

**Querschnittsbereich 1:
Epidemiologie, Medizinische Biometrie und
Medizinische Informatik**

- Übungsmaterial -

Erstellt von Mitarbeitern des IMISE und des ZKS Leipzig

**2. Übung
Diagnostik**

© Universität Leipzig
WS 2018/19

**Ein erfundenes Beispiel
(H.P.Beck-Bornholt und H.-H.Dubben)**

- In einem wenig erschlossenen exotischen Land ist die seltene Krankheit K (canine overhoe) endemisch. Sie haben dort Ihren Urlaub verbracht. Die Ansteckungsgefahr für Touristen ist zwar gering, dennoch entschließen Sie sich, bei einem Spezialisten einen Test auf Infektion mit K durchführen zu lassen, da die Heilungschancen bei einer Früherkennung deutlich besser sind als nach dem Ausbruch der Krankheit.
- Der Test fiel positiv aus, und nun sind Sie in Sorge!
- Welche Informationen werden Sie sich besorgen, um die Wertigkeit des Testergebnisses zu beurteilen?

Ein erfundenes Beispiel (2)

Auf Ihre Anfrage erfahren Sie

zum Test:

- Bei 99 von 100 Infizierten spricht der Test an.
- Auch bei 2 von 100 Nichtinfizierten spricht der Test an.

zur Infektionshäufigkeit:

- Unter Touristen wird von 1000 einer infiziert.

Ein erfundenes Beispiel (3)

Bei Ihnen fiel der Test positiv aus.

Wie groß ist die Wahrscheinlichkeit, dass Sie sich tatsächlich infiziert haben?

A	99%	
B	98%	
C	etwa 95%	
D	etwa 50%	
E	etwa 5%	
F	2%	
G	1%	

Ein erfundenes Beispiel (4)

Wie können wir die richtige Antwort einfach ermitteln?

1. Schritt: Konstruktion einer fiktiven Population von Touristen, von denen jeder 1000. infiziert ist
2. Schritt: Berechnung der Anzahl der Testpositiven unter Infizierten und Nichtinfizierten
3. Schritt: Berechnung der Anzahl der Testnegativen
4. Schritt: Berechnung der Wahrscheinlichkeit, bei positivem Testausgang tatsächlich infiziert zu sein

Ein erfundenes Beispiel (6)

	Test negativ (T-)	Test positiv (T+)	Zeilensumme
Nicht infiziert (K-)			
Infiziert (K+)			
Spaltensumme			

Unter Testpositiven sind Infizierte zu erwarten. Damit ist die gesuchte Wahrscheinlichkeit, bei positivem Testausgang tatsächlich infiziert zu sein, gegeben durch $P(K+ | T+) = \frac{\dots}{\dots} =$ also%.

Sie können sich also fühlen.

Irgendwie unbefriedigend ist die Situation dennoch.

Ein erfundenes Beispiel (7)

Wie sicher ist ein negativer Testausgang in diesem Beispiel?

	Test negativ (T-)	Test positiv (T+)	Zeilensumme
Nicht infiziert (K-)			
Infiziert (K+)			
Spaltensumme			

Unter Testnegativen ist nur ein Infizierter zu erwarten. Damit ist die gesuchte Wahrscheinlichkeit, bei **negativem** Testausgang tatsächlich **nicht** infiziert zu sein, gegeben durch $P(K- | T-) = \frac{\dots}{\dots} \approx$ also%.

Bei einem negativen Testausgang kann man sich also fühlen.

Ein erfundenes Beispiel (8)

PPW und NPW hängen stark von der Prävalenz ab.

Für Sensitivität = 99% und Spezifität = 98% ergibt sich zum Beispiel für den PPW die folgende Tabelle:

Prävalenz	PPW
0,01%	0,5%
0,1%	4,7%
1%	33,3%
10%	84,6%
20%	92,5%

Was ist ein „Diagnostischer Test“ ?

Ein „Diagnostischer Test“ ist ein Verfahren, durch das für einen Patienten aufgrund des **Vorhandenseins oder Nichtvorhandenseins eines Symptoms** oder Symptomkomplexes entschieden wird, ob der Patient an einer bestimmten Erkrankung **erkrankt oder nicht erkrankt** ist.

Symptome (Krankheitszeichen) sind bestimmte Ausprägungen eines (direkt oder indirekt) beobachtbaren Merkmals, z.B.

- Schmerzen, z.B. stechender Schmerz im linken Oberbauch
- Körpertemperatur, z.B. $> 39^{\circ}\text{C}$ (starkes Fieber)
- Veränderungen im Röntgenbild des Thorax, z.B. Schatten

Korrektheit diagnostischer Tests

Da Symptome nie gleichzeitig vollkommen sensitiv und vollkommen spezifisch sind, treten Fehlklassifikationen auf:

		Test	
		positiv (T+)	negativ (T-)
Patient	krank (K+)	 <i>richtig positiv (RP)</i>	falsch negativ (FN) 
	gesund (K-)	falsch positiv (FP) 	 <i>richtig negativ (RN)</i>

Güte eines diagnostischen Tests

Die Güte eines diagnostischen Tests wird durch die beiden Parameter Sensitivität und Spezifität charakterisiert:

- **Sensitivität** $P(T+/K+)$
Wahrscheinlichkeit für ein positives Ergebnis des Testes T unter der Voraussetzung, dass die Krankheit K vorliegt
- **Spezifität** $P(T-/K-)$
Wahrscheinlichkeit für ein negatives Ergebnis des Testes T unter der Voraussetzung, dass die Krankheit K nicht vorliegt

Sensitivität und Spezifität

Fieber bei Pneumonie

- Fieber tritt bei Pneumonie regelmäßig auf, d.h. Fieber **sensitiv** bzgl. Pneumonie
- Fieber häufig bei vielen Krankheiten, d.h. Fieber **nicht spezifisch** für Pneumonie

Zahnschmerz bei Karies

- Karies kann lange ohne Zahnschmerz bleiben, d.h. Zahnschmerz **nicht sensitiv** bzgl. Karies
- Zahnschmerz oft Anzeichen für Karies und selten für anderes, d.h. Zahnschmerz **spezifisch** für Karies

Bestimmung der Güte eines diagnostischen Tests

Voraussetzung zur Bestimmung von Sensitivität und Spezifität eines diagnostischen Tests ist die Kenntnis der „korrekten Diagnose“.

Hierfür ist ein so genannter „**Goldstandard**“ erforderlich.

		Test		Σ
		T+	T-	
Krankheit (Goldstandard)	K+	RP	FN	RP+FN
	K-	FP	RN	RN+FP
Σ		RP+FP	RN+FN	N

$$Sens = \frac{RP}{RP + FN}$$

$$Spez = \frac{RN}{RN + FP}$$

Wertigkeit eines Testergebnisses

Bei der **Anwendung diagnostischer Tests** will man nicht vom Krankheitszustand auf das Testergebnis schließen, sondern es interessiert der umgekehrte **Schluss vom Testergebnis auf den Krankheitszustand**:

- **Positiver Prädiktiver Wert (PPW) $P(K+/T+)$**
Wahrscheinlichkeit „erkrankt“ zu sein, wenn Test positiv
- **Negativer Prädiktiver Wert (NPW) $P(K-/T-)$**
Wahrscheinlichkeit „gesund“ zu sein, wenn Test negativ

Wertigkeit eines Testergebnisses

Bei der **Bewertung eines Testergebnisses** ist außer der Güte des Tests auch die **Prävalenz** der Erkrankung in der Population, der der Patient angehört, zu berücksichtigen.

Informationen über die Population, der ein Patient angehört, erhält man im Wesentlichen durch die **Anamnese**.

Bestimmung der Wertigkeit eines Testergebnisses

Entspricht der Anteil der Erkrankten $RP+FN$ der Prävalenz der Erkrankung in der Population, in der der Test eingesetzt werden soll, so können die Vorhersagewerte wie unten angegeben bestimmt werden:

		Test		Σ
		T+	T-	
Krankheit (Goldstandard)	K+	RP	FN	RP+FN
	K-	FP	RN	RN+FP
Σ		RP+FP	RN+FN	N

$$PPW = \frac{RP}{RP + FP}$$

$$NPW = \frac{RN}{RN + FN}$$

Bestimmung der Wertigkeit eines Testergebnisses

Entspricht der Anteil der Erkrankten $RP+FN$ in der Untersuchungspopulation nicht der Prävalenz der Erkrankung in der Population, in der der Test eingesetzt werden soll, so müssen die Vorhersagewerte über die Bayessche Formel bestimmt werden:

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

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

Beispiel 1 für diagnostische Tests (1)

Das Auftreten des Allels ϵ_4 in dem für das Apolipoprotein E codierenden Gen APOE ist streng mit dem Auftreten von Alzheimer assoziiert. Um den diagnostischen Wert des Auftretens des Allels ϵ_4 im Vergleich zu den klinischen Kriterien für Alzheimer zu überprüfen, wurden an 2188 Patienten aus Zentren für Demenzerkrankungen die pathologische Diagnose als Goldstandard mit der klinischen Diagnose entsprechend standardisierter Kriterien und dem Auftreten des Allels ϵ_4 im APOE-Gen verglichen.

(Mayeux et al., NEJM 1998)

Beispiel 1 für diagnostische Tests (2)

- Von den 2188 untersuchten Patienten wurden bei der pathologischen Untersuchung 1770 Fälle mit Alzheimer identifiziert. Die restlichen Fälle waren Patienten mit anderen Demenzerkrankungen. Die klinische Diagnose Alzheimer stimmte bei 1643 Patienten mit der pathologischen Diagnose überein. Bei 228 Patienten waren weder die Kriterien für die pathologische Diagnose Alzheimer noch für die klinische Diagnose Alzheimer erfüllt.
- Unter den pathologisch gesicherten Fällen mit Alzheimer konnte bei 628 kein $\epsilon 4$ Allel im APOE-Gen gefunden werden. Dagegen fand man bei den Fällen, bei denen pathologisch das Vorliegen von Alzheimer ausgeschlossen werden konnte, 133 Patienten die das $\epsilon 4$ Allel hetero- oder homozygot besaßen.

Beispiel 1 für diagnostische Tests (3)

- Stellen Sie für beide Kriterien (klinische Diagnose und Auftreten des $\epsilon 4$ Allels im APOE-Gen) die den Zusammenhang zwischen diesen Kriterien und der pathologisch gesicherten Diagnose beschreibenden 2×2 – Tabellen auf.
- Berechnen Sie Sensitivität und Spezifität beider Kriterien.
- Berechnen Sie positiven und negativen prädiktiven Wert beider Kriterien. Was ist bei der Anwendung dieser Vorhersagewerte zu beachten?

Beispiel 1 für diagnostische Tests (4)

Güte und Wertigkeit der klinischen Kriterien

Sens =

Spez =

PPW =

NPW =

Beispiel 1 für diagnostische Tests (5)

Güte und Wertigkeit des genetischen Kriteriums

Sens =

Spez =

PPW =

NPW =

Beispiel 1 für diagnostische Tests (6)

- Die errechneten Vorhersagewerte gelten nur für Populationen von Demenzerkrankten, wie sie in den der Untersuchung zu Grunde liegenden Zentren für Demenzerkrankungen vorliegen. Denn die Tabellen, aus denen die Vorhersagewerte errechnet wurden, enthalten die Prävalenz von Alzheimererkrankten in diesen Populationen.
- Entspricht die Prävalenz der Demenzerkrankten in der Population, in der der Test eingesetzt werden soll, nicht der in den Tabellen der vorliegenden Studie, so müssen die Vorhersagewerte über die Bayessche Formel berechnet werden.

Beispiel 1 für diagnostische Tests (7)

- Um die Spezifität zu erhöhen, muss die Zahl der falsch positiven Testergebnisse verringert werden:

$$Spez = \frac{RN}{RN + FP}$$

- Dies kann erreicht werden, wenn die Diagnose Alzheimer nur für die Fälle ausgesprochen wird, bei denen sowohl die klinischen Kriterien als auch das genetische Kriterium für Alzheimer sprechen.
- D.h. eine klinische Diagnose Alzheimer sollte durch eine Untersuchung des APOE-Gens abgesichert werden. Wird bei der genetischen Untersuchung kein $\epsilon 4$ Allel gefunden, so wird von der Diagnose Alzheimer Abstand genommen.

Beispiel 2 für diagnostische Tests (1)

- Eine Lebensmittelaufsichtsbehörde untersuchte 400 Joghurtbecher, die in Geschäften aufgrund des überschrittenen Mindesthaltbarkeitsdatums sichergestellt wurden.
- Bei 80 Bechern war der Deckel gewölbt. Nach dem Öffnen prüfte man den Inhalt auf Zeichen der Ungenießbarkeit (oberflächliche Kulturen, auffallender Geruch, eigenartige Färbung, unangenehmer Geschmack). Dabei wurden 20 Portionen als genießbar eingestuft.
- Bei der Prüfung des Inhalts der Becher ohne Deckelwölbung nach den selben Kriterien wurden 140 Proben für ungenießbar befunden.

Beispiel 2 für diagnostische Tests (2)

- a) Stellen Sie eine Kreuztabelle mit Häufigkeitsangaben auf, aus der die oben beschriebene Beziehung zwischen diagnostischen Befunden und tatsächlichem Zustand hervorgeht.
- b) Berechnen Sie aus den Daten Prävalenz, Sensitivität sowie positiven prädiktiven Wert (PPW).
- c) An welchem der Parameter Prävalenz, Sensitivität, Spezifität, PPW, NPW und welcher Ausprägung (möglichst hoch / niedrig) sind folgende Personen interessiert?
 - I. Ein Händler mit verdorbener Ware, von der er möglichst viel verkaufen will.
 - II. Ein Kunde, der seinen Joghurt nach äußerlicher Inspektion des Bechers kauft.

Beispiel 2 für diagnostische Tests (3)

a)

b) Prävalenz =
Sensitivität =
PPW =

c) (I)
(II)