

# STATISTIKA

## VĚDA, KTERÁ HLEDÁ SOUVISLOSTI MEZI JEVY A STAVY

FUNKCÍ STATISTIKY je měřit, srovnávat, vysvětlovat-

**SNAHA SNIŽOVAT VARIABILITU JEVŮ**

**(VARIABILITA JEVU = NEJISTOTA)**

## **Aplikace statistických metod.**

1) Okruhy systémových problémů:

**A. charakteristika vstupu nebo výstupu:** - používáme deskriptivní statistické charakt. (průměr, modus, medián, s, R) nebo charakteristiky intrasystémové diachronní (klouzavé průměry, rozptyl časové řady, autokorelace)

**B. analýza vztahů mezi vstupy, výstupy:** - zajímá nás např. úsporná redukce dimenzionality výstupního prostoru (predikce některého z výstupů), používáme vícerozměrné statistiky (korelace, regrese, faktorová analýza)

**C. analýza vztahů vstup - výstup:** - má nejbližší k ověření kauzálních vztahů - má nejbližší k ověření kauzálních vztahů

- **longitudinální vnitroskupinový experiment** využívá analýzy rozptylu, mnohorozměrnou analýzu rozptylu, kovariance, F-test, atd.

- **meziskupinový experiment:** analýzu rozptylu, F-test, t-test

**D. Popis a analýza vnitřního stavu:** - popisem stavového vektoru  $s (s_1, s_2, \dots, s_n)$  se snažíme stanovit či odhadnout vnitřní stav

- využíváme modely lineárních strukturálních vztahů s latentními proměnnými, longitudinální faktorovou analýzu, faktorovou analýzu skupin proměnných

**E. Stavová analýza vztahů vstup-výstup:** - rozbor vztahů mezi vstupními proměnnými společně s nutnými stavovými proměnnými a výstupními proměnnými

- využíváme longitudinální faktorové analýzy (z kombinace vstupu a hodnot vnitřních vlastností odhadnout výstup)

## MÍRY CENTRÁLNÍ TENDENCE – ukazují „střed souboru „

- **Aritmetický průměr** –  $\bar{X}$  **průměrná hodnota souboru**
- **MEDIÁN** - **X** **prostřední hodnota souboru - rozděluje výsledky na dvě poloviny podle počtu pozorování, měření ...**
- **MODUS** - **X** **nejčastější hodnota, soubor nemusí mít modus, = x krát stejná hodnota (interval)**

Jak daleko jsou hodnoty od centra = MÍRA VARIABILITY

= HOMOGENITA souboru

---

VARIAČNÍ ROZPĚTÍ – R

Citlivé na extrémní data (bere jen krajní polohy - data)

Všechny hodnoty

- ROZPTYL = odchylka  $^2$  – ( $s^2$ ) součet hodnot (velká, ale kladná hodnota)
- SMĚRODATNÁ ODCHYLKA – s nejčastěji používaný ukazatel variability jevu, možno s ní dále počítat
- VARIAČNÍ KOEFICIENT – VK = relativizovaná sm. dch. – srovnání dvou souborů o nestejných proměnných (odch. v procentech prům.) VK =  $100s/x$

### 3) Základní statistické vzorce:

1) Aritmetický průměr:	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
2) Směrodatná odchylka:	$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cong \frac{1}{6} R_i$
3) Rozptyl:	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
4) Variační rozpětí:	$R = x_{\max} - x_{\min}$
5) Korelační koeficient:	$r_{jk} = \frac{s_{jk}^2}{s_j s_k} = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 (x_{ki} - \bar{x}_k)^2}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ki} - \bar{x}_k)^2}}$
6) Spearmanův koeficient korelace pořadí:	$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$ , kde d je diference (rozdíl) mezi pořadími
7) Modus	<b>MOD = nejčastější hodnota</b>
8) Medián	<b>MED = prostřední hodnota (průměr dvou prostředních hodnot)</b>
9) Relativní četnost:	$f_k = \frac{n_k}{n}$ , x 100 dostáváme pravděpodobnost, že výsledek je v daném intervalu
10) Relativní kumulativní četnost:	$F_k = \frac{\sum_{i=1}^k n_i}{n}$
11) Procentily:	$P = \frac{\sum_{i=1}^k n_i}{n} \cdot 100$ , udává kolik procent lidí se nachází pod TO



**VŠE PŘEDPOKLÁDÁ NORMÁLNÍ ROZLOŽENÍ DAT !!!**

**Lze počítat - malé soubory 2 – 50 dat – Shapiro-Wilkův test**

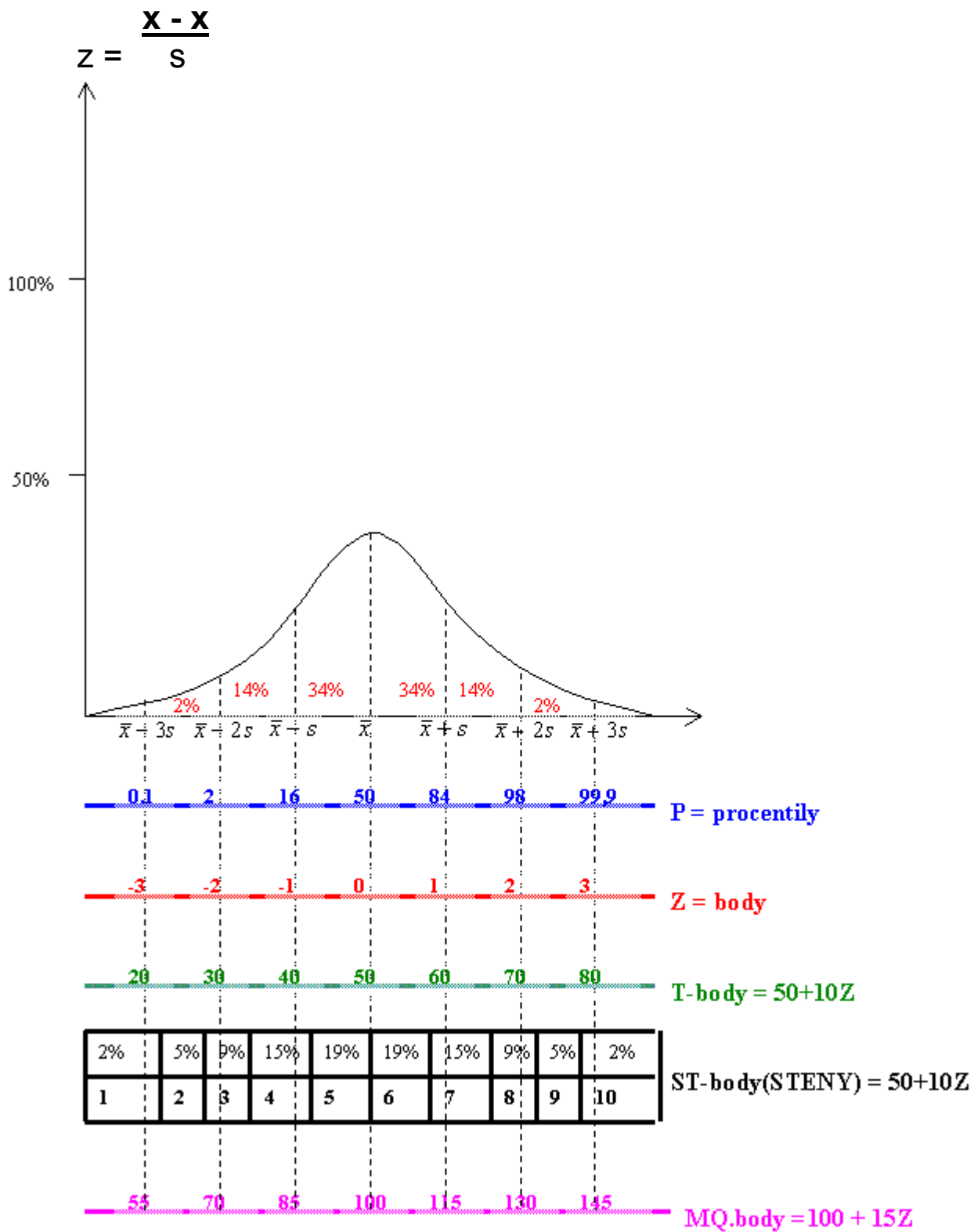
**REGRESNÍ PŘÍMKA - vede hodnotami tak, že součet čtverců odchylek je nejmenší**

**PREDIKCE - předpověď výkonů na základě dostupných dat -  
PRODLOUŽENÍ REGRESNÍ PŘÍMKY**

**MNOHONÁSOBNÁ KORELACE – PŘESNĚJŠÍ PREDIKCE NA ZÁKLADĚ  
ZNALOSTI DVOU NEZÁVISLÝCH PROMĚNNÝCH**

**Statistická významnost ještě nic neznamena,  
při měření 32 osob (vrcholoví sportovci, víc jich u nás není)  
r 0,46 stat. významné při P - 0.99, koef. determinace – ( $r^2$ )  
je cca 0,19 = cca  
20%, které umíme vysvětlit !!!!!**

#### 4. Vytvoření bodových stupnic pomocí Z-bodů:

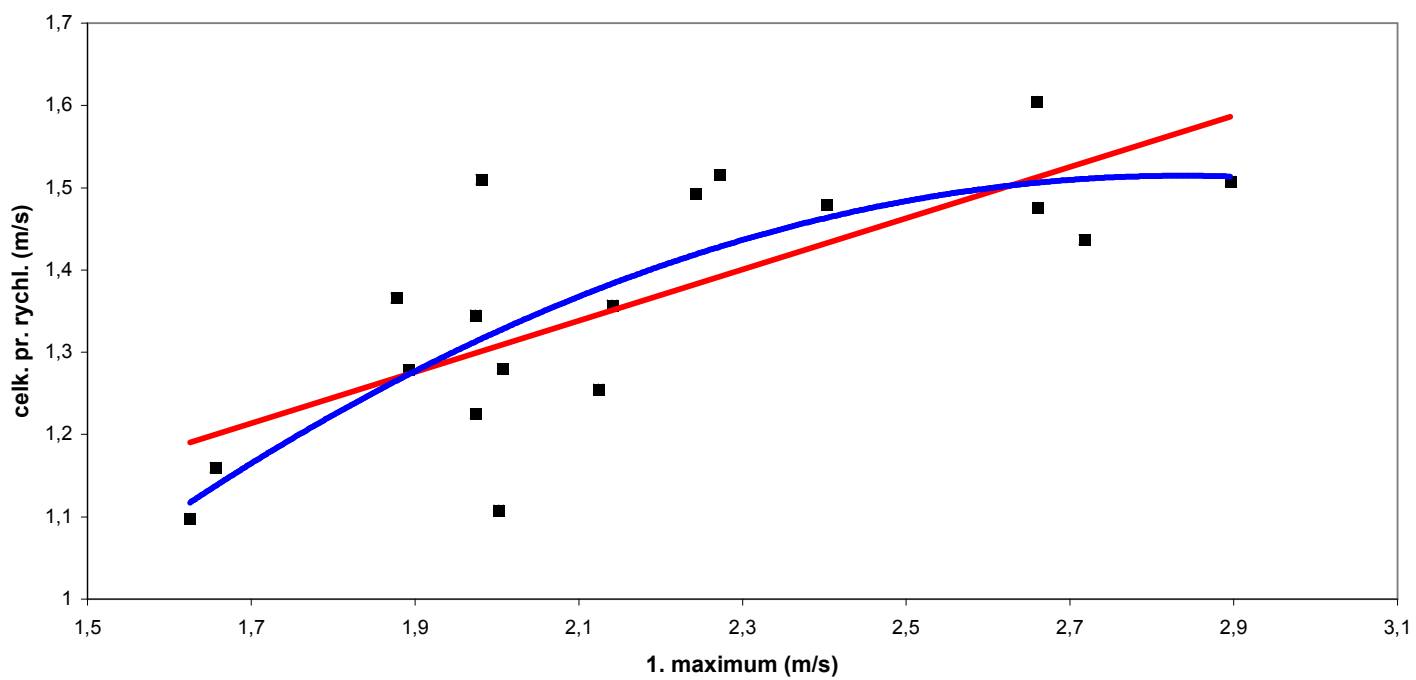


## Dvouvýběrový F-test pro rozptyl

	<i>Soubor 1</i>	<i>Soubor 2</i>
Stř. hodnota	5	4,666666667
Rozptyl	7,5	6,25
Pozorování	9	9
Rozdíl	8	8
<b>F</b>	<b>1,2</b>	
P(F<=f) (1)	0,401385866	
<b>F krit (1)</b>	<b>3,438103136</b>	

## Dvouvýběrový t-test s rovností rozptylů

	<i>Soubor 1</i>	<i>Soubor 2</i>
Stř. hodnota	5	4,666666667
Rozptyl	7,5	6,25
Pozorování	9	9
Společný rozptyl	6,875	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	16	
<b>t stat</b>	<b>0,269679945</b>	
P(T<=t) (1)	0,395426819	
<b>t krit (1)</b>	<b>2,58349246</b>	
P(T<=t) (2)	0,790853638	
<b>t krit (2)</b>	<b>2,920787665</b>	



■ celk. prům. rychl./1.max.

— lineární

— polynomický

$$y = 0,3114x + 0,6844$$

$$R^2 = 0,5584$$

$$y = -0,2701x^2 + 1,5331x - 0,6607$$

$$R^2 = 0,614$$