

# Etapy stat.šetření

---

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

# Statistická indukce – 4. etapa

---

## **Teorie odhadů**

- Odhad průměru základního souboru
- Odhad pravděpodobnosti

## **Testování statistických hypotéz**

- Srovnání dvou průměrů
- Srovnání pravděpodobností

## **Hodnocení závislosti**

- Závislost kvantitativních veličin
- Závislost kvalitativních veličin

**ZOBECNĚNÍ VÝSLEDKŮ VÝBĚROVÉHO ŠETŘENÍ NA CELÝ  
ZÁKLADNÍ SOUBOR**

**! PRAVDĚPODOBNOST, NÁHODNÁ VELIČINA A JEJÍ ROZDĚLENÍ**

---

# Pravděpodobnost

---

- Všechny **STAT výroky** – pravděpodobnostní charakter .
  
  - Jejich věrohodnost vyjadřujeme
    - pstí, že tento výrok platí → **SPOLEHLIVOST**
    - pstí, že daný výrok neplatí → **RIZIKO**
  
  - Dva základní pojmy teorie pravděpodobnosti – **náhodný jev** a **pravděpodobnost**.
  
  - **Pravděpodobnost náhodného jevu** – kvantitativní charakteristika, která je mírou častosti výskytu zkoumaného jevu.
-

# Pravděpodobnost

---

*PST někt.náhodných jevů umíme **vypočítat**, např. výhra ve Sportce, karetní hry ;  
vznik teorie pravděpodobnosti vázán na **hazardní hry** → později užití v  
jiných oblastech lidské činnosti*

□ **Klasická definice**       $P(A) = m/n$

*m = počet příznivých výsledků v experimentu*

*n = počet všech možných výsledků*

Příklady: kostka, mince

karty

sportka

narození chlapce, dívky

□ **Komplex podmínek** – souhrn předpisů, za nichž se experiment provádí.

---

# Pravděpodobnost

---

- u většiny jevů v lékařské praxi PST *neznáme* (např. výskyt onemocnění v populaci, pooperační komplikace, účinek léku)

→ **pouze odhadujeme** – pomocí **RELATIVNÍ ČETNOSTI**,

kt. zjišťujeme opakováním *pokusů* *n. pozorování* + *měříme*:

**m/n**            *m = počet příznivých výsledků v experimentu*  
*n = počet všech možných výsledků*

- blíží se skutečné PSTI tím více, čím větší počet pokusů a pozorování provedeme
-

# Pravděpodobnost

---

Pravděpodobnost reakce pacienta na určitou léčbu?

Pouze **odhad** pomocí **relativních četností**

$n$  pacientů

$x$  vyléčeno

Podíl  $x/n$  *odhaduje pravděpodobnost vyléčení  $P$*

---

# Vlastnosti pravděpodobnosti

---

1.  $0 < \text{nebo} = P(A) < \text{nebo} = 1$
2.  $P(A) = 0$  pro jevy nemožné
3.  $P(A) = 1$  pro jevy jisté

## Pravidla pro počítání

### a) Pravidlo pro sčítání

A, B disjunktní – vzájemně se vylučují

$$P(A \text{ nebo } B) = P(A) + P(B)$$

A, B nejsou disjunktní

$$P(A \text{ nebo } B) = P(A) + P(B) - P(A \text{ i } B)$$

---

# Vlastnosti pravděpodobnosti

---

## b) Pravidlo pro násobení

A, B nezávislé

$$P(A \text{ i } B) = P(A) \cdot P(B)$$

A, B závislé

$$P(A \text{ i } B) = P(A) \cdot P(B/A) \\ = P(B) \cdot P(A/B) \text{ př. dvě nemoci u člověka}$$

$P(A/B)$ ,  $P(B/A)$  jsou tzv. **podmíněné pravděpodobnosti**.

$$\text{Platí: } P(A/B) = P(A \text{ i } B) / P(B)$$

$$P(B/A) = P(A \text{ i } B) / P(A)$$

Pravidlo pro sčítání i násobení se dá rozšířit na více jevů.

Podmíněné pravděpodobnosti jsou užitečné pro hodnocení **rizika** nemoci v populaci

---



# Vlastnosti pravděpodobnosti

---

Např. Ze srovnání pravděpodobností

$P(\text{Ca})$ ,  $P(\text{Ca}/\text{K})$ ,  $P(\text{Ca}/\text{N})$

Lze usuzovat na riziko kouření (K,N) na výskyt karcinomu plic (Ca)

Např.  $P(\text{Ca}/\text{K})/P(\text{Ca})$  udává, kolikrát je větší pravděpodobnost výskytu karcinomu plic u kuřáků než v celé populaci

$P(\text{Ca}/\text{K})/P(\text{Ca}/\text{N})$  udává, kolikrát je větší pravděpodobnost výskytu karcinomu plic u kuřáků než u nekuřáků

---

# Výběrový/základní soubor

---

## Výběrový soubor

- reprez. náhodný výběr
- výběrové (empirické) rozdělení četností
- popis rozdělení: tabulka, graf
- stat. ukazatele = výběrové charakteristiky: **m, s, p** (ozn. latinkou)
- jsou to charakteristiky náhodných veličin, tzn. mění se výběr od výběru + je nutné počítat s chybami (výběrové, náhodné)

## Základní soubor

- soubor, který nás zajímá
- teoretické rozdělení četností (matematický model)
- popis rozdělení: pravděpodobnostní rozdělení
- stat. ukazatele = parametry:  **$\mu, \sigma, \pi$**  (ozn. řeckou abecedou)
- jsou to konstanty, zpravidla neznámé, pro

$$n \rightarrow \infty \text{ platí, že } m \rightarrow \mu, \\ s \rightarrow \sigma, \quad p \rightarrow \pi$$

Statistická indukce = usuzování z vlastností výběru na vlastnosti základního souboru

---

# Empirické a pravděpodobnostní rozdělení

---

- každá veličina, kterou zkoumáme, je ovlivněna řadou nepatrných **náhodných vlivů**, což způsobuje její **variabilitu** – tzn. veličina nabývá u různých subjektů různých hodnot.
  - měříme-li veličinu ve výběrovém souboru, pak rozložení hodnot této veličiny znázorňujeme na základě empiricky zjištěných četností
  - každá veličina má své pravděpodobnostní (teoretické) rozdělení  
Pravděpodobnostní rozdělení (pravděpodobnostní křivka) vyjadřuje očekávání, jak často se budou jednotlivé hodnoty vyskytovat v nekonečně velkém souboru
-

# Empirické a pravděpodobnostní rozdělení

---

## □ Pravděpodobnostní (teoretické) rozdělení

v takovém rozložení jsou na ose  $x$  všechny hodnoty, kterých může veličina potenciálně nabývat a na ose  $y$  jsou zaneseny pravděpodobnosti, se kterými se dané hodnoty vyskytují

X

- v empirickém rozdělení (polygon četností) jsou popsány četnosti, se kterými se naměřené hodnoty vyskytovaly ve výběrovém souboru
-

# Typy pravděpodobnostních rozdělení

---

- přírodním i společenským jevům vlastní *různorodost* a *mnohotvárnost*  $X$  v praxi lze vystačit s poměrně malým množstvím modelů.

*Pozn.*

- *s veličinou zacházíme jako s normálně rozdělenou, pokud nemáme dostatečné důvody pro vyvrácení této domněnky*
- *rozložení většiny veličin lze převést na normální rozdělení*

## Diskrétní veličiny

- binomické rozdělení (jev – nejev)
- rovnoměrné rozdělení
- Poissonovo rozdělení (vzácné jevy)

## Spojitě veličiny

- normální rozdělení
  - Studentovo t-rozdělení
  - Snedecorovo F-rozdělení
  - Chí-kvadrát rozdělení
-

# Normální rozdělení

---

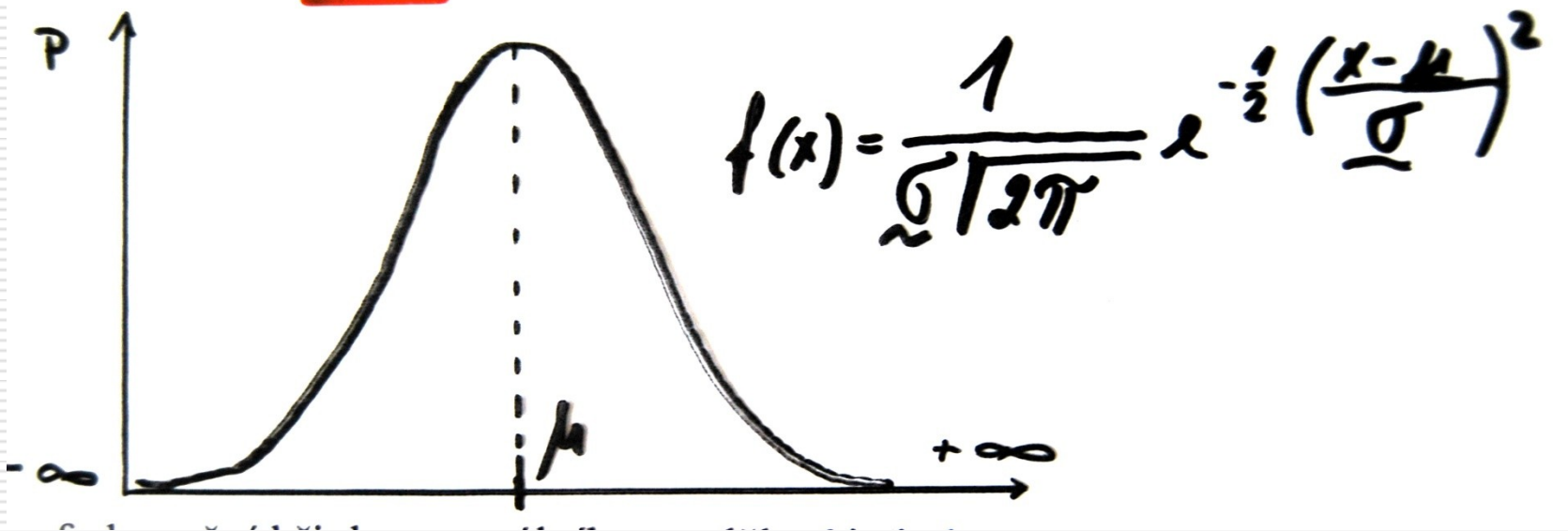
Spojité znameno  $\rightarrow$  *normální rozdělení* – je-li  
vytvářen nahromaděním velkého počtu  
nepatrných, nezávislých příčin nahodilého  
charakteru

---

# Normální rozdělení

---

- matematický model rozdělení četností **náhodné veličiny**



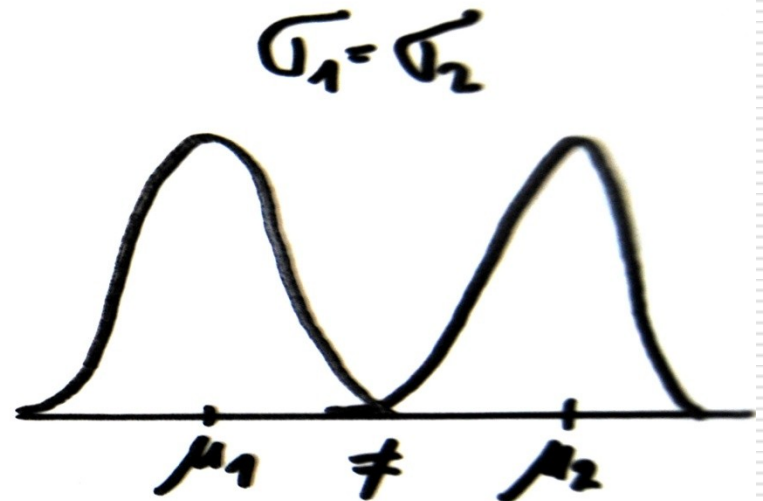
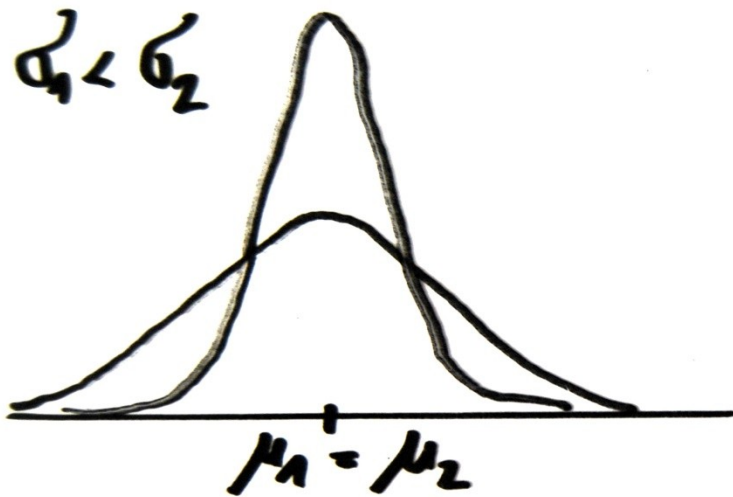
- frekvenční křivka normálního rozdělení je jednoznačně určena dvěma parametry:  $\mu$ ,  $\sigma$   
 $\mu$  určuje polohu křivky (analogie **m**)  $\mu$   
 $\sigma$  určuje tvár křivky (analogie **s**) **sigma**
-

# Normální rozdělení

---

$\mu$  určuje polohu křivky (analogie **m**) **mí**

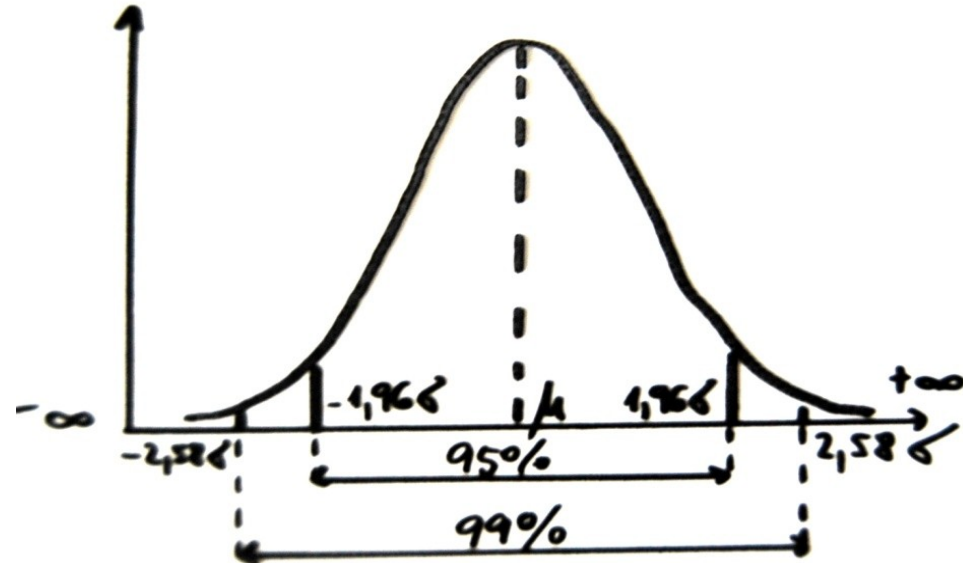
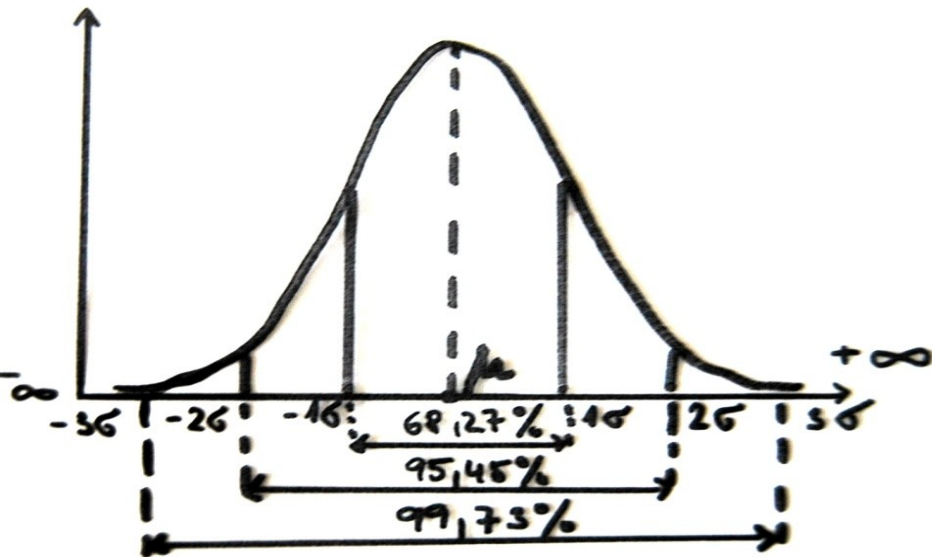
$\sigma$  určuje tvár křivky (analogie **s**) **sigma**





# Vlastnosti normálního rozložení

- Frekvenční křivky normálního rozložení mají pro různé veličiny různý tvar ( $\sigma$ ) a polohu ( $\mu$ )
- Pro všechny ale platí, že intervaly, ve kterých se odhadovaná proměnná nachází s pravděpodobností 95 nebo 99%, lze vyjádřit jako odchylky od  $\mu$  v násobcích  $\sigma$  :



# Odhady parametrů (1)

---

## 1) Bodové odhady

$$\begin{array}{l} \mu \doteq m \\ \pi \doteq p \\ \sigma \doteq s \end{array}$$

### Požadavky na bodové odhady:

- a) **Konzistence** – s rostoucím VS se výběrová charakteristika více blíží k parametru
- b) **Nestrannost** – odhady parametru provedené na základě různých VS kolísají kolem hodnoty neznámého parametru na obě strany
- c) **Minimální rozptyl** – uvedené kolísání musí být co nejmenší

### Nevýhody bodových odhadů:

- neznáme jejich **spolehlivost a přesnost**

---

# Odhady parametrů (2)

---

## 2/ Intervalové odhady

- Neznámý parametr odhadujeme intervalem vytvořeným kolem tzv. nejlepšího nestranného bodového odhadu
- **Interval spolehlivosti** (konfidenční interval)
- **Spolehlivost** si určujeme sami – buď 95% nebo 99%

jde o pravděpodobnost, že odhadovaný parametr se nachází v daném intervalu

**95% CI (-;-)**

**99% CI (-;-)**

---

# Odhady parametrů (3)

---

- doplněk spolehlivosti vyjadřuje **riziko odhadu** – tj. riziko, že odhadovaný parametr leží mimo interval

při spolehlivosti 95% →

**riziko odhadu 5%**

při spolehlivosti 99% →

**riziko odhadu 1%**

---

# Odhad průměru základního souboru (parametru $\mu$ ) [mý]

---

1. Nejlepší bodový odhad parametru  $\mu$  je výběrový průměr  $m$
2. V souborech, kde  $n > 30$ , se výběrový průměr chová jako náhodná veličina, která má normální rozdělení
3. V souborech, kde  $n < 30$ , používáme model Studentova rozdělení (konstanty 1,96, příp. 2,58 se nahrazují jinými – viz. skripta, str. 25)
4. Každý výběrový průměr je zatížen chybou – jde o tzv. **standardní chybu** průměru  $SE_m$ , kterou odhadujeme ze vztahu: (střední chyba)

Závěr:

$$SE_m = \frac{s}{\sqrt{n-1}}$$

95% CI

$$m \pm 1,96 \cdot \frac{s}{\sqrt{n-1}}$$

99% CI

$$m \pm 2,58 \cdot \frac{s}{\sqrt{n-1}}$$

# Vlastnosti odhadu

---

- 1) **Spolehlivost** – volí se předem, jde o stanovení pravděpodobnosti, obvykle 0,95 nebo 0,99
- 2) **Přesnost** – je dána délkou intervalu, čím kratší je interval, tím je vyšší přesnost odhadu

$$m \pm 1,96 \cdot \frac{s}{\sqrt{n-1}}$$

Obě vlastnosti spolu souvisí

Přesnost odhadu lze ovlivnit:

- a) snížením či zvýšením **P** spolehlivosti
  - b) snížením či zvýšením **n** (velikost souboru)
  - c) snížením či zvýšením **s** (homogenita souboru)
-

# Příklad 1:

---

Odhadněte průměrnou vitální kapacitu plic mužů  
40-50 letých na podkladě výběrového šetření  
200 mužů, u kterých jsme zjistili:

$$m = 4,82$$

$$s = 0,67$$

$$n = 200$$

---

# Příklad 2:

---

## Skupina A:

Odhadněte průměrnou hladinu hemoglobinu v populaci zdravých mužů z náhodného výběru 100 jedinců s průměrnou hodnotou  $m = 152,4$  g/l a směrodatnou odchylkou  $s = 18,2$  g/l se spolehlivostí: a) 95% b) 99%

## Skupina B:

Odhadněte průměrnou hladinu hemoglobinu v populaci zdravých mužů z náhodného výběru 35 jedinců s průměrnou hodnotou  $m = 152,4$  g/l a směrodatnou odchylkou  $s = 18,2$  g/l se spolehlivostí: a) 95% b) 99%

## Skupina C:

Odhadněte průměrnou hladinu hemoglobinu v populaci zdravých mužů z náhodného výběru 100 jedinců s průměrnou hodnotou  $m = 152,4$  g/l a směrodatnou odchylkou  $s = 14,8$  g/l se spolehlivostí: a) 95% b) 99%

---



# Odhad pravděpodobnosti ZS (parametru $\pi$ ) [pí]

1. Nejlepší bodový odhad je relativní četnost

$$P = \frac{k}{n} \rightarrow \tilde{\pi}$$

n = počet pozorování

k = počet pozorování, u nichž nastal sledovaný jev

2. Pro pravděpodobnosti sice platí binomické rozdělení, ale pokud platí

$$n \cdot p \cdot (1-p) > 9$$

můžeme vycházet z normálního rozdělení

3. **Standardní chybu SE odhadujeme ze vztahu:**

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

číslo

$$SE = \sqrt{\frac{p(100-p)}{n}} \%$$

Pro 95% CI  $\tilde{\pi} = p \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$

99% CI  $\tilde{\pi} = p \pm 2,58 \sqrt{\frac{p(1-p)}{n}}$

# Příklad:

---

Odhadněte pravděpodobnost výskytu zrakové vady u studentů LF na základě výběrového šetření u 200 studentů

$$n = 200$$

$$k = 80$$

$$p = 0,40 \text{ (40\%)}$$

---

# Příklad:

---

Ve výběru 100 šestiměsíčních zdravých dětí náhodně vybraných z brněnské populace byl sledován hemoglobin v g%.

$$n = 100$$

$$m = 13,10$$

$$s = 1,90$$

- 1) Určete interval, ve kterém se pohybuje hemoglobin u 95% vyšetřených dětí.
  - 2) Odhadněte průměrné množství hemoglobinu v základním souboru se spolehlivostí 0,95. Jaká je přesnost tohoto odhadu?
  - 3) U kolika dětí musíme provést šetření, aby přesnost odhadu průměru byla při spolehlivosti 0,95 nejméně  $\pm 0,2$ .
-