

Statistická analýza dat

□ **Kontakt na vyučující:**

Mgr. Helena Klimusová

Psychologický ústav FF MU

A. Nováka 1, Brno

budova C – 5. patro

□ tel. 549494035

□ e-mail klimusov@fss.muni.cz

□ konzultační hodiny v JS 2004:
pondělí 14.00-15.00

Statistická analýza dat

- cíl kurzu
 - organizace výuky
 - osnova kurzu
 - doporučená literatura
 - hodnocení
-

Cíl kurzu

- ❑ obeznámit se s metodami a základními postupy statistické analýzy
 - ❑ naučit se používat statistický software SPSS
 - ❑ získat znalosti potřebné pro popis, analýzu a interpretaci výsledků kvantitativního výzkumu
 - ❑ tyto znalosti aplikovat při analýze dat bakalářských a diplomových prací
 - ❑ naučit se číst a porozumět metodám a výsledkům statistické analýzy ve výzkumných zprávách publikovaných v odborném tisku
-

Doporučená literatura

- ❑ Healey, Joseph F. *Statistics: A Tool for Social Research*. 3. vyd. Belmont, California: Wadsworth, 1993.
 - ❑ Disman, Miroslav. *Jak se vyrábí sociologická znalost*. Praha: Karolinum, 1993.
 - ❑ Svoboda, Helmut. *Moderní statistika*. 1. vyd. Praha: Svoboda, 1977.
 - ❑ Gregor, Miroslav, Přadková, Svatomíra, Spěšná, Daniela. *Statistika pro sociology*. 2. vyd. Brno MU 1993.
-

Statistika - úvod

- vymezení statistiky
 - úloha statistiky v psychologickém výzkumu
 - základní pojmy - měření, proměnné (diskrétní a spojité proměnné; podle úrovně měření); popisná a indukční statistika; populace a vzorek
 - příprava dat před analýzou
-

Definice statistiky

- Statistika je vědní obor zabývající se zkoumáním jevů, které mají hromadný charakter.
 - Statistika je soubor metod, které nám umožňují rozhodnutí v případě nejistoty.
-

Definice statistiky

- Analýza dat náhodného charakteru, která předpokládá použití pokročilejšího aparátu matematiky a určitého matematického modelu (teorie pravděpodobnosti).
-

Definice statistiky

- *Statistická analýza jsou záhadné, někdy až prapodivné manipulace s daty získanými pomocí experimentu, jejichž cílem je zastrčit ten fakt, že výsledky nemají pro lidstvo žádný zobecnitelný význam. Obvykle jsou při ní používány počítače, což této proceduře propůjčuje další auru nereálnosti.*
-

Role statistiky ve výzkumu

- při výzkumu v sociálních vědách jsou střídavě používány deduktivní a induktivní postupy
 - statistika se uplatňuje u induktivních postupů
 - (deduktivní postupy – z teorie jsou odvozeny hypotézy a z nich postup pozorování/měření; induktivní postupy: z pozorování jsou odvozeny obecné závěry, které mohou modifikovat teorii atd.)
-

Proměnné

- při výzkumu psychologové získávají, měří a analyzují **proměnné**
-

Proměnné

- při výzkumu psychologové získávají, měří a analyzují **proměnné**
 - proměnná je objekt, který může nabývat různých hodnot (na rozdíl od konstanty)
-

Proměnné

- jsou rozlišovány různé **typy proměnných**
-

Proměnné

- jsou rozlišovány různé **typy proměnných**
 - podle jejich role ve výzkumném plánu
-závislé, nezávislé, intervenující
 - kvalitativní (kategoriální) a kvantitativní (ty dále na spojité a diskrétní)
 - podle úrovně měření – nominální, pořadové, intervalové a poměrové
-

Závislé a nezávislé proměnné

- příklad experimentu: chceme zjistit účinnost prostředku na zlepšení paměti u seniorů ve srovnání s placebem
-

Závislé a nezávislé proměnné

- **nezávislá** proměnná – podávaný prostředek (testovaný lék x placebo)
 - **závislá** proměnná – výkon v testu paměti (např. počet vybavených slov)
 - možné **intervenující** proměnné – ochota užívat léky, podmínky při testu
-

Závislé a nezávislé proměnné

- jiný příklad: Milgramův experiment s poslušností vůči autoritě

Závislé a nezávislé proměnné

- **nezávislá** proměnná – viditelnost a slyšitelnost žáka a přítomnost autority
 - **závislá** proměnná – nejvyšší hodnota uděleného šoku
 - možné **intervenující** proměnné – pohlaví učitele/žáka atd.
-

Závislé a nezávislé proměnné

- **nezávislá** proměnná – ta, se kterou experimentátor manipuluje (independent variable)
 - **závislá** proměnná – proměnná měřená experimentátorem (dependent variable)
 - **intervenující** proměnná – e. se snaží její vliv eliminovat (znáhodňováním, vyrovnáváním) (intervening variable)
-

Kvalitativní a kvantitativní data

- ❑ **kvalitativní (kategoriální)** – zjišťujeme hodnotu znaku – kategorii (např. pohlaví, náboženské vyznání, rodinný stav, barva očí, vzdělání); často fungují jako nezávislé proměnné (faktory); (categorical data)
 - ❑ **kvantitativní (metrická) data** – zjištěná měřením pomocí nějakého nástroje (v širším smyslu); skór v testu, tělesná váha...; obvykle závislé proměnné (measurement data)
-

Úroveň měření

- kategoriální proměnné jsou měřeny buď na nominální nebo pořadové úrovni
 - **nominální** – čísla jsou hodnotám proměnné přiřazena náhodně; pouze hodnoty označují, ale nelze s nimi jako s čísly zacházet (jediné operace jsou = a \neq) (nominal scales)
-

Úroveň měření

- **pořadová** (ordinální) – hodnoty je možno uspořádat podle velikosti (např. pořadí v závodě, postojové škály), ale není možno stanovit, o kolik se liší (ordinal scales)
 - možné početní operace jsou stále $=$ a \neq , ale také $<$ a $>$
-

Úroveň měření

- **intervalová** – hodnoty je možno uspořádat a vzdálenosti mezi nimi (intervaly) jsou shodné (např. rok narození, teplota) (interval scales)
 - možné početní operace: $=$ a \neq , $<$ a $>$, $+$ a $-$
-

Úroveň měření

- **poměrová** – stejné vlastnosti jako intervalová a navíc se mezi hodnotami vyskytuje přirozená 0 (indikující absenci znaku) – tj. má smysl se ptát, kolikrát je hodnota větší než jiná hodnota (ratio scales)
 - např. věk, počet správně řešených úloh, počet dětí
 - kromě předchozích početních operací je možné také násobení a dělení
-

Úroveň měření

- kvantitativní data je možno zredukovat na kategoriální (např. místo hodnoty krevního tlaku údaj normotenze x hypertenze)
 - někdy je obtížné stanovit hraniční hodnoty (cut-off scores)
-

Účel typologie proměnných

- má pomoci při rozhodování o tom, jak nejlépe zobrazit, shrnout či analyzovat data
-

Populace a výběr

- data mohou být získávána na populaci nebo na vzorku populace
-

Populace a výběr

- ❑ **populace** (základní soubor) - úplný souhrn lidí, objektů nebo věcí, které jsou předmětem našeho zkoumání
 - ❑ je dána přesným stanovením jeho prvků
 - ❑ prvky mohou být určeny buď jejich výčtem nebo vymezením některých společných vlastností (např. územní příslušnost, věk atd. – např. všechny děti s diagnostikovanou poruchou pozornosti a trvalým bydlištěm v JM kraji)
-

Populace a výběr

- **výběr** (výběrový soubor, vzorek) – výzkumníci se snaží, aby byl reprezentativní vůči příslušné populaci (pomocí postupů náhodného výběru)
 - stejný soubor objektů může za určitých okolností představovat jednu populaci a jindy výběr
-

Parametry a statistiky

- pokud provádíme měření na celé populaci, jsou výsledky nazývány **parametry** populace
 - bývají označovány řeckými písmeny (např. průměr μ)
-

Parametry a statistiky

- výsledky z měření na vzorku se nazývají **statistiky**
 - většinou jsou užívány k odhadu populačních parametrů pomocí postupů **statistické indukce (inference)**
-

Popisná a indukativní statistika

- **popisná (deskriptivní) statistika** – třída technik, které slouží k popisu proměnných (např. typická hodnota proměnné, distribuce hodnot atd.)
 - **induktivní (inferenční) statistika** – umožňuje rozhodnutí, zda zjištění ze vzorku platí i na populaci (a s jakou pravděpodobností) – odhady intervalu spolehlivosti, testování hypotéz
-

Popisná a indukční statistika

- indukční statistika se dále dělí na **parametrické** a **neparametrické** postupy
 - parametrickými testy ověřujeme platnost hypotézy o parametrech základního souboru (např. průměr) – vyžadují měření na intervalové úrovni
-

Příprava dat před analýzou

- uspořádání dat
 - kontrola přesnosti dat
 - vytvoření struktury databáze
 - vkládání dat do PC
 - transformace dat
-

Uspořádání dat

- výzkumná data mohou pocházet z různých zdrojů: dotazníky, rozhovory, měření při experimentech (pretest, posttest), pozorování
 - je třeba všechny formuláře, dotazníky, záznamové archy atd. řádně označit, aby je vždy bylo možno i zpětně dohledat
-

Uspořádání dat

- původní data se u profesionálních výzkumů po určitou dobu **archivují** (min. 5-7 let)
 - je vhodné vytvořit si databázi – **přehled** všech použitých zdrojů dat (např. v MS Access nebo Excel, Statistica, SPSS)
-

Kontrola přesnosti dat

- někdy vhodné provést již při sběru dat – pak je možno příp. chybějící nebo jinak problematické odpovědi vyřešit na místě
 - **kontrola:** čitelnosti odpovědí, úplnosti odpovědí, důležitých informací (datum, jméno výzkumníka atp.)
 - je vhodné z dalších kroků vyloučit nevalidní odpovědi (na všechny otázky stejné atd.)
-

Vytvoření struktury databáze

- ❑ data je možno vkládat přímo do statistických programů nebo do databázových programů (a potom je převést)
 - ❑ rozsáhlé výzkumné projekty mají vytištěnu podrobnou kódovací příručku (*codebook*) s popisem proměnných
-

Vytvoření struktury databáze

pro každou proměnnou:

- ❑ jméno proměnné (např. *pohlaví*)
 - ❑ popis proměnné (např. *celkový skór dotazníku deprese*)
 - ❑ popis hodnot proměnné (např. *1=chlapci, 2=dívky*)
 - ❑ způsob kódování chybějících hodnot (např. 9)
 - ❑ formát proměnné (numerický, textový, datum...)
-

Vkládání dat do PC

- pro **kontrolu přesnosti vkládání** existují nejrůznější postupy
 - např. dvojitě vkládání (*double entry*)
 - speciální program, který při druhém vkládání porovnává data s původním údajem a upozorní na chyby; není příliš rozšířeno
-

Vkládání dat do PC

- ❑ jednodušší je náhodně zkontrolovat určitý počet případů (při velkém procentu chyb kódovat znovu a porovnat)
 - ❑ vždy je nutno alespoň zkontrolovat, zda rozsah zadaných hodnot odpovídá předpokladům
-

Transformace dat

- **obrácení položek** –některé položky bývají formulovány v opačném směru než zbytek škály;
 - jejich hodnoty je třeba překódovat (např. u 5 bodové škály se bude 1 rovnat 5, 2 bude 4 atd. až 5 bude 1)
-

Transformace dat

- vzorec pro obrácení škály

nová hodnota = nejvyšší hodnota + 1 - stará hodnota

- v SPSS je možno provést pomocí příkazů COMPUTE nebo RECODE
-

Transformace dat

- **výpočet celkových skóre** – většinou součtem hodnot několika proměnných (příp. průměr atp.)
 - v příkazu COMPUTE se zadá příslušný vzorec,
např. pro součet **celk = p1 + p2 + p3**
 - nebo rovnou funkce (*sum*, *mean*)
celk = sum (p1 TO p3)
-

Transformace dat

- ❑ **sloučení hodnot proměnných** –
např. ze 7 bodové škály chceme udělat 3 bodovou; nebo z věku zadaného v letech vytvořit 3 věkové skupiny
 - ❑ příkaz RECODE
-

Kontrolní otázky

- ❑ klasifikace proměnných podle jejich role ve výzkumném plánu
 - ❑ rozdíly mezi nominální, ordinální, intervalovou a poměrovou úrovní měření; možné početní operace
 - ❑ vymezení populace a výběru, parametrů a statistik
 - ❑ vymezení popisné a indukční statistiky
 - ❑ postup přípravy dat před analýzou
-