**Statistický rozcestník ANEB co s daty**

**1. příprava výzkumného šetření je nejdůležitější část**

**2. sběr a analýza dat slouží k zamítnutí/nezamítnutí předem stanovených úkolů práce a hypotéz**

**3. vždy mít na paměti věcné hledisko výzkumu, zejména v souvislosti s interpretací statistických výsledků**

|  |  |  |
| --- | --- | --- |
| **Chci s daty provést** | **Způsob** | **Umožní mi zjistit** |
| První náhled na data | Základní popisná statistika* průměr, směrodatná odchylka, rozptyl, N, medián, kvartily a další míry polohy a variability
* tabulky četností: absolutní, relativní, kumulativní
* grafy: krabicový, histogram, bodový
 | * chybná měření, extrémy
* homogenitu souboru
* chybějící data
* trend v datech
 |
| Otestovat normalitu | * Kolmogorov-Smirnov test, Shapiro-Wilks test
 | * rozhodnutí, zda použít parametrické nebo neparametrické testy
 |
| Zjistit, zda výběry/skupiny jsou shodné nebo ne | * 2 skupiny/proměnné: t-testy
* 3 a více skupin/proměnných: Analýza rozptylu (ANOVA)
	+ T-testy i ANOVA má svou parametrickou i neparametrickou variantu!
 | * konstatovat statisticky nebo věcně (size of effect) významný rozdíl

Př. došlo ke zlepšení výbušné síly po intervenci ?(pretest-posttest)Př. která ze dvou tréninkových metod je úspěšnější?Př. mezi kterými skupinami je statisticky významný rozdílPř. byl zkoumán výsledný čas v motorickém testu v závislosti na typu suplementace sportovce (faktor A) a na způsobu tréninku (faktor B) |
| Zjistit závislost více proměnných (spojité) | * korelace, index determinace
* faktorová analýza
 | * těsnost lineárního vztahu mezi proměnnými
* může existovat jasný vztah ale nelineární, který nezachytíme pomocí korelace nebo faktorové analýzy
* korelace neznamená kauzalitu!!!

Př. závisí výkon v běhu na 100 m s výkonem do skoku do dálky?Př. závisí ekonomika běhu na povrchu? |
| Zjistit závislost více proměnných (kategoriální-např. dotazník) | * test nezávislosti chí-kvadrát v kontingenční tabulce
* vícerozměrné kontingenční tabulky - asociační stromy
* shluková analýza
* regresní a klasifikační stromy (CART, CHAID)
 | * sílu a směr vztahu

Př. závisí bolestivost zad na věku a způsobu zaměstnání?Př. mezi kterými proměnnými z dotazníku existuje nejsilnější vazba? |
| Redukovat velký počet vstupních dat | * faktorová analýza
* analýza hlavních komponent
 | * zda za naměřenými daty není nějaká latentní struktura (POZOR na interpretaci)

Př. lze 10 disciplín desetiboje popsat menším počtem faktorů? |
| Vysvětlit závislou proměnnou několika nezávislými, provést předpověď | * lineární regrese
* regresní a klasifikační stromy (CART, CHAID)
* časové řady
* neuronové sítě
 | * příspěvek jednotlivých nezávislých proměnných k popisu proměnné závislé

Př. Popsat trend výkonnosti v atletických disciplínách a provést předpověď výkonů na olympiádě v Riu 2016 |