

# Přednáška II.

## Vztah pravděpodobnosti, statistiky a biostatistiky

- ➡ Statistika vychází z pravděpodobnosti
- ➡ Podmíněná pravděpodobnost, Bayesův vzorec
- ➡ Senzitivita, specificita, prediktivní hodnoty
- ➡ Frekventistická a Bayesovská statistika



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Opakování – klíčové principy biostatistiky



# Statistika, biostatistika a analýza dat

## Statistika

- Primárně je zaměřena na vývoj metod a algoritmů pro řešení teoretických problémů.
- Nicméně i statistika je vždy primárně motivována reálnými problémy.
- Vychází z teorie pravděpodobnosti.

## Biostatistika

- Propojení znalosti statistických metod a dané problematiky v řešení biologických a klinických úloh.
- Na prvním místě není teoretický vývoj, ale aplikace.

## Analýza dat

- Velmi obecná oblast bez jasné definice.
- Prostupuje různými odvětvími.
- Zahrnuje komplexní postupy hodnocení dat (čištění, kódování).
- Nemusí být založena na statistice.

# Biostatistika vychází ze statistiky

- ➡ Biostatistika je **aplikace statistických metod** v řešení biologických a klinických problémů.
- ➡ Snahou je **získat z pozorovaných dat užitečnou informaci**. V popředí zájmu je pozorovaná variabilita mezi studovanými subjekty, kterou chceme vysvětlit.

# Statistický pohled na problém

- ➔ **Cílová populace** – chceme postihnout konkrétní problém.
- ➔ Získáme **experimentální vzorek** cílové populace (pozorování), která převedeme na **číselné vyjádření** (data). Vzorek by měl být reprezentativní a náhodný.
- ➔ Předpokládáme **pravděpodobnostní chování** (model) tohoto vzorku (tedy i cílové populace).
- ➔ Konkrétní problém vyjádříme ve vybraném modelu jako **hypotézu**.
- ➔ **Zhodnotíme hypotézu** na základě vybraného modelu a pozorovaných dat.



# Statistika vychází z pravděpodobnosti

- ➡ Teorie pravděpodobnosti se zabývá **modelováním náhody**.
- ➡ Lze nějak ale vyjádřit, co je to náhoda?

# Statistika vychází z pravděpodobnosti

- ➡ Teorie pravděpodobnosti se zabývá **modelováním náhody**.
- ➡ Lze nějak ale vyjádřit, co je to náhoda?

Objektivní nepředvídatelnost?

Nedostatek informací?

# Statistika vychází z pravděpodobnosti

- ➡ Teorie pravděpodobnosti se zabývá **modelováním náhody**.
- ➡ Lze nějak ale vyjádřit, co je to náhoda?

Objektivní nepředvídatelnost?

**Nedostatek informací?**

*“Chance is only ignorance of the connections between phenomena.”*

Pierre Simon de Laplace

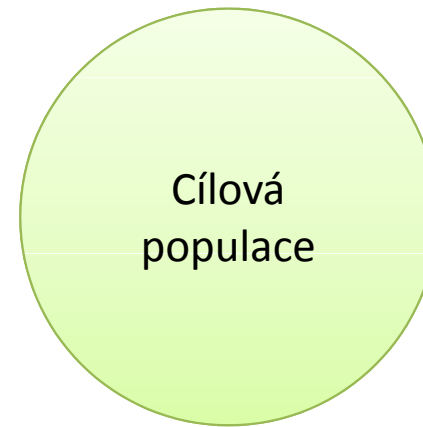


# Statistika vs. pravděpodobnost

**Statistika**

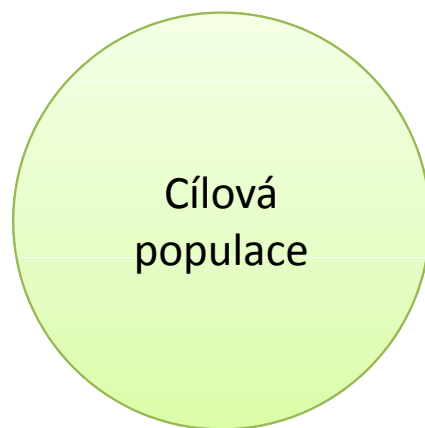


**Pravděpodobnost**



# Statistika vs. pravděpodobnost

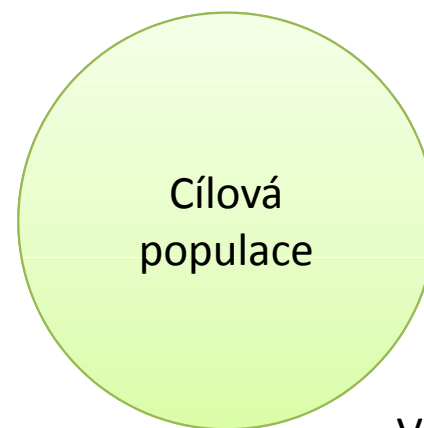
## Statistika



Cílem statistiky je získání informace o cílové populaci na základě pozorovaného experimentálního vzorku.



## Pravděpodobnost



V teorii pravděpodobnosti se ptáme na pravděpodobnost získání konkrétního výsledku, máme-li danou strukturu cílové populace.



# Značení

- ➔ **Základní prostor** ( $\Omega$ ) – množina všech možných výsledků experimentu
- ➔ **Elementární jev** ( $\omega$ ) – konkrétní výsledek experimentu
- ➔ **Náhodný jev** ( $A$ ) – podmnožina základního prostoru
- ➔ **Množina všech jevů** ( $\mathcal{A}$ ) – množina (všech) podmnožin základního prostoru
- ➔  $\emptyset$  představuje jev nemožný,  $\Omega$  zase jev jistý
- ➔ Množinové operace mají v teorii pravděpodobnosti svůj význam:
  1.  $\omega \in A$  - jev  $A$  nastane, když nastane  $\omega$
  2.  $\omega \notin A$  - jev  $A$  nenastane, když nastane  $\omega$
  3.  $A \subset B$  - nastání jevu  $A$  implikuje nastání jevu  $B$
  4.  $A \cap B$  - nastání jevu  $A$  a zároveň jevu  $B$
  5.  $A \cup B$  - nastání jevu  $A$  nebo jevu  $B$
  6.  $A \cap B = \emptyset$  - jevy  $A$  a  $B$  se navzájem vylučují, jsou *disjunktní*
  7.  $A^c$  - nastání jevu opačného k jevu  $A$

# Pravděpodobnost

➡ Pravděpodobnost lze definovat jako funkci, která přiřazuje náhodnému jevu reálné číslo mezi 0 a 1. Je to tedy funkce  $P: \mathbf{A} \rightarrow [0,1]$ . Musí platit následující:

$$1. P(\emptyset) = 0, P(\Omega) = 1$$

$$2. A \subseteq \Omega \Rightarrow 0 \leq P(A) \leq 1$$

$$3. A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

# Definice pravděpodobnosti

➔ **Klasická definice pravděpodobnosti:** předpokládáme, že  $\Omega$  je konečná a všechny  $\omega$  jsou stejně pravděpodobné. Pak

$$A \subseteq \Omega \Rightarrow P(A) = \frac{|A|}{|\Omega|}$$

kde  $|A|$  je počet prvků množiny  $A$  (počet elementárních jevů jevu  $A$ ).

➔ **Axiomatická definice pravděpodobnosti:**  $\Omega$  je libovolná množina elementárních jevů,  $\mathbf{A}'$  je množina měřitelných jevů ( $\mathbf{A}'$  je podmnožina  $\mathbf{A}$ ).  
Funkce  $P: \mathbf{A}' \rightarrow [0,1]$ , která splňuje

$$1. P(\Omega) = 1$$

$$2. A_1, A_2, \dots \in \mathbf{A}': A_i \cap A_j = \emptyset, \forall i \neq j \Rightarrow P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

se nazývá pravděpodobnost. Trojice  $(\Omega, \mathbf{A}', P)$  se nazývá pravděpodobnostní prostor.

# Definice pravděpodobnosti – najděte rozdíly

➔ **Klasická definice pravděpodobnosti:** předpokládáme, že  $\Omega$  je **konečná** a všechny  $\omega$  jsou **stejně pravděpodobné**. Pak

$$A \subseteq \Omega \Rightarrow P(A) = \frac{|A|}{|\Omega|}$$

kde  $|A|$  je počet prvků množiny  $A$  (počet elementárních jevů jevu  $A$ ).

➔ **Axiomatická definice pravděpodobnosti:**  $\Omega$  je **libovolná** množina elementárních jevů,  $\mathbf{A}'$  je množina měřitelných jevů ( $\mathbf{A}'$  je podmnožina  $\mathbf{A}$ ). Funkce  $P: \mathbf{A}' \rightarrow [0,1]$ , která splňuje

$$1. P(\Omega) = 1$$

$$2. A_1, A_2, \dots \in \mathbf{A}': A_i \cap A_j = \emptyset, \forall i \neq j \Rightarrow P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

se nazývá pravděpodobnost. Trojice  $(\Omega, \mathbf{A}', P)$  se nazývá pravděpodobnostní prostor.

# Co to znamená?

- ➔ Axiomatická definice připouští i **nespočetný základní prostor**, tedy nespočetnou množinu elementárních jevů.
- ➔ **Příklady**: hod kostkou × měření výšky lidské postavy
  
- ➔ Axiomatická definice připouští **různou pravděpodobnost** různých **elementárních jevů**.
- ➔ **Příklady**: hod kostkou × měření výšky lidské postavy

# Nezávislost jevů

➡ Dva jevy  $A$  a  $B$  jsou nezávislé právě tehdy, když platí

$$P(A \cap B) = P(A)P(B)$$

➡ Jsou-li dva jevy  $A$  a  $B$  nezávislé, pak i

$A^c$  je nezávislé na  $B$

$A$  je nezávislé na  $B^c$

$A^c$  je nezávislé na  $B^c$



# Nezávislost jevů

➡ Dva jevy  $A$  a  $B$  jsou nezávislé právě tehdy, když platí

$$P(A \cap B) = P(A)P(B)$$

➡ Jsou-li dva jevy  $A$  a  $B$  nezávislé, pak i

$A^c$  je nezávislé na  $B$

$A$  je nezávislé na  $B^c$

$A^c$  je nezávislé na  $B^c$

➡ **Příklad:** Uvažujme opět hod kostkou a jevy  $A = \{1, 3, 5\}$  a  $B = \{4, 5, 6\}$ .

$$P(A \cap B) = 1/6 \neq 1/4 = P(A)P(B)$$

➡ Jevy  $A$  a  $B$  tedy nejsou nezávislé.

# Podmíněná pravděpodobnost

➡ Máme-li jev  $B$  s pravděpodobností  $P(B) > 0$ , pak podmíněnou pravděpodobnost jevu  $A$  za podmínky nastoupení jevu  $B$  definujeme jako

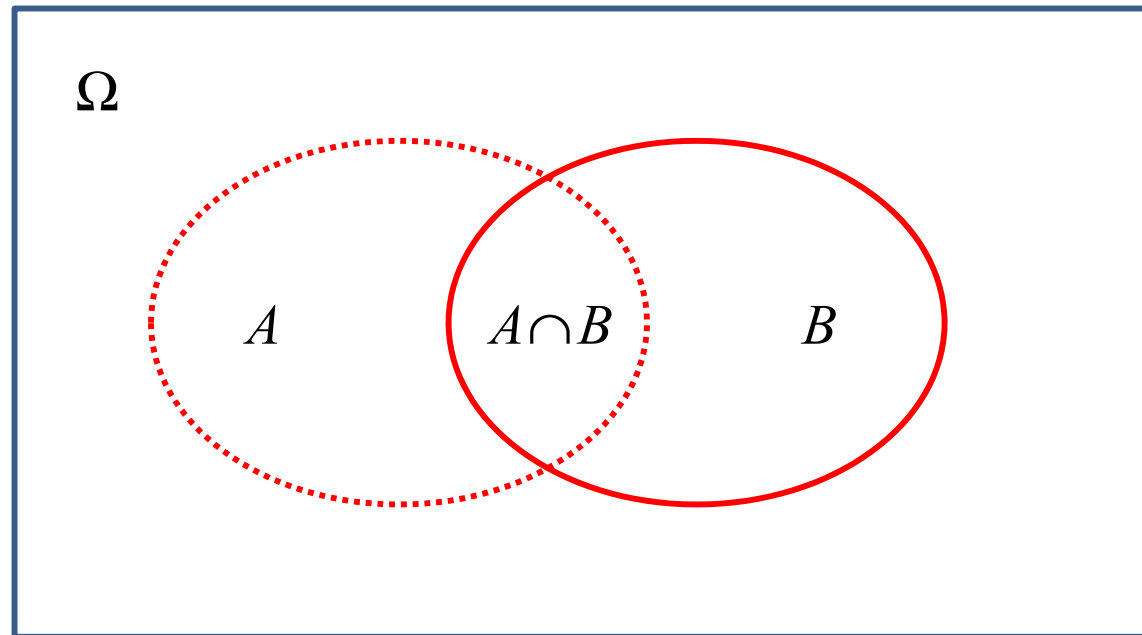
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

➡ Pro nezávislé jevy  $A$  a  $B$  platí

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

# Podmíněná pravděpodobnost

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



# Podmíněná pravděpodobnost

➡ **Příklad:** Osoba  $X$  má všechny typické příznaky chřipky. Pravděpodobnost, že se jedná o klasickou chřipku je  $0,7$  (jev  $A$ ), prasečí chřipku  $0,2$  (jev  $B$ ), ptačí chřipku  $0,05$  (jev  $C$ ) a dosud neznámou formu  $0,05$  (jev  $D$ ). Diagnostický test prokázal, že klasická chřipka to není. Jaká je nyní pravděpodobnost, že se jedná o novou formu chřipky?

# Podmíněná pravděpodobnost

➔ **Příklad:** Osoba X má všechny typické příznaky chřipky. Pravděpodobnost, že se jedná o klasickou chřipku je 0,7 (jev  $A$ ), prasečí chřipku 0,2 (jev  $B$ ), ptačí chřipku 0,05 (jev  $C$ ) a dosud neznámou formu 0,05 (jev  $D$ ). Diagnostický test prokázal, že klasická chřipka to není. Jaká je nyní pravděpodobnost, že se jedná o novou formu chřipky?

➔ **Řešení:**

$$P(D|A^c) = \frac{P(D \cap A^c)}{P(A^c)} = \frac{P(D)}{P(A^c)} = \frac{0,05}{0,3} = 0,167$$

# Celková pravděpodobnost a Bayesův vzorec

➡ Můžeme-li rozdělit základní prostor na  $k$  po dvou disjunktních podmnožin ( $H_i, i = 1, \dots, k$ ), pro které zároveň platí, že jejich sjednocení je celý základní prostor (tzv. systém hypotéz), pak pravděpodobnost jevu  $A$  lze získat jako

$$P(A) = \sum_{i=1}^k P(A|H_i)P(H_i)$$

**Vzorec pro celkovou pravděpodobnost**

➡ Dále platí

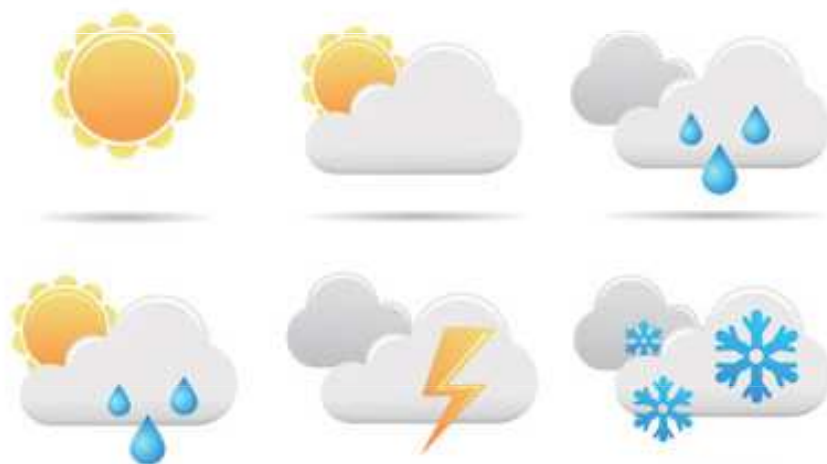
$$P(H_j | A) = \frac{P(A \cap H_j)}{P(A)} = \frac{P(A|H_j)P(H_j)}{\sum_{i=1}^k P(A|H_i)P(H_i)}$$

**Bayesův vzorec**









# Počasí a podmíněná pravděpodobnost

➡ Co má počasí společného s pravděpodobností?



# Počasí a podmíněná pravděpodobnost

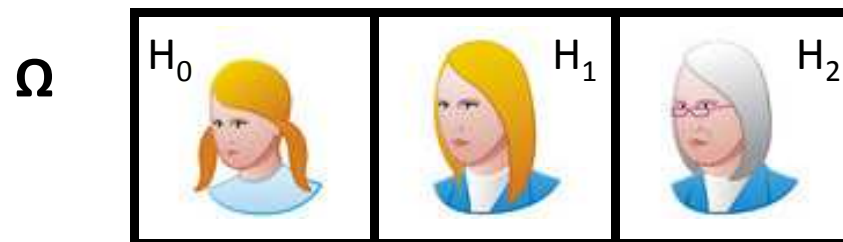
- ➔ Co má počasí společného s pravděpodobností?
- ➔ U každého jevu ( $A$ ) se můžeme ptát na jeho pravděpodobnost za slunečného počasí, za deště, za bouřky, atd. Celkovou pravděpodobnost jevu  $A$  potom můžeme získat jako součet přes tyto možnosti.
- ➔ Tyto stavy lze chápat jako **výchozí hypotézy** ovlivňující výsledek, přičemž vždy nastává (platí) pouze jeden z těchto stavů (hypotéz). **Pokud pozorujeme jev  $A$ , můžeme se zpětně ptát na platnost těchto hypotéz (s použitím Bayesova vzorce).**

$\Omega$	$H_0$ 	$H_1$ 	$H_2$ 
	$H_3$ 	$H_4$ 	$H_5$ 



# Celková pravděpodobnost – jiný příklad

- ➔ Populaci můžeme rozdělit dle věku na tři skupiny: děti ( $H_0$ ), dospělé v produktivním věku ( $H_1$ ) a dospělé v postproduktivním věku ( $H_2$ ), přičemž známe rozdělení populace, tedy známe  $P(H_0)$ ,  $P(H_1)$  a  $P(H_2)$ .



- ➔ Označme jev  $A$ : stane se úraz.
- ➔ Známe-li pravděpodobnost úrazu u dítěte,  $P(A|H_0)$ , u dospělého v produktivním věku,  $P(A|H_1)$ , a u dospělého v postproduktivním věku,  $P(A|H_2)$ , jsme schopni pomocí vzorce pro celkovou pravděpodobnost spočítat  $P(A)$ .

# Bayesův vzorec

➡ **Příklad:** Uvažujme populaci mužů nekuřáků ve věku 50 – 60 let, u kterých sledujeme výskyt chronického kašle (jev  $A$ ). Dle stavu plic můžeme muže zjednodušeně rozdělit na zdravé (jev  $H_1$ ), nemocné plicním karcinomem (jev  $H_2$ ) a nemocné sarkoidózou (jev  $H_3$ ). Pravděpodobnosti výskytu jednotlivých plicních onemocnění jsou známy, navíc známe i pravděpodobnosti výskytu chronického kašle dle stavu plic:

$$P(H_1) = 0,991, P(H_2) = 0,001, P(H_3) = 0,008$$

$$P(A|H_1)=0,002, P(A|H_2)=0,900, P(A|H_3)=0,950$$

➡ Zajímá nás, s jakou pravděpodobností bude u pacienta s chronickým kašlem při podrobnějším vyšetření diagnostikován karcinom plic.

# Bayesův vzorec

- ➔ **Příklad:** Uvažujme populaci mužů nekuřáků ve věku 50 – 60 let, u kterých sledujeme výskyt chronického kašle (jev  $A$ ). Dle stavu plic můžeme muže zjednodušeně rozdělit na zdravé (jev  $H_1$ ), nemocné plicním karcinomem (jev  $H_2$ ) a nemocné sarkoidózou (jev  $H_3$ ). Pravděpodobnosti výskytu jednotlivých plicních onemocnění jsou známé, navíc známe i pravděpodobnosti výskytu chronického kašle dle stavu plic:

$$P(H_1) = 0,991, \quad P(H_2) = 0,001, \quad P(H_3) = 0,008$$
$$P(A|H_1)=0,002, \quad P(A|H_2)=0,900, \quad P(A|H_3)=0,950$$

- ➔ Zajímá nás, s jakou pravděpodobností bude u pacienta s chronickým kašlem při podrobnějším vyšetření diagnostikován karcinom plic.

## ➔ Řešení:

$$P(H_2 | A) = \frac{P(A \cap H_2)}{P(A)} = \frac{P(A|H_2)P(H_2)}{\sum_{i=1}^3 P(A|H_i)P(H_i)}$$
$$P(H_2 | A) = \frac{0,900 \times 0,001}{0,002 \times 0,991 + 0,900 \times 0,001 + 0,950 \times 0,008} = 0,086$$

# Význam podmíněné pravděpodobnosti v biostatistice

- ➔ Princip podmíněné pravděpodobnosti je v biostatistice velmi častý – máme **systém hypotéz** (nejčastěji dvou) o vlastnostech cílové populace a pozorovaná data. Na jejich základě pak rozhodujeme o platnosti stanovených hypotéz.
- ➔ Přímé použití podmíněné pravděpodobnosti lze demonstrovat na příkladu binárních **diagnostických testů**:
  - ➔ Osoba ve skutečnosti má (jev  $H$ ) nebo nemá (jev  $H^c$ ) sledované onemocnění.
  - ➔ Diagnostický test u dané osoby indikuje přítomnost (jev  $A^+$ ) nebo nepřítomnost (jev  $A^-$ ) sledovaného onemocnění.
  - ➔ Nás zajímají diagnostické schopnosti testu.

# Senzitivita, specificita

		Skutečnost – přítomnost nemoci	
		Ano (H)	Ne (H <sup>c</sup> )
Výsledek diagnostického testu	Pozitivní (A <sup>+</sup> )	T	U
	Negativní (A <sup>-</sup> )	V	W

➔ **Senzitivita testu:** schopnost testu rozpoznat skutečně nemocné osoby, tedy pravděpodobnost, že test bude pozitivní, když je osoba skutečně nemocná.

➔ Senzitivita testu =  $P(A^+ | H) = T / (T + V)$ .

➔ **Specificita testu:** schopnost testu rozpoznat osoby bez nemoci, tedy pravděpodobnost, že test bude negativní, když osoba není nemocná.

➔ Specificita testu =  $P(A^- | H^c) = W / (U + W)$ .

# Pozitivní a negativní prediktivní hodnota

		Skutečnost – přítomnost nemoci	
		Ano (H)	Ne (H <sup>c</sup> )
Výsledek diagnostického testu	Pozitivní (A <sup>+</sup> )	T	U
	Negativní (A <sup>-</sup> )	V	W

➔ **Prediktivní hodnota pozitivního testu:** pravděpodobnost, že osoba je skutečně nemocná, když je test pozitivní.

➔ Prediktivní hodnota pozitivního testu =  $P(H|A^+) = T / (T + U)$ .

➔ **Prediktivní hodnota negativního testu:** pravděpodobnost, že osoba není nemocná, když je test negativní.

➔ Prediktivní hodnota negativního testu =  $P(H^c|A^-) = W / (V + W)$ .

# Shrnutí

		Skutečnost – přítomnost nemoci	
		Ano (H)	Ne (H <sup>c</sup> )
Výsledek diagnostického testu	Pozitivní (A <sup>+</sup> )	T	U
	Negativní (A <sup>-</sup> )	V	W

T + U → **Prediktivní hodnota pozitivního testu**

V + W → **Prediktivní hodnota negativního testu**

$T + V$   
↓  
**Senzitivita testu**

$U + W$   
↓  
**Specifická testu**

# Senzitivita, specificita

➔ **Příklad:** Zajímá nás přesnost vyšetření jater ultrazvukem, tedy schopnost vyšetření UTZ identifikovat maligní ložisko v pacientových játrech. Přesnost je vztažena k histologickému ověření odebrané tkáně. Výsledky jsou dány tabulkou:

Vyšetření UTZ	Histologické ověření		Celkem
	Maligní	Benigní	
Maligní	32	2	34
Benigní	3	24	27
Celkem	35	26	61

➔ Senzitivita testu =  $P(A^+ | H) = ?$

➔ Specificita testu =  $P(A^- | H^c) = ?$



# Senzitivita, specificita

➔ **Příklad:** Zajímá nás přesnost vyšetření jater ultrazvukem, tedy schopnost vyšetření UTZ identifikovat maligní ložisko v pacientových játrech. Přesnost je vztažena k histologickému ověření odebrané tkáně. Výsledky jsou dány tabulkou:

Vyšetření UTZ	Histologické ověření		Celkem
	Maligní	Benigní	
Maligní	32	2	34
Benigní	3	24	27
Celkem	35	26	61

➔ **Senzitivita testu** =  $P(A^+ | H) = 32 / 35 = 91,4 \%$  (IS = 75,8 – 97,8)

➔ **Specificita testu** =  $P(A^- | H^c) = 24 / 26 = 92,3 \%$  (IS = 73,4 – 98,7)

# Bayesův vzorec pro výpočet prediktivních hodnot

➡ Obě prediktivní hodnoty testu lze vypočítat s pomocí charakteristik testu, senzitivity a specificity, a celkové prevalence onemocnění v cílové populaci.

**Senzitivita testu**  $P(A^+ | H)$

**Specificita testu**  $P(A^- | H^c)$

**Prevalence**  $P(H)$

**Prediktivní hodnota  
pozitivního testu**

$$P(H | A^+) = \frac{P(A^+ | H)P(H)}{P(A^+ | H)P(H) + P(A^+ | H^c)P(H^c)}$$

**Prediktivní hodnota  
negativního testu**

$$P(H^c | A^-) = \frac{P(A^- | H^c)P(H^c)}{P(A^- | H^c)P(H^c) + P(A^- | H)P(H)}$$

# Pozitivní a negativní prediktivní hodnota

➔ **Příklad:** Zajímají nás pozitivní a negativní prediktivní hodnoty diagnostického testu na HIV pozitivitu, u kterého výrobce garantuje 98% senzitivitu a 99% specificitu.

1. Uvažujme jihoafrickou zemi s prevalencí HIV pozitivních cca 20 %:

$$P(A^+ | H) = 0,98; P(A^- | H^c) = 0,99; P(H) = 0,2.$$

# Pozitivní a negativní prediktivní hodnota

➔ **Příklad:** Zajímají nás pozitivní a negativní prediktivní hodnoty diagnostického testu na HIV pozitivitu, u kterého výrobce garantuje 98% senzitivitu a 99% specifitu.

1. Uvažujme jihoafrickou zemi s prevalencí HIV pozitivních cca 20 %:

$$P(A^+ | H) = 0,98; P(A^- | H^c) = 0,99; P(H) = 0,2.$$

## Prediktivní hodnota pozitivního testu

$$P(H | A^+) = \frac{P(A^+ | H)P(H)}{P(A^+ | H)P(H) + P(A^+ | H^c)P(H^c)} = \frac{0,98 \times 0,20}{0,98 \times 0,20 + (1 - 0,99) \times (1 - 0,20)} = 96,1\%$$

## Prediktivní hodnota negativního testu

$$P(H^c | A^-) = \frac{P(A^- | H^c)P(H^c)}{P(A^- | H^c)P(H^c) + P(A^- | H)P(H)} = \frac{0,99 \times (1 - 0,20)}{0,99 \times (1 - 0,20) + (1 - 0,98) \times 0,20} = 99,5\%$$

# Pozitivní a negativní prediktivní hodnota

➔ **Příklad:** Zajímají nás pozitivní a negativní prediktivní hodnoty diagnostického testu na HIV pozitivitu, u kterého výrobce garantuje 98% senzitivitu a 99% specificitu.

2. Uvažujme evropskou zemi s prevalencí HIV pozitivních cca 0,2 %:

$$P(A^+ | H) = 0,98; P(A^- | H^c) = 0,99; P(H) = 0,002.$$

# Pozitivní a negativní prediktivní hodnota

➔ **Příklad:** Zajímají nás pozitivní a negativní prediktivní hodnoty diagnostického testu na HIV pozitivitu, u kterého výrobce garantuje 98% senzitivitu a 99% specificitu.

2. Uvažujme evropskou zemi s prevalencí HIV pozitivních cca 0,2 %:

$$P(A^+ | H) = 0,98; P(A^- | H^c) = 0,99; \mathbf{P(H) = 0,002}.$$

## Prediktivní hodnota pozitivního testu

$$P(H | A^+) = \frac{P(A^+ | H)P(H)}{P(A^+ | H)P(H) + P(A^+ | H^c)P(H^c)} = \frac{0,98 \times 0,002}{0,98 \times 0,002 + (1 - 0,99) \times (1 - 0,002)} = 16,4\%$$

## Prediktivní hodnota negativního testu

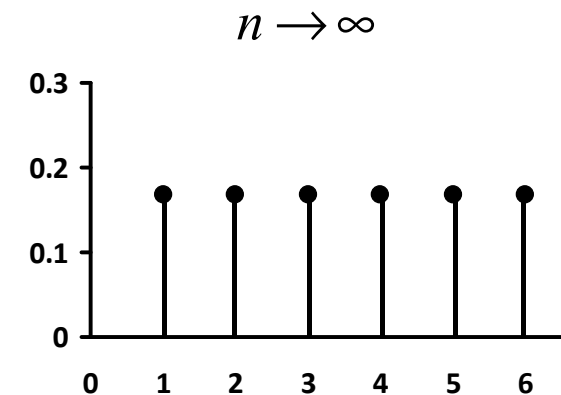
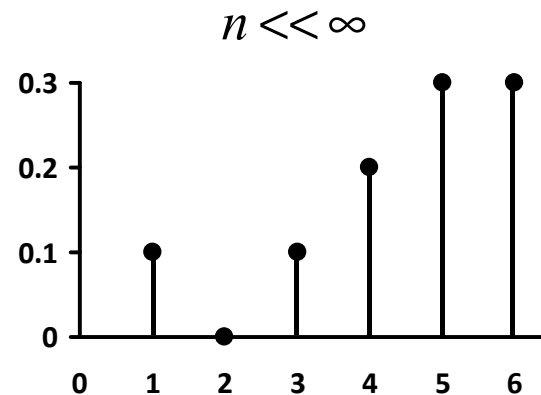
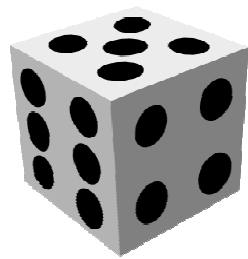
$$P(H^c | A^-) = \frac{P(A^- | H^c)P(H^c)}{P(A^- | H^c)P(H^c) + P(A^- | H)P(H)} = \frac{0,99 \times (1 - 0,002)}{0,99 \times (1 - 0,002) + (1 - 0,98) \times 0,002} = 99,9\%$$

# Dva směry statistiky

- ➔ Ve statistice existují dva hlavní filozofické směry: **frekventistický** a **Bayesovský**.
- ➔ Liší se v pohledu na pravděpodobnostní chování neznámých hodnot, které se snažíme odhadnout.
- ➔ **Frekventistická statistika**: všechny neznámé hodnoty považujeme za konstantní (parametry). Na základě dat se snažíme tuto hodnotu „lokalizovat“.
- ➔ **Bayesovská statistika**: všechny neznámé hodnoty mají pravděpodobnostní chování (rozdělení pravděpodobnosti). Na základě dat se snažíme toto pravděpodobnostní chování „upřesnit“.

# Frekventistická statistika

- ➔ Neznámou charakteristiku cílové populace (konstantu) se snažíme **odhadnout** pouze na základě pozorovaných dat.
- ➔ Důležitý je předpoklad reprezentativnosti vzorku – **pracujeme pouze s daty** jako obrazem neznámé charakteristiky. Bude-li špatný vzorek, bude špatný i odhad (výsledky mohou být velmi odlišné od známých hodnot).
- ➔ Často pracuje s **asymptotickým chováním**, kdy velikost vzorku jde do nekonečna; řada odhadů a testů je odvozena právě pro tyto situace.





# Bayesovská statistika

- ➔ Neznámá charakteristika cílové populace má pravděpodobnostní chování, které se snažíme pomocí pozorovaných dat **upřesnit**.

$$P(H | A) = \frac{P(A | H)P(H)}{P(A)} \propto P(A | H)P(H)$$

- ➔ Předpoklad reprezentativnosti vzorku je stále důležitý, ale již nepracujeme pouze s daty – pracujeme i s tzv. **apriorní pravděpodobností**,  $P(H)$ , což je náš vstupní předpoklad o chování neznámé charakteristiky.
- ➔ Nevýhodou je neznalost apriorní pravděpodobnosti.

# Reklama na další týdny...

Středem zájmu statistiky a biostatistiky je tzv. náhodná veličina.

