

# Statistika – úvod (1)

---

- Vědecky zaměření lékaři – snaha zaznamenávat, hodnotit, vysvětlovat a využít pozorované skutečnosti
  - s proměnlivostí biologických objektů si nevěděli rady
- hledání trvalých a obecně platných zákonitostí
-

# Statistika – úvod (2)

---

→ umožnila až aplikace metod **MATEMATICÉ STATISTIKY** (v lékařském výzkumu používány zejména po 2.svět.válce)

Nezbytné charakteristiky věd.výzkumu:

- přesnost**
  - správnost**
  - spolehlivost**
-

# Statistika – úvod (3)

---

**Těžiště stat.metod** spočívá v

- racionálním přístupu k řešení problému
- plánování výzkumu
- správné interpretaci a objektivizaci závěrů

x nejen v aritmetických operacích

---

# Statistika – úvod (4)

---

Ve starověku – soupis občanů, půdy + všeho, co tvořilo základ síly státu

STATISTIKA – prošla vývojem od pouhé registrace k analytickému hodnocení

poč. 20.stol. → náplň STAT: hodnocení **Ø veličin** + obohaceno o studium jejich **variability**

- vypracována řada MATEMAT.MODELU
  - základem STAT: **TEORIE PRAVDĚPODOBNOSTI**  
→ umožňuje z info získat tu složku, kt.je zákonitá + má trvalou platnost
-

# Statistika – úvod (5)

---

- věda, jejímž předmětem studia jsou **výsledky HROMADNÉHO POZOROVÁNÍ** → jejich sběr, analýza + využití pro rozhodování a předpovědi
- **Hromadné jevy** – např. narození, úmrtí, onemocnění, příznaky nemoci aj.
- Průběh pokusu – mnoho nepatrných vlivů, kt. se vymykají naší kontrole

důsledek : *nepředvídatelné kolísání jednotlivých výsledků*

---

# Statistika – úvod (6)

---

- Hromadné jevy, kt. nelze před provedením pokusu n. pozorování zcela přesně předvídat → **HROMADNÉ NÁHODNÉ JEVI**

*Pozn.: NÁHODY nelze vyloučit, ale lze je studovat exaktními metodami*

(vědecký x nevědecký výklad)

→ Studium náhodných jevů pomáhá *lépe pronikat k podstatě jevů a zákonitostem přírody*

---

# Statistika – úvod (shrnutí 1)

---

- vychází z moderní epidemiologické metody
- zabývá se **sběrem, popisem a analýzou** dat

**DATA** – zjištěné (naměřené) vlastnosti stat.jednotek

→ hodnoty, kt.se vyznačují variabilitou (důsledek působení velkého počtu drobných **náhodných** vlivů)

**NÁHODA** – přirozený jev, kt.lze zkoumat exaktními metodami teorie pravděpodobnosti (neodpovídá-li variabilita dat variabilitě, kt.způsobuje náhoda, pak to lze stat.metodami určit)

---

# Statistika – úvod (shrnutí 2)

---

→ věda, kt.se zabývá studiem **hromadných náhodných jevů**

**Jev** (*v teorii pravděpodobnosti*) – výsledek pokusu nebo pozorování

**Hromadné jevy** – mohou nastat jako výsledky neomezeně opakovatelných pokusů; výsledky pozorování na rozsáhlých souborech vzájemně rovnocenných prvků (např. narození, úmrtí, symptomy nemocí atd.)

**Hromadné náhodné jevy** – nelze předvídat výsledek (před provedením pokusu n.pozorování)

---



# Statistická úvaha

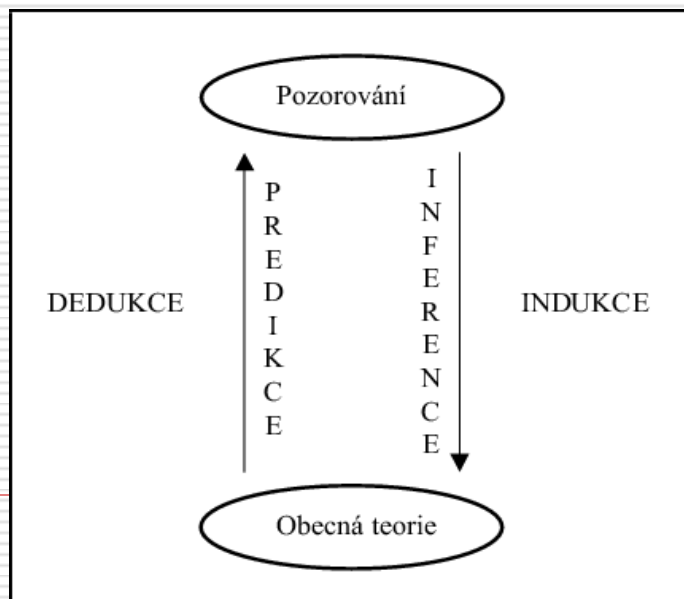
---

Aplikace statistických metod se úzce váže na záměry a úvahy vědeckého pracovníka: **deduktivní úvaha**

**induktivní úvaha**

**Deduktivní úvaha:** od obecných zkušeností k jednotlivým (konkrétní)

**Induktivní úvaha:** od jednotlivých zkušeností k obecným.



# Role STAT. v medicíně

---

využití na úrovni:

- **populační** – úroveň a vývoj zdrav.situace; zvažuje zdrav.stav lidí, determinanty zdraví + možnost jejich ovlivnění
  - **Individuální** – stanovení správné diagnózy (mnoho kvalitativ.+kvantitativ.údajů) + odhad prognózy léčby
-

# Etapy stat.šetření

---

1. Plán šetření
  2. Sběr dat
  3. Popis a technické zpracování, tzv. **deskriptivní statistika**
  4. Rozbor, závěry, interpretace, tzv. **induktivní statistika**
-

# Ad 1. Plán (1)

---

- ujasnění problému + volba metody šetření
- přesně definovat cíl výzkumu (vědec: kriticky prověřit x diletant. dokázat)
- zvolit vhodnou stupnici měření (nominální x pořadová x intervalová)
- stanovíme určující znaky (vymezují stat.jednotku věcně, místně, časově), jimiž definujeme stat.jednotku šetření  
→ **vymezíme základní soubor**

určíme:

- metodu výzkumu
  - způsob výběru
  - velikost výběrového souboru
  - skladbu kontrolní skupiny
-

# Ad 1. Plán (2)

---

- formulujeme **testovanou hypotézu**
  - zvolíme **typ statistického testu**
  - uvážíme, zda přínos odpovídá nákladům
  - zvážíme vhodnost pilotní studie
-

# Ad 1. Plán (3)

---

Plán stat.šetření předurčuje všechny ostatní etapy  
+ použitelnost výsledků v praxi.

J.W.Goethe:

**„Kdo splete první knoflík, ten se pořádně nezapne.“**

---

# Ad 2. Sběr dat

---

podle rozsahu výběru a cíle → volba **vhodné metody záznamu**

- pracovní protokol
- dotazník – nabídnuté odpovědi n. otevřené otázky
- záznam přímo do počítače

*Pozn.: u laboratorních měření – systematické chyby, náhodné chyby*

*u populačního šetření – přiměřený počet otázek, jednoznačná formulace, řešit neúplnost dat (zdroj možného zkreslení, např. malá návratnost)*

---

# Ad 3. Popis stat.souboru (DESKRIPTIVNÍ STATISTIKA)

---

- *smyslem je především uspořádat (roztřídit) získané údaje + vypočítat takové **ukazatele**, kt.by souhrnně charakterizovaly studovaný soubor vcelku*
  - *Význam: popis + východisko statistického rozboru směřujícího k poznání pravidelností, souvislostí, závislostí a zákonitostí + umožňuje **usuzovat z vlastností zkoumaného výběrového souboru na soubor základní***
-



# Ad 4. Rozbor a závěry (STATISTICKÁ INDUKCE)

---

- Vychází z *teorie pravděpodobnosti*
  - Zabývá se *zobecněním zjištěných výsledků*
  - Využívá celé řady *matematických modelů* (např. rozdělení náhodné proměnné)
  - Umožňuje vhodně zvolenými stat. testy *ověření testované hypotézy*
- způsob třídění i volba stat. ukazatelů může podstatně ovlivnit možnosti, správnost, přínos obecných závěrů

zaměříme se na 3 nejdůležitější oblasti:

- **odhady neznámých parametrů** pro celý soubor
  - **srovnání** dvou parametrů
  - **hodnocení závislosti** dvou veličin
-

# Statistické šetření (1)

---

- **vyčerpávající (úplné)** → sledujeme všechny prvky stat.souboru
- **výběrové** → vybíráme urč.část stat.jednotek → **VÝBĚR**, kt.podrobujeme vlastnímu stat.šetření

usuzujeme-li induktivně z vlastností stat. Jednotek výběr.souboru na vlastnosti všech stat.jednotek → hovoříme o **statistické indukci**

objektivizuje induktivní závěry pomocí teorie pravděpodobnosti

---

# Statistické šetření (2)

---

**Základní soubor** → souhrn prvků (osob, případů nemoci, pokusů), jejichž *vlastnosti chceme poznat*

**Výběrový soubor** → vybraná část ZS, kterou *budeme skutečně testovat* (měření, dotazníky, testy)

**NÁHODNÝ VÝBĚR** – postup, kdy každý prvek ZS má na začátku stejnou p<sub>st</sub>, že bude vybrán do VS

---

# Náhodný výběr

---

## ➤ podmínka reprezentativnosti

podle způsobu provedení:

- **Prostý náhodný výběr** (losování)
- **Systematický náhodný výběr** (mechanický)
- **Stratifikovaný (oblastní) náhodný výběr**
- **Párový výběr** (mačování)

*Pozn.: reprezentativnost může být porušena i při sběru dat, např.*

- *neúplné chybějící údaje*
  - *nevhodně zvolené otázky*
  - *nejednoznačnost odpovědi atd.*
-

# Deskriptivní statistika (popis souboru)

---

- **Stat.třídění** – rozdělení stat.souboru do skupin (tříd) podle předem určených třídících znaků; metoda třídění umožňuje zkoumat strukturu souboru + určit typ rozdělení sledovaných veličin
    - třídění jednostupňové
    - třídění vícestupňovécíl: uspořádat + zpřehlednit velký soubor dat
  - **Výsledky třídění** – tabulky + grafy  
cíl: znázornit rozložení četnosti sledovaných znaků
  - **Statistické charakteristiky – zhuštěná info o celém souboru**  
cíl: charakterizovat sledované znaky pomocí výstižných ukazatelů
-

# Třídění kvalitativních veličin

---

- nelze měřit číselně x pouze klasifikovat do různých kategorií (pohlaví, věk..)
- Kategorie jsou předem dány
  - jde o výčet všech hodnot, kt.může veličina nabývat (barva očí – modrá, hnědá, zelená..)

- 1. Nominální** → lze vyjádřit pouze slovně, nelze seřadit
    - 1. alternativní** – existují pouze 2 varianty (kuřák x nekuřák, muž x žena)
    - 2. množné** – existují > dvě varianty (diagnózy, barva vlasů)
  - 2. Ordinální** → lze je seřadit dle kriterií  
(ZŠ – SŠ – VŠ; silnýkuřák – slabý kuřák – nekuřák)
-

# Třídění kvantitativních veličin

---

→ lze vyjádřit pouze číselně

1. **Diskrétní** – vyjádřeny celými čísly
2. **Spojité** – desetinná čísla (výška, hmotnost...)

*Pozn.: v praxi lze spojité znaky převést na diskrétní*

- Kategorie (třídy) vytváříme teprve na základě získaných dat
  - Dochází k redukci dat ve prospěch přehlednosti
-

# Vytváření intervalů

---

## 1. Rozpětí

od největší naměřené hodnoty odečtu k nejmenší

$$6,59 - 3,08 = 3,51$$

## 2. Stanovení počtu intervalů

závisí na mnoha faktorech (velikost souboru, podrobnost...); stanovíme, kolik kategorií chceme mít (5-20)

*např. 10*

## 3. Délka intervalu

rozpětí / počet intervalů

$$3,51 / 10 = 0,351 \text{ (délka jednoho intervalu) } = 0,40$$

pravidlo:           a) okrouhlé číslo  
                          b) ne víc deset.míst než měřená veličina

## 4. Hranice intervalu

počátek – od nejmenšího čísla  $3,08 \rightarrow 3,00$

*1.int.: 3,00 – 3,39*

*2.int.: 3,4 – 3,79 Pozn.: nesmí se překrývat*

---



# Prezentace dat

---

➤ **Četnost**

*kolik z naměřených hodnot spadá do jednotl. intervalů*

➤ **Kumulativní četnost**

*součet všech předchozích intervalů*

➤ **Relativní četnost**

*% z celk. počtu měření*

➤ **Kumulativní relativní četnost**

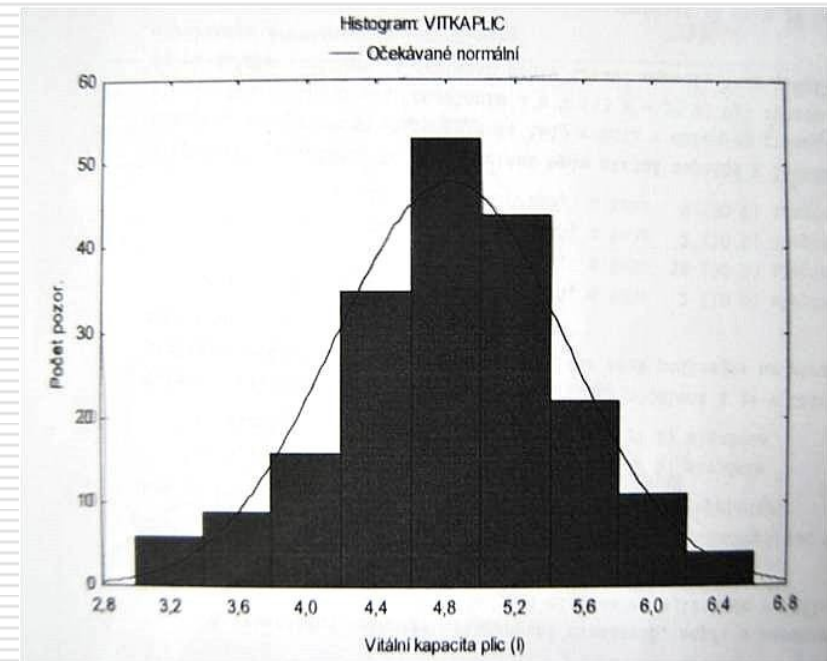
*obdoba kumulativní četnosti v %*

---

# Tabulka: Vitální kapacita plic

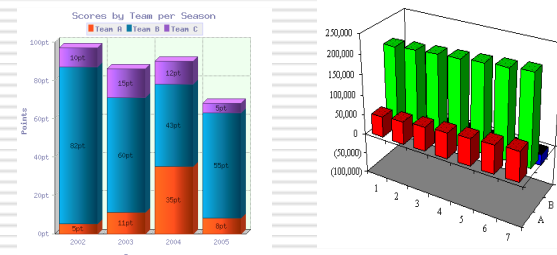
		Tabulka četnosti: VITKAPLIC			
OD	DO	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
3,000000	<=x<3,400000	6	6	3,00000	3,0000
3,400000	<=x<3,800000	9	15	4,50000	7,5000
3,800000	<=x<4,200000	16	31	8,00000	15,5000
4,200000	<=x<4,600000	35	66	17,50000	33,0000
4,600000	<=x<5,000000	53	119	26,50000	59,5000
5,000000	<=x<5,400000	44	163	22,00000	81,5000
5,400000	<=x<5,800000	22	185	11,00000	92,5000
5,800000	<=x<6,200000	11	196	5,50000	98,0000
6,200000	<=x<6,600000	4	200	2,00000	100,0000
6,600000	<=x<7,000000	0	200	0,00000	100,0000
ChD		0	200	0,00000	100,0000

Popisné statistiky (Tabulka_vitální kapacita.sta)						
Proměnná	Průměr	Sm. odch.	Minimum	Maximum	N	Počet ChD
VITKAPLIC	4,825550	0,667008	3,080000	6,590000	200	0

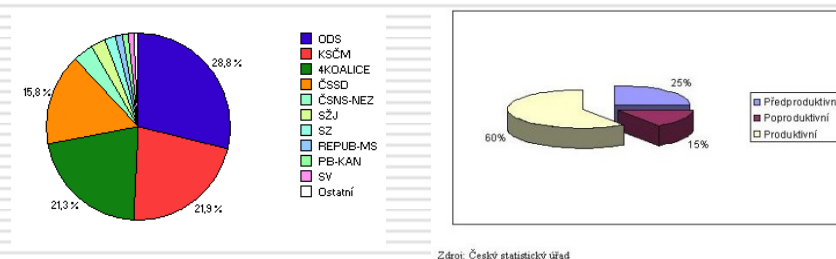


# Kvalitativní veličiny

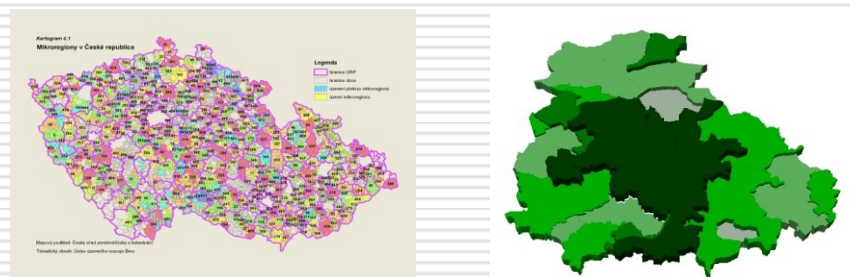
- Sloupcový graf (sloupce oddělené mezerou)



- Výšečový graf (struktura)

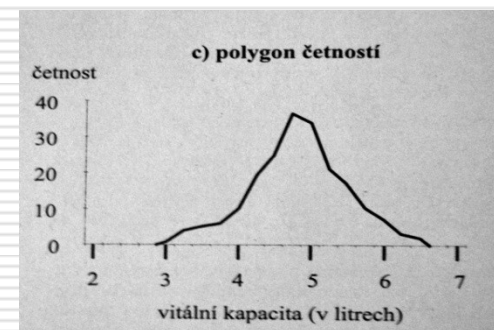
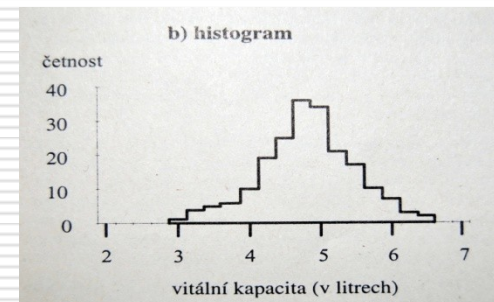
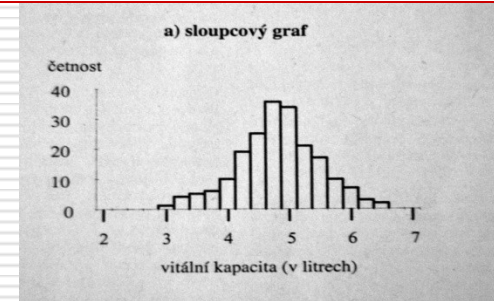


- Kartogram (regionální srovnání)



# Kvantitativní veličiny

- Sloupcový graf
- Histogram (pouze obrysy sloupců)
- Polygon (středky sloupců se spojí křivkou)



# Grafy znázorňující frekvenci rozložení veličiny

---

osa **X** : naměřené hodnoty sledování veličiny

osa **Y** : četnost (abs. nebo v %) intervalů

## Tvar rozložení četností:

- Symetrické x asymetrické
  - Jednovrcholové x vícevrcholové
  - Podoba s teoretickými modely rozložení četností
-

# Statistické charakteristiky

---

- Vyjadřujeme jimi ty vlastnosti stat.jednotek a stat.souboru jako celku, kt.považujeme za podstatné
- Používáme, chceme-li podat souhrnnou zhuštěnou informaci o stat.souboru + srovnat zjištěné výsledky

## Statistické charakteristiky

**a) Ukazatelé polohy**

**b) Ukazatelé variability**

---

# Ukazatelé polohy (střední hodnoty)

---

→ většina hodnot, jež mohou náhodné veličiny nabývat, se kupí kolem některého pevného bodu – *středu rozdělení četností*, kt.charakterizuje polohu stat.souboru na číselné ose

K nejčastěji užívaným ukazatelům polohy patří:

- **Aritmetický průměr**
  - **Medián**
  - **Modus**
-

# Ukazatelé variability

---

- hodnoty náhodné proměnné kolísají v určitém rozmezí kolem středních hodnot (měnlivost, variabilita)
- tyto ukazatelé kvantifikují míru tohoto kolísání tohoto kolísání (rozptýlení)

## Nejběžnější ukazatelé variability:

- **Rozpětí**
  - **Rozptyl**
  - **Směrodatná odchylka**
  - **Variační koeficient**
-



# Volba ukazatele

---

1. Tvar rozložení (sym. X asym.)

2. Typ sledovaného znaku

- Nominální: modus
  - Ordinální: modus, medián, percentil
  - Intervalové: medián, modus, percentil, průměr
-

# Ukazatele polohy (střední hodnoty)

---

**1. Aritmetický průměr** (ne u asym. – náchylný k extrémním hodnotám)  
*součet sledovaných hodnot vydělený počtem sledovaných jednotek*

## **2. Medián**

*hodnota, kt. je právě uprostřed všech pozorování seřazených podle velikosti (sudý počet – průměr 2 prostředních)*

## **3. Modus**

*hodnota s největší četností (nejčastější)*

---

# Ukazatele variability

---

## 1. Rozpětí

$X_{\max.} - X_{\min.}$ ; pro  $N \leq 10$

**2. Rozptyl:** *průměr čtverců odchylek arit.průměru od jednotliv.pozorování*

**3. Směrodatná odchylka:** *odmocněný roztpyl*

- Ukazatel variability udávaný ve stejných jednotkách jako sledovaný znak
- O kolik se většina hodnot sledovaného znaku odchyluje od průměru

**4. Variační koeficient**

- Relativní míra variability
  - *Jaký podíl tvoří směrodat.odchylka z průměru*
-

# Využití variačního koeficientu

---

- 1) Je-li v. k.  $>50\%$ , pak je soubor natolik nesourodý, že nemá smysl charakterizovat ho aritmetickým průměrem.
- 2) Slouží ke srovnání variability 2 souborů, jejichž průměry se značně liší.

Př.: VC u mužů a žen

M:  $m = 4,80$                        $s = 0,66$                       v. k. =  $13,8\%$

Ž:  $m = 3,90$                        $s = 0,42$                       v. k. =  $10,8\%$

- 3) Slouží ke srovnání variability znaků uváděných v různých jednotkách.

Př.: VC, tělesná výška, hmotnost u mužů

VC:  $m = 4,80$  l                       $s = 0,66$                       v. k. =  $13,8\%$

Výška:  $m = 178$  cm                       $s = 4$                       v. k. =  $2,2\%$

Hmotnost:  $m = 82$  kg                       $s = 6$                       v. k. =  $7,3\%$

---

# Ukazatele variability pro asym.rozložení četností

---

- Variabilitu vyjadřujeme pomocí dvou kvantilů – percentilů, decilů, kvartilů
- Kvantily dělí soubor uspořádaný dle velikosti na části obsahující stejný podíl z celk.počtu jednotek
- Variabilita – určení intervalu, ve kt.se pohybuje 80% (P10-P90), příp. 50% (P25-P75)

## Postup:

1. Určíme hodnotu pozorování, kt.představuje 10.percentil  
→ **DOLNÍ HRANICE INTERVALU**
  2. Určíme hodnotu pozorování, kt.představuje 90.percentil  
→ **HORNÍ HRANICE INTERVALU**
-