. Der diagnostische Test
3. Der statistische Test
4. Das Problem
5. Lösungen (?)
6. Fazit/Diskussion

1. Einleitung

Was Ihnen erspart bleibt:

Thema meines Anschubanspruchs: (FKZ: 5/08)

Regressionsanalyse mit diskreten, nicht-binären Zielgrößen und korrelierten Beobachtungen zur Analyse biometrischer und epidemiologischer Daten

$$\ln\left(\frac{\pi_{ijr}}{\pi_{ijR}}\right) = \theta_r + \phi_r \beta x'_{ij} + u_{ir}$$

Modellgleichung für ein Stereotype Regressionsmodell mit gemischten Effekten zur Modellierung von korrelierten diskreten, nicht-binären Zielgrößen

Was Ihnen nicht erspart bleibt:

Bedingte Wahrscheinlichkeiten:

$P(A | B)$: Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn wir wissen, dass B wahr ist. („Wahrscheinlichkeit für A gegeben B“)

Beispiel: Verdecktes Ziehen einer Karte aus einem Skatblatt (32 Karten)

Ereignis A: Karte ist ein Herz $\rightarrow P(A) = 8/32 = 1/4$

Ereignis B: Karte ist ein As $\rightarrow P(B) = 4/32 = 1/8$

$P(A | B)$: Wahrscheinlichkeit für ein Herz, wenn wir wissen, dass die Karte ein As ist $\rightarrow P(A | B) = 1/4$

$P(B | A)$: Wahrscheinlichkeit für ein As, wenn wir wissen, dass die Karte ein Herz ist $\rightarrow P(B | A) = 1/8$

Im allgemeinen gilt: $P(A | B) \neq P(B | A)$ (und auch $\neq P(A \cap B)$)

2. Der diagnostische Test

		Krankheit	
		K+	K-
Diagnostischer Test	T+	RP	FP
	T-	FN	RN

Maßzahlen für die Güte eines diagnostischen Tests:

Sensitivität: $P(T+ | K+)$

Wahrscheinlichkeit für positives Testergebnis bei einem Kranken

Spezifität: $P(T- | K-)$

Wahrscheinlichkeit für negatives Testergebnis bei einem Gesunden

Aber:

Diese Werte nützen uns am Krankenbett nichts, denn da haben wir nur das Testergebnis (T+ bzw. T-), wollen aber wissen, ob die Krankheit vorliegt (K+ bzw. K-)

Von Interesse also: $P(K+ | T+)$ bzw. $P(K- | T-)$ (Prädiktive Werte)

Satz von Bayes:

$$P(K+ | T+) = \frac{P(T+ | K+) P(K+)}{P(T+ | K+) P(K+) + P(T+ | K-) P(K-)}$$
$$= \frac{\text{Sensitivität x Prävalenz}}{\text{Sensitivität x Prävalenz} + (1\text{-Spezifität}) \times (1\text{-Prävalenz})}$$

Problem: Prädiktive Werte sind prävalenzabhängig!

Beispiel: Bundesweites Screening auf HIV-Virus (mit ELISA-Test)

Sensitivität: $P(T+ | K+) = 99.5\%$

Spezifität: $P(T- | K-) = 99.5\%$

Prävalenz: $P(K+) = 0.1\%$

Wahrscheinlichkeit, dass jemand mit positivem HIV-Test tatsächlich HIV-positiv ist:

$$P(K+ | T+) = \frac{P(T+ | K+) P(K+)}{P(T+ | K+) P(K+) + P(T+ | K-) P(K-)}$$
$$= \frac{0.995 \times 0.001}{0.995 \times 0.001 + 0.005 \times 0.999} = 16.6\%$$

Von 6 Test-Positiven ist nur einer tatsächlich HIV-positiv!!!

3. Der statistische Test

		Nullhypothese (Wahrheit)	
		+	-
Statist. Test	ST+ (H0 akzeptieren)	(H0 richtig) 😊	(H0 falsch) Fehler 2. Art
	ST- (H0 verwerfen)	Fehler 1. Art	😊

Fehler 1. Art: $\alpha = P(\text{ST-} \mid H_0+)$

Wahrscheinlichkeit für Verwerfen von H_0 , wenn Sie in Wirklichkeit vorliegt

Fehler 2. Art: $\beta = P(\text{ST+} \mid H_0-)$

Wahrscheinlichkeit für Annahme von H_0 , wenn Sie in Wirklichkeit nicht vorliegt

Durchführung eines statistischen Tests:

Lege α vor Beginn des Experimentes fest (z.B. $\alpha = 0.05$)

Nach der Durchführung des Experiments (Daten: x) berechnen wir eine Teststatistik $T(x)$ und entscheiden:

Verwerfe H_0 , falls $P(T > T(x) \mid H_0) < \alpha$

$P(T > T(x) \mid H_0)$ ist der p-Wert.

Der p-Wert ist die Wahrscheinlichkeit dafür, dass, unter der Annahme dass die Nullhypothese richtig ist, die Teststatistik ein Wert annimmt, der noch extremer als der beobachtete Wert ist.

4. Das Problem

Der p-Wert macht eine Aussage über die Wahrscheinlichkeit der Daten, nicht über die Wahrscheinlichkeit der Hypothese!!!

Wir möchten $P(H_0 \mid \text{Daten})$, bekommen aber $P(\text{Daten} \mid H_0)$!!!

**Lösung (analog zum diagnostischen Test):
Anwendung des Satz von Bayes:**

$$P(H_{0+} \mid \text{Daten}) \approx \frac{P(\text{Daten} \mid H_{0+}) P(H_{0+})}{P(\text{Daten} \mid H_{0+}) P(H_{0+}) + P(\text{Daten} \mid H_{0-}) P(H_{0-})}$$

Großes Problem: Wie groß ist $P(H_{0+})$? (A priori - Wahrscheinlichkeit für die Nullhypothese)

5. Lösungen (?)

Beispiel: Randomisierte klinische Studie mit zwei Therapien (Verum/Placebo) und binärer Zielgröße (Heilung ja/nein)

	Heilung		
	Ja	Nein	
Verum	58	42	100
Placebo	42	58	100
	100	100	200

Ergebnis: OR [95%-KI]: 1.91 [1.09; 3.34]
 $\chi^2 = 5.12 \Rightarrow$ p-Wert = 0.024

Mögliche Lösung:

Nullhypothese und Alternative sind gleichwahrscheinlich ($P(H_0+) = 0.5$)

Für unsere Beispielstudie gilt dann:

$$P(H_0+ \mid \text{Daten}) = 0.135$$

Ist die Annahme $P(H_0^+) = 0.5$ realistisch?

Verband forschender Arzneimittelhersteller (2003):
Für ein Medikament, das zugelassen wird, werden 4.7 Substanzen in der Phase I getestet.

Unter der Annahme, dass ein Arzneimittel erst in die Phase I kommt, wenn die Firma auch sicher ist, dass es einen Effekt hat, wäre hier: $P(H_0^+) = 0.78$

Für unsere Beispielstudie würde dann gelten: $P(H_0^+ | \text{Daten}) = 0.354$

Zur Erinnerung: Der p-Wert war 0.024!

Spannende Frage:

Wie groß muss $P(H_0+)$ sein, damit der p-Wert gleich $P(H_0+ | \text{Daten})$ ist?

Antwort (für unser Beispiel): $P(H_0+) = 13\%$

Berechtigte Frage: Ist es überhaupt ethisch, eine Studie zu machen, wenn man sich zu 87% sicher ist, dass die Therapie einen Effekt hat?

Weitere Lösungen (ohne Angabe von $P(H_0+)$!)

- **Minimum Bayes Factor** (Goodman, 1999):

$$P(H_0+ \mid \text{Daten}) = \frac{\exp(-\chi^2/2)}{(1+\exp(-\chi^2/2))}$$

In unserem Beispiel: $P(H_0+ \mid \text{Daten}) = 0.071$

- **Calibration of p-values** (Sellke/Bayarri/Berger, 1999):

Für p-Wert $< 1/e$ (= 0.368) gilt:

$$P(H_0+ \mid \text{Daten}) = \frac{-e \cdot \text{p-Wert} \cdot \log(\text{p-Wert})}{(1 - e \cdot \text{p-Wert} \cdot \log(\text{p-Wert}))}$$

In unserem Beispiel: $P(H_0+ \mid \text{Daten}) = 0.194$

6. Fazit/Diskussion

- **Der p-Wert sagt uns nicht das, was wir eigentlich wissen wollen, denn er macht keine Aussage über die Sicherheit der zugrunde liegenden Hypothesen.**
- **Die Sicherheit, die uns der p-Wert vorgaukelt, ist im allgemeinen zu groß.**
- **Wie sollen wir (die klinische Forschung) mit diesem Problem umgehen?**

Für's Unter-der-Decke-Halten ist es bereits zu spät:



THEMEN DER ZEIT

Klinische Studien und Statistik

Von der Wahrscheinlichkeit des Irrtums

Wolfgang Weihe

Ein „Galileischer Dialog“ über ein statistisches Problem

Das nachfolgend wiedergegebene Gespräch findet auf der Terrasse eines großen internationalen Hotels in Venedig statt. Teilnehmer sind drei Ärzte, die sich anlässlich eines Kongresses hier getroffen hatten. Sie diskutierten über die Be-

frage, inwieweit uns klinische Studien helfen können, für unsere Patienten die richtige Therapie zu finden.

Sagredo: Genau. Du hattest die unglaubliche Behauptung aufgestellt, klinische Studien in der Form, wie sie derzeit durchgeführt werden, könnten

der Wahl für die Verlässlichkeit der Ergebnisse. Nichts könnte klarer und deutlicher sein.

Sagredo: Ich muss Simplicio Recht geben. Du willst sicher nur sagen, dass Studien schlecht geplant oder schlampig durchgeführt werden oder dass man

Deutsches Ärzteblatt, 101(13), 26.3.2004, 834-838

Komplettes Abschaffen von p-Werten und nur noch Berichten von Konfidenzintervallen?

- Spricht auch über die Größe des Effektes.
- Führt den p-Wert aber über die Hintertür wieder ein.

Kompletter Umstieg in die Bayes'sche Statistik?

- A priori – Wahrscheinlichkeiten ($P(H_0+)$) müssen tatsächlich vorher spezifiziert werden! (Vorschlag von Spiegelhalter: Spezifizierung von 3 A priori – Wahrscheinlichkeiten: skeptisch, optimistisch, realistisch)
- Approximative Methoden (Minimum Bayes Factor, Calibration) sind erste Schritte in die richtige Richtung.