

## 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 PROVĚST	ZPŮSOB	UMOŽNÍ MI ZJISTIT
První náhled na data	Základní popisná statistika <ul style="list-style-type: none"> <li>• průměr, směrodatná odchylka, rozptyl, N, medián, kvartily a další míry polohy a variability</li> <li>• tabulky četností: absolutní, relativní, kumulativní</li> <li>• grafy: krabicový, histogram, bodový</li> </ul>	<ul style="list-style-type: none"> <li>• chybná měření, extrémy</li> <li>• homogenitu souboru</li> <li>• chybějící data</li> <li>• trend v datech</li> </ul>
Otestovat normalitu	<ul style="list-style-type: none"> <li>• Kolmogorov-Smirnov test, Shapiro-Wilks test</li> </ul>	<ul style="list-style-type: none"> <li>• rozhodnutí, zda použít parametrické nebo neparametrické testy</li> </ul>
Zjistit, zda výběry/skupiny jsou shodné nebo ne	<ul style="list-style-type: none"> <li>• 2 skupiny/proměnné: t-testy</li> <li>• 3 a více skupin/proměnných: Analýza rozptylu (ANOVA)                             <ul style="list-style-type: none"> <li>○ T-testy i ANOVA má svou parametrickou i neparametrickou variantu!</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• konstatovat statisticky nebo věcně (size of effect) významný rozdíl</li> <li>• PŘ. došlo ke zlepšení výbušné síly po intervenci ?(pretest-posttest)</li> <li>• PŘ. která ze dvou tréninkových metod je úspěšnější?</li> <li>• PŘ. mezi kterými skupinami je statisticky významný rozdíl</li> <li>• PŘ. byl zkoumán výsledný čas v motorickém testu v závislosti na typu suplementace sportovce (faktor A) a na způsobu tréninku (faktor B)</li> </ul>
Zjistit závislost více proměnných (spojité)	<ul style="list-style-type: none"> <li>• korelace, index determinace</li> <li>• faktorová analýza</li> </ul>	<ul style="list-style-type: none"> <li>• těsnost lineárního vztahu mezi proměnnými</li> <li>• může existovat jasný vztah ale nelineární, který nezachytíme pomocí korelace nebo faktorové analýzy</li> <li>• korelace neznamená kauzalitu!!!</li> <li>• PŘ. závisí výkon v běhu na 100 m s výkonem do skoku do dálky?</li> <li>• PŘ. závisí ekonomika běhu na povrchu?</li> </ul>
Zjistit závislost více proměnných (kategoriální-např. dotazník)	<ul style="list-style-type: none"> <li>• test nezávislosti chí-kvadrát v kontingenční tabulce</li> <li>• vícerozměrné kontingenční tabulky - asociační stromy</li> <li>• shluková analýza</li> <li>• regresní a klasifikační stromy (CART, CHAID)</li> </ul>	<ul style="list-style-type: none"> <li>• sílu a směr vztahu</li> <li>• PŘ. závisí bolestivost zad na věku a způsobu zaměstnání?</li> <li>• PŘ. mezi kterými proměnnými z dotazníku existuje nejsilnější vazba?</li> </ul>
Redukovat velký počet vstupních dat	<ul style="list-style-type: none"> <li>• faktorová analýza</li> <li>• analýza hlavních komponent</li> </ul>	<ul style="list-style-type: none"> <li>• zda za naměřenými daty není nějaká latentní struktura (POZOR na interpretaci)</li> <li>• PŘ. lze 10 disciplín desetiboje popsat menším počtem faktorů?</li> </ul>
Vysvětlit závislou proměnnou několika nezávislými, provést předpověď	<ul style="list-style-type: none"> <li>• lineární regrese</li> <li>• regresní a klasifikační stromy (CART, CHAID)</li> <li>• časové řady</li> <li>• neuronové sítě</li> </ul>	<ul style="list-style-type: none"> <li>• příspěvek jednotlivých nezávislých proměnných k popisu proměnné závislé</li> <li>• PŘ. Popsat trend výkonnosti v atletických disciplínách a provést předpověď výkonů na olympiádě v Rio 2016</li> </ul>