



# Statistika pro sociální pracovníky

# Obsah:

- 1. Úvod do statistické analýzy
- 2. Frekvenční tabulky a grafy
- 3. Centrální tendence a variabilita
- 4. Normální distribuce
- 5. Úvod do testování hypotéz
- 6. Populace a vzorek, statistická inference
- 7. Výběr statistického testu
- 8. Korelace
- 9. Regresní analýza
- 10. Kontingenční tabulky (křížení)
- 11. T-test a analýza rozptylu



# Kapitola 1: Úvod do statistické analýzy

- Proč studovat statistiku? (3)
- Použití statistické analýzy (4)
- Základní metodologické pojmy (5-)
- Úrovně měření
- Úrovně měření a analýza dat
- Další klasifikace měření
- Kategorie (druhy) statistické analýzy
- Shrnutí a otázky pro studium

# Proč studovat statistiku v sociální práci?

- Sociální pracovníci (SoPci) provádějí výzkumné studie (jak sbírat, třídít, organizovat data a učinit závěry)
- SoPci přispívají ke znalostní základně profese šířením nálezů svých výzkumů (prokázat důvěryhodnost zjištění)
- SoPci se v praxi opírají o výsledky cizích výzkumů (Porozumění výsledků statistické analýzy zvyšuje pravděpodobnost zvolení správné intervence)
- SoPci potřebují zhodnotit efektivitu své intervence (efektivita a odpovědnost intervence)



# Použití statistické analýzy

- Výběr, design a úprava nástrojů sběru dat
- Výběr a velikost výzkumného vzorku
- Zhodnotit zda nástroje měření produkují validní a reliabilní výsledky
- Shrnout vlastnosti specifického výzkumného vzorku nebo populace
- Odhadnout vlastnosti populace, z které jsme vybrali vzorek
- Rozhodnout zda vztahy a souvislosti nalezené ve vzorku mohou být zobecněny na populaci

# Základní metodologické pojmy

- Populace vs. Vzorek
- Deskriptivní vs. Inferenční statistika
- Data (6)
- Informace (7)
- Proměnné a konstanty (8-9)
- Konceptualizace (9)
- Operacionalizace (10)
- Reliabilita (konzistence a opakovatelnost) (11)
- Validita (platnost) (12)
- Výzkumná hypotéza (13)

# Vzorek vs. populace

- Sbíráme informace (data) o vzorku nebo o populaci
- Oba termíny nepředstavují jen lidi nebo události, ale i sadu naměřených hodnot o těchto lidech / událostech
- naměřené hodnoty všech klientů obslužených v průběhu jednoho roku v konkrétním zařízení soc. služeb = populace
- Vzorek = podskupina hodnot z větší populace hodnot. Údaje o vybrané polovině klientů obslužených v průběhu jednoho roku

# Deskriptivní vs. Inferenční statistika

- Deskriptivní (=popisná) = používáme pokud je našim cílem pouze popsat a interpretovat sadu hodnot (data) která držíme v ruce (ať už vzorek nebo populaci)
- Inferenční (=deduktivní, usuzovací) = používáme pokud je naším záměrem použít data která držíme k tomu, abychom se pokusili porozumět většímu množství dat (děláme odhady o vlastnostech populace na základě vzorku).



# Data (pl.), datum (s.)

- = Naměřené hodnoty sesbírané při výzkumu ještě před tím než byly jakkoli analyzovány (př. skóre škály spokojenosti klientů)
- Nástroje sběru dat = např. rozhovory, obsahové analýzy, přímé a zúčastněné pozorování atd.
- Analýza dat posbíraných někým jiným pro jiné účely: analýza sekundárních dat
- Kvalitní závěry vyžadují kvalitní data

# Informace

- = Interpretace analyzovaných data
- Př. „Při snižování zneužívání látek je intervenční metoda A u účastníků výzkumu úspěšnější než metoda B“
- nebo „je velmi teplo“
- oproti tomu 102° F je datum

# Proměnná vs. konstanta

- Při výzkumu limitujeme sběr dat pouze na takové data, které jsou nezbytné a užitečné pro naši výzkumnou otázku
- = vlastnost která se liší v kvantitě nebo kvalitě u lidí které zkoumáme
- Př. úroveň vzdělání, pohlaví, sexuální orientace, etnicita, úroveň sebedůvěry atd.
- Konstanta = vlastnost která se neliší v kvantitě nebo kvalitě u lidí které zkoumáme
- Př. Smrtelnost (lidé) nebo národnost (Češi)
- Př. Vzorek adolescentních dívek závislých na kokainu

# Kategorie a hodnoty proměnných

- Různá změřením proměnných mohou být vyjádřena buďto slovně nebo čísly
- Kategorie proměnné (categories) = různé formy které proměnná může nabývat, vyjádřené slovně př. Pohlaví: muž / žena
- Hodnoty proměnné (values) = různá změřením proměnné vyjádřená čísly, která reflektují kvantifikovatelný rozdíl př. Věk: 20, 24, 60

# Frekvence kategorií a hodnot

- = Četnost (=kolikrát) s jakou se vyskytuje daná hodnota nebo kategorie ve zkoumaném skupině osob (vzorku)
- Příklad: ve skupině máme 12 mužů a 16 žen: frekvence pro kategorii „muž“ proměnné „pohlaví“ je 12, zatímco frekvence pro kategorii „žena“ stejné proměnné je 16

# Konceptualizace

- 1. Vybrat nedůležitější proměnné vzhledem k výzkumné otázce (důkladně prostudovat literaturu!)
- 2. Definovat co přesně každá proměnná znamená
- 3. Načrtnout druh vztahu mezi proměnnými

Př. Je intervence A lepší než intervence B ve snižování autistického chování tříletých dětí u nichž byl diagnostikován autismus? (typ a obsah intervence, snižování?, projevy autistického chování, závažnost?, Intervence ovlivňuje chování)

# Operacionalizace

- = specifikovat jak přesně budeme měřit proměnné které jsme konceptualizovali
- Jedna proměnná – často více indikátorů
- Př. Úroveň asertivity:
  - a) specifické chování klienta pozorovaného doktorem v různých podmínkách
  - b) subjektivní diagnóza klienta sociálním pracovníkem
  - c) sebe-analýza klientem
  - d) klientovy odpovědi na dotazník, který je ve vědecké komunitě považovaný za důvěryhodně měřící asertivitu
- Př. Úroveň spokojenosti se zaměstnáním...

# Reliabilita

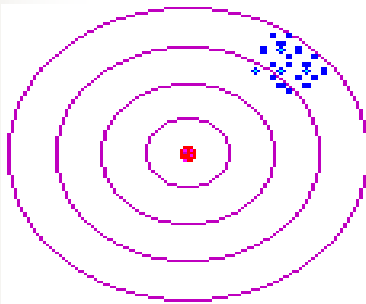
- = stupeň konzistence a opakovatelnosti měření, stupeň chybovosti měření, podíl skutečného skóre na naměřeném skóre, spolehlivost (unidimenzionalita) škály
- Př. Nástroj který měří klinickou depresi a vykazuje konzistentní výsledky napříč lidmi a v různých situacích, oproti tomu jiný nástroj může být ovlivněn tím kdo jej používá, prostředím, časem....



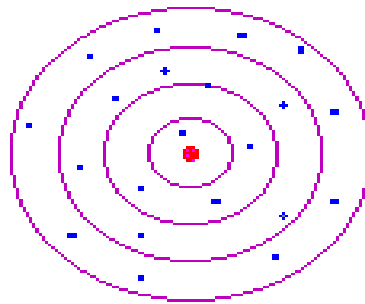
# Validita

- = měříme skutečně to co věříme že měříme (validita měření/konstruktu) X interní validita (design), validita výsledku (analýza), externí validita (výběr vzorku)
- Samotná reliabilita nástroje neznamena že je nástroj správný
- Např. Nástroj kteří měří klinickou depresi může opakovaně udávat konzistentní výsledky ale ty mohou být opakovaně špatné (např. měřit úroveň sebevědomí, energie nebo fyzického zdraví)

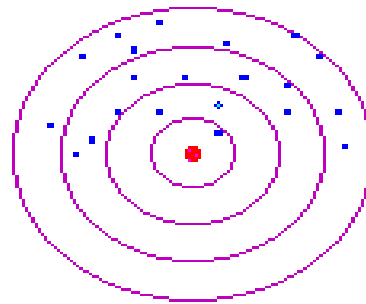
# Vztah mezi reliabilitou a validitou



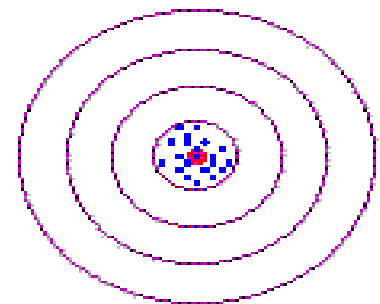
**Reliable  
Not Valid**



**Valid  
Not Reliable**



**Neither Reliable  
Nor Valid**



**Both Reliable  
And Valid**

|        |           | Concept  |   |
|--------|-----------|--|---|
|        |           | same   | different   |
| Method | same      | <b>reliability</b><br>verbal = verbal<br>written = written | <b>discriminant</b><br>verbal = math<br>written = written       |
|        | different | <b>convergent</b><br>verbal = verbal<br>written = observed | <b>very discriminant</b><br>verbal = math<br>written = observed |

# Výzkumná hypotéza

- = tvrzení o vztahu mezi proměnnými
- Formulace v budoucím čase protože předpovídá co najdeme (co očekáváme) když zanalyzujeme data výzkumu
- Způsob jakým je stanovena umožňuje její ověření statistickou analýzou
- Př. H1: **Lidé kteří jsou v depresi budou mít jinou kvalitu spánku než lidé bez depresí** (*testujeme tak že změříme kvalitu spánku operacionalizovaného např. jako počet hodin spánku lidí s diagnosou deprese a porovnáme s kvalitou spánku lidí u kterých nebyla diagnostikována deprese*)
- POZOR: H1 neříká nic o podstatě/směru vztahu (zda deprese ovlivňuje spánek či naopak), ani o jeho formě (negativní, pozitivní vztah, přímý/nepřímý, relace/kauzalita), pouze souvislost mezi proměnnými

## Výzkumná hypotéza 2: (ne)závislá proměnná

- Alter. H1: **Narušený spánek způsobuje depresi**
- Navrhuje přímý, kauzální vztah, jedna proměnná ovlivňuje druhou
- Nezávislá (independent, explaining, predictor) proměnná = ovlivňující proměnná, ta s kterou manipulujeme  
(Př. Pravý experiment (druh výzkumného designu): narušíme spánek několika lidí abychom zjistili jestli dostanou deprese (nelze z etického hlediska))
- Závislá (dependent, explained, criterion) proměnná = předpokládá se že je ovlivňována nezávislou proměnnou (deprese ovlivňovaná narušeným spánkem)

# Úrovně měření

- Po konceptualizaci, operacionalizaci a stanovení výzkumné hypotézy je třeba určit, jak přesně byly proměnné měřeny
- Snaha o maximální přesnost, někdy není vhodné z etického nebo praktického hlediska (př. výše příjmu( návratnost))
- Určení úrovně měření proměnné – nezbytnost k výběru statistické metody analýzy
- Čtyři úrovně měření: nominální, ordinální, intervalová, poměrová

# Nominální (nominal)

- Proměnná je rozdělena do kategorií (podtříd), které jsou nespojité (vzájemně se odlišují)
- Mezi podtřídami neexistují kvantifikovatelné rozdíly, pouze kvalitativní – tudíž nelze řadit podle velikosti
- Příklad: Pohlaví, rasa, rodinný status
- Hodnoty (čísla) přidělená kategoriím reflektují pouze kvalitativní rozdíly (Příklad: Jaké je vaše pohlaví? 1. muž 2. žena)
- Podmínky: kategorie musí být rozdílné, vzájemně neslučitelné (každý respondent pasuje pouze do jediné kategorie) a vyčerpávající (pro každého respondenta je k dispozici odpověď/kategorie)

# Ordinální (ordinal)

- Kategorie proměnné představují nejen rozdílnost, ale mají i kvantitativní význam – lze je smysluplně seřadit (*rank-order*) od největší po nejmenší.
- Příklad: Stupeň vzdělání, účinnost intervence, úroveň klientovy spokojenosti se službou (škály)
- Příklad: Otázka: Jak byste ohodnotil práci svého sociálního pracovníka? 1. velmi dobrý, 2. Dobrý, 3. Slušný, 4. Slabý, 5. velmi slabý nebo Jaký je váš nejvyšší stupeň vzdělání v oboru sociální práce? 1. Vyšší odborný, 2. Bc, 3. Mgr., 4. PhD
- Pozor: protože intervaly mezi jednotlivými kategoriemi nejsou stejné, nelze říci, že rozdíl mezi 1. a 3. je stejný jako mezi 2. a 4.

# Intervalová (interval)

- Umisťuje hodnoty proměnné na pravidelně rozdělené kontinuum, kde jednotlivé hodnoty jsou od sebe stejně vzdáleny – má shodnou jednotku měření např. jeden rok, teplotní stupeň
- Můžeme říct nejen že jeden respondent má určité proměnné více než druhý (ordinální) ale i o kolik
- Rozdíl mezi 1 a 4 je významově stejný jako mezi 3 a 6, a sice 3. Rozdíl v inteligenci mezi IQ 105 a 100 je stejný jako mezi IQ 125 a 120.
- Ale: nemá absolutní nulu - nelze identifikovat bod, ve kterém neexistuje žádné (nulové) množství proměnné, a proto nelze říct, že 2 je dvakrát více než 1, ale pouze o jednotku více  
Př. Protože 0 stupňů Farenheita neznamena absenci tepla, nelze říct že při 60 stupních F je dvakrát tepleji než při 30 stupních F.
- Aritmetické operace: sčítání a odčítání



# Poměrová (ratio)

- Existence fixní, absolutní a nelibovolně zvolené nulové hodnoty
- Hodnoty proměnné představují skutečné množství vlastnosti
- Můžeme říct nejen o kolik má jeden respondent více jednotek vlastnosti než druhý, ale také kolikrát více
- Příklad: Věk, počet dětí v rodině, počet případů kolikrát klient absolvoval skupinovou terapii
- Absolutní nula umožňuje všechny aritmetické operace:  
+, -, \*, /
- Příklad: Země s  $\emptyset$  porodností 4 děti/pár má 2x větší porodnost než země s  $\emptyset$  porodností 2 děti/pár
- Intervalová nebo poměrová? Pokud je logicky možné jít do mínusu, pak intervalová, pokud nelze, pak poměrová.

# Úroveň měření a analýza dat

- Jak přesně je proměnná naměřena ovlivňuje možnosti použití statistických technik
- Někdy ač proměnná vypadá jako intervalová, je pouze ordinální. Př. Schopnost řídit auto operacionalizovaná jako počet dopravních nehod za 10 let: poměrová?  
Ne, protože ačkoli má opravdovou nulu a stejně veliké intervaly, tak nelze říci, že rozdíl ve schopnostech mezi osobou která měla 4 nehody a osobou se 3 nehodami je stejný jako rozdíl mezi osobou s 1 nehodou a žádnou nehodou. Co dělat? Lze vytvořit skupiny kde 0-2 nehody, 3-5 atd. a užívat proměnnou jako ordinální.
- Přesnější proměnné lze transformovat na méně přesné (př. Věk na Věk\_skupiny (Starší a Mladší)), ne však opačně.

# Jiné klasifikace proměnných

- Nespojité (discrete) = může nabývat jen konečné množství hodnot př. Počet sourozenců, člověk může mít 2 nebo 3 sourozence, ale ne 2,16
- Spojité (continuous) = může teoreticky nabývat jakékoli hodnoty př. výška studentů soc.práce. Mezi jakýmkoli dvěma naměřenými hodnotami může být teoreticky další hodnota

# Jiné klasifikace proměnných (2)

- Dichotomické (dichotomous) =druh nespojité proměnné, která má pouze dvě kategorie př. Pohlaví (muž nebo žena) nebo výsledek voleb (vyhrál nebo prohrál)
- Binární (binary) = druh dichotomické proměnné, kdy připisujeme hodnoty 1 přítomnosti a 0 absenci proměnné př. Vlastnictví auta (1= mám auto, 0 = nemám)
- Falešné (dummy) = druh dichotomické proměnné, kdy nominální nebo ordinální proměnnou transformujeme do takového množství proměnných, kolik má původní proměnná kategorií (k) (V analýze však použijeme o jednu méně (k-1) proměnných).

Př. Pohlaví transformujeme do muže (=1 pokud muž, =0 pokud žena), nebo do ženy (=1 pokud žena, 0 = pokud muž)

Užití dummy:

- 1) když chceme použít nominální proměnnou v technikách vyžadující proměnné vyšší úrovně (např. lineární regrese)
- 2) když chceme testovat nelinearitu vztahu (efektu)

# Druhy statistické analýzy

- Podle počtu proměnných
  - a) univariační/jednorozměrná (univariate) = zkoumá distribuci hodnot jediné proměnné
  - b) bivariační/dvourozměrná (bivariate) = zkoumá vztah mezi dvěma proměnnými
  - c) multivariační/vícerozměrná (multivariate) = zkoumá vztah mezi třemi a více proměnnými zároveň

# Druhy statistické analýzy (2)

## ■ Podle účelu

### ■ a) deskriptivní

- = pomocí tabulek a grafů shrnuje základní informace ze syrových (původních) dat a pomáhá tak jejich interpretaci a komunikaci navenek
- nepřekračujeme hranice vzorku

### ■ b) inferenční

- = určuje jak bezpečné je zobecňovat (odhadovat, usuzovat) výsledky ze vzorku (statistiky) na charakteristiky populace (parametry)
- - jak je pravděpodobné že výsledek ze vzorku platí i v populaci