



STATISTICKÝ ROZCESTNÍK aneb CO S DATY

Martin Sebera

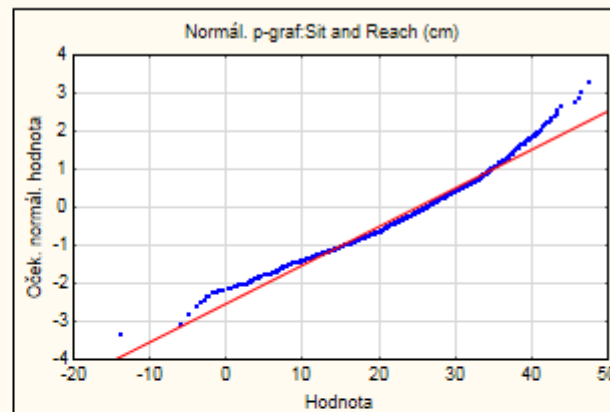
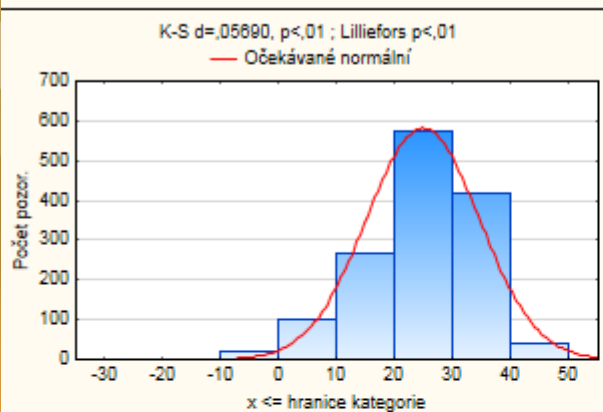
Pravidla výzkumu z pohledu analýzy dat

- 1. příprava výzkumného šetření je nejdůležitější část**
- 2. sběr a analýza dat slouží k zamítnutí/nezamítnutí předem stanovených úkolů práce a hypotéz (explorační vs. konfirmační přístup)**
- 3. vždy mít na paměti věcné hledisko výzkumu, zejména v souvislosti s interpretací statistických výsledků**

První náhled na data – popisná statistika

- průměr, sm. odchylka, medián, kvartily aj.
- četnosti: absolutní, relativní, kumulativní
- grafy: krabicový, histogram

Souhrn: Sit and Reach (cm)



Souhrnné statistiky: Sit and Reach (cm)

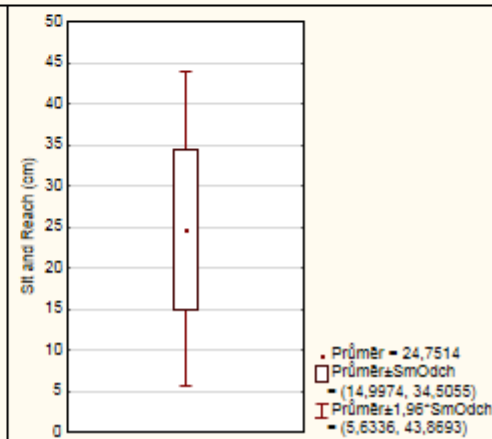
N platných=1423,000000

Průměr= 24,751441

Minimum=-14,000000

Maximum= 47,300000

Sm.odch.= 9,754011

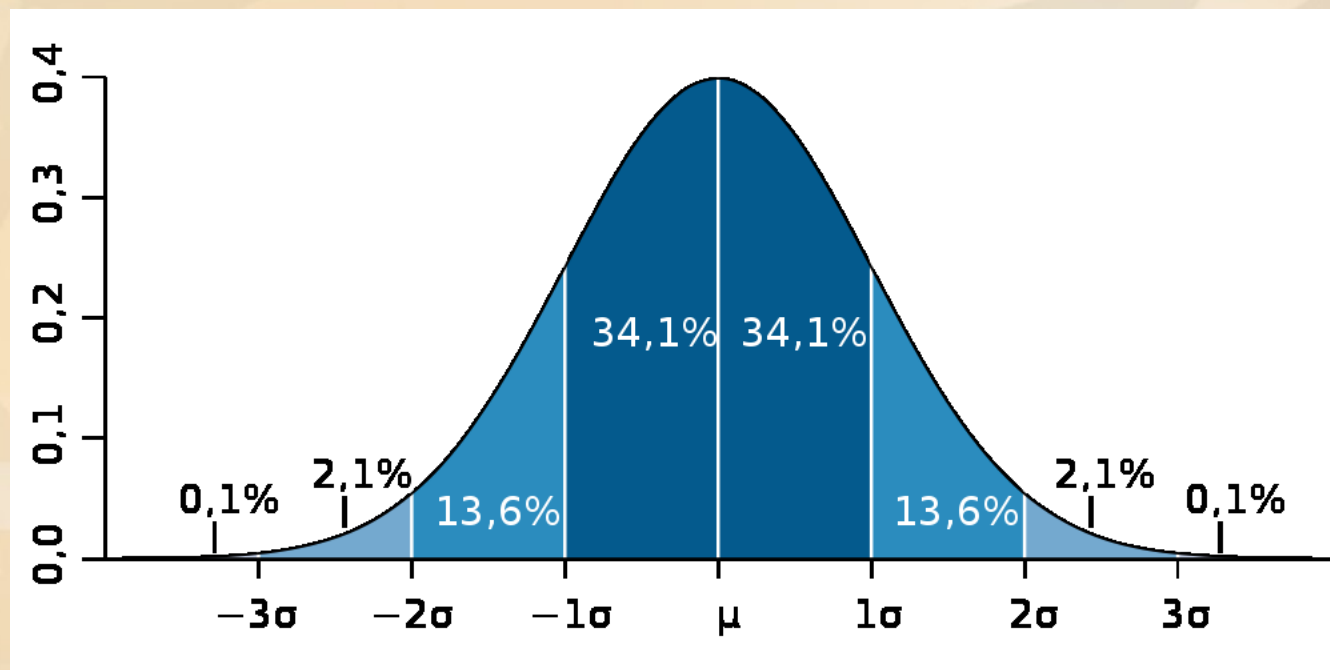


Proč?

- chybná měření, extrémny
- homogenitu souboru
- chybějící data

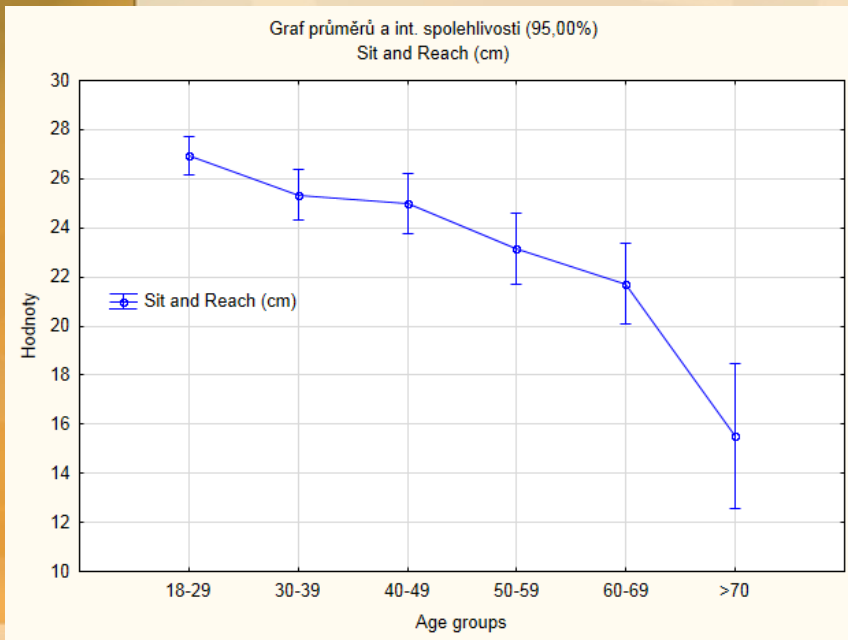
Normalita

- Kolmogorov-Smirnov a Shapiro-Wilks test
- **Proč?**
rozhodnutí, zda použít parametrické nebo
neparametrické testy



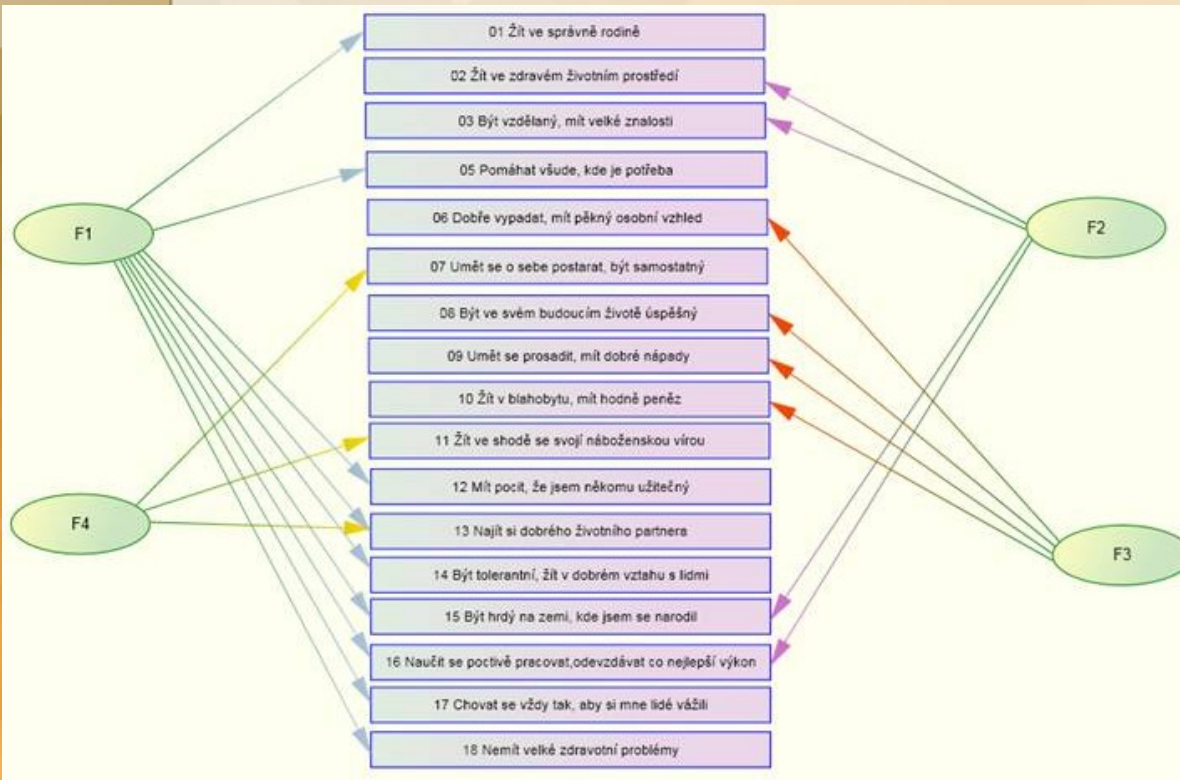
Výběry/skupiny jsou shodné nebo ne

- 2 skupiny/proměnné: t-testy
- 3 a více: Analýza rozptylu (ANOVA)
- Proč? došlo ke zlepšení výbušné síly po intervenci ? (pretest-posttest)
- suplementace sportovce (faktor A) a na způsobu tréninku (faktor B)



Závislost více proměnných (spojité)

- korelace, index determinace, faktorová analýza (lineární vztah)
- **Proč? Př. závisí výkon v běhu na 100 m s výkonem do skoku do dálky?**

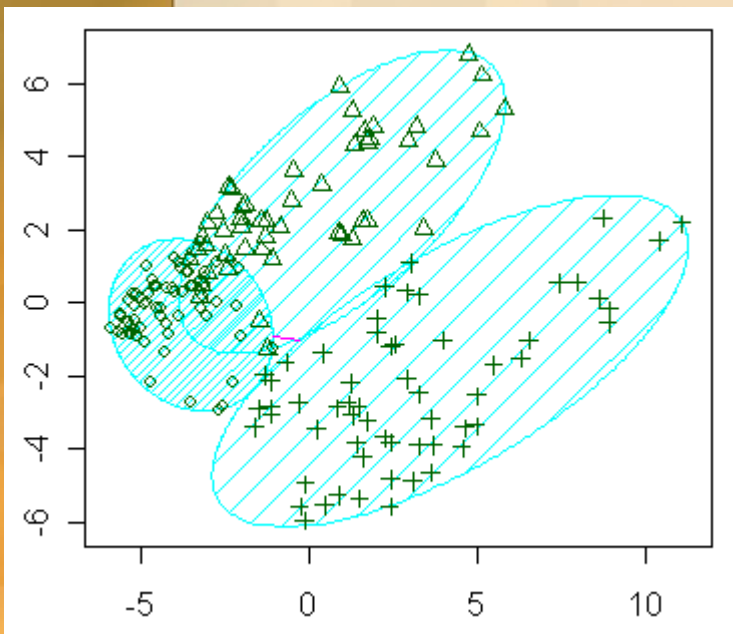


Př. závisí ekonomika běhu na povrchu?

**korelace
neznamená
kauzalitu!!!**

Závislost (kategorieální-dotazník)

- test nezávislosti chí-kvadrát v kontingenční tabulce, vícerozměrné kontingenční tabulky - asociační stromy
- Shluková analýza
- **Proč: sílu a směr vztahu**

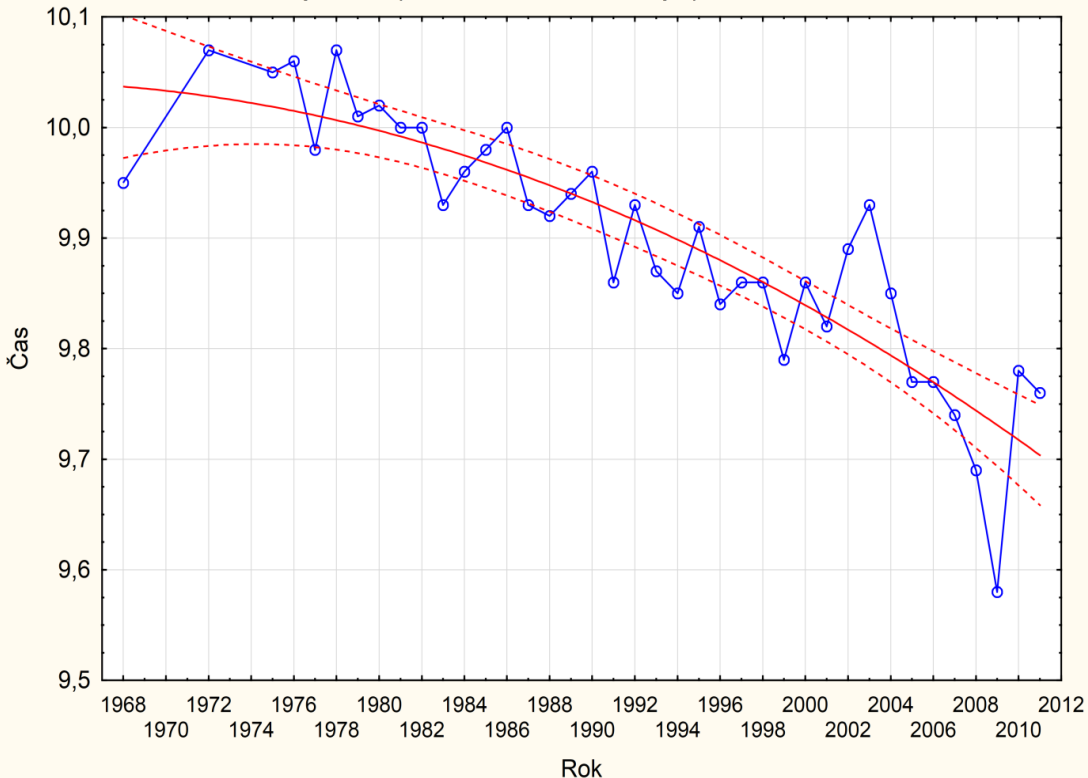


- **Př. závisí bolestivost zad na věku a způsobu zaměstnání?**
- **Př. mezi kterými proměnnými z dotazníku existuje nejsilnější vazba?**

Vytvořit model a předpověď

- Lineární regrese
- Proč? příspěvek jednotlivých nezávislých proměnných k popisu proměnné závislé

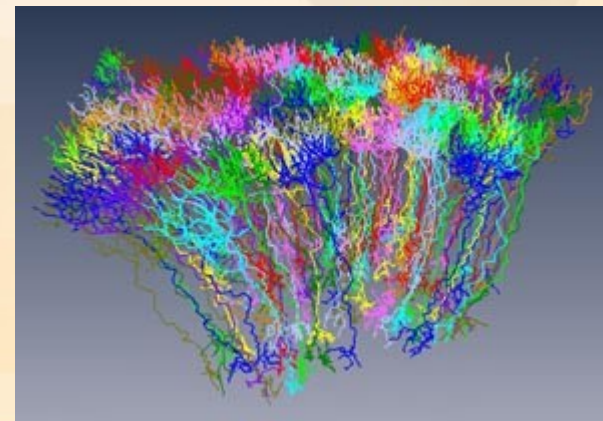
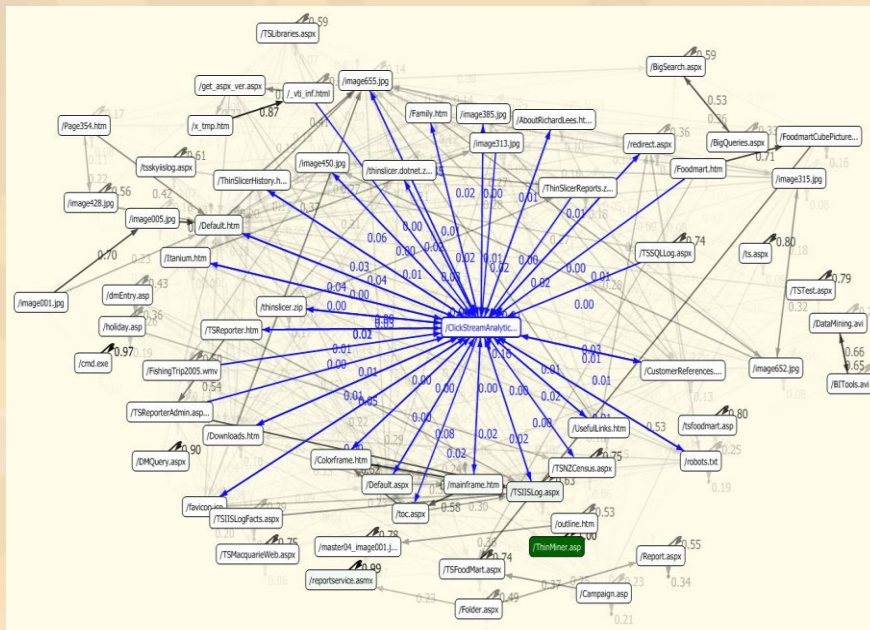
Vývoj nejlepšího výkonu v běhu na 100 m
Vyrovnání parabolou s 95 % intervaly spolehlivosti



- Př. Popsat trend výkonnosti v atletických disciplínách a provést předpověď výkonů na olympiádě v Rio 2016

A další...

- Časové řady
- Reliabilita testů
- Analýza síly testu
- Neuronové sítě



Zdroje:

- Cyhelský, L., Kahounová, J., & Hindls, R. (2001). *Elementární statistická analýza*. (2. dopl. vyd., 318 s.) Praha: Management Press.
- Hendl, J. (2006). *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. (Vyd. 2., opr., 583 s.) Praha: Portál.
- Meloun, M., & Militký, J. (1998). *Statistické zpracování experimentálních dat*. (2. vyd., xxi, 839 s.) Praha: East Publishing.
- Sebera, M. *Vícerozměrné statistiky*, 2013 (v tisku)
- Sebera, M. *Časové řady*, 2013 (v tisku)
- Zvonař, M., Pavlík, J., Sebera, M., Vespalec, T. & Štochl, J. *Vybrané kapitoly z antropomotoriky*. Brno: Masarykova univerzita, 2010.