

# Kvantitativní přístupy v politologii

POL 593

# Statistika?

Andrew Lang o politikovi:

“Používá statistiku jako opilý člověk pouliční lampu – spíš na podporu než na osvětlení.”

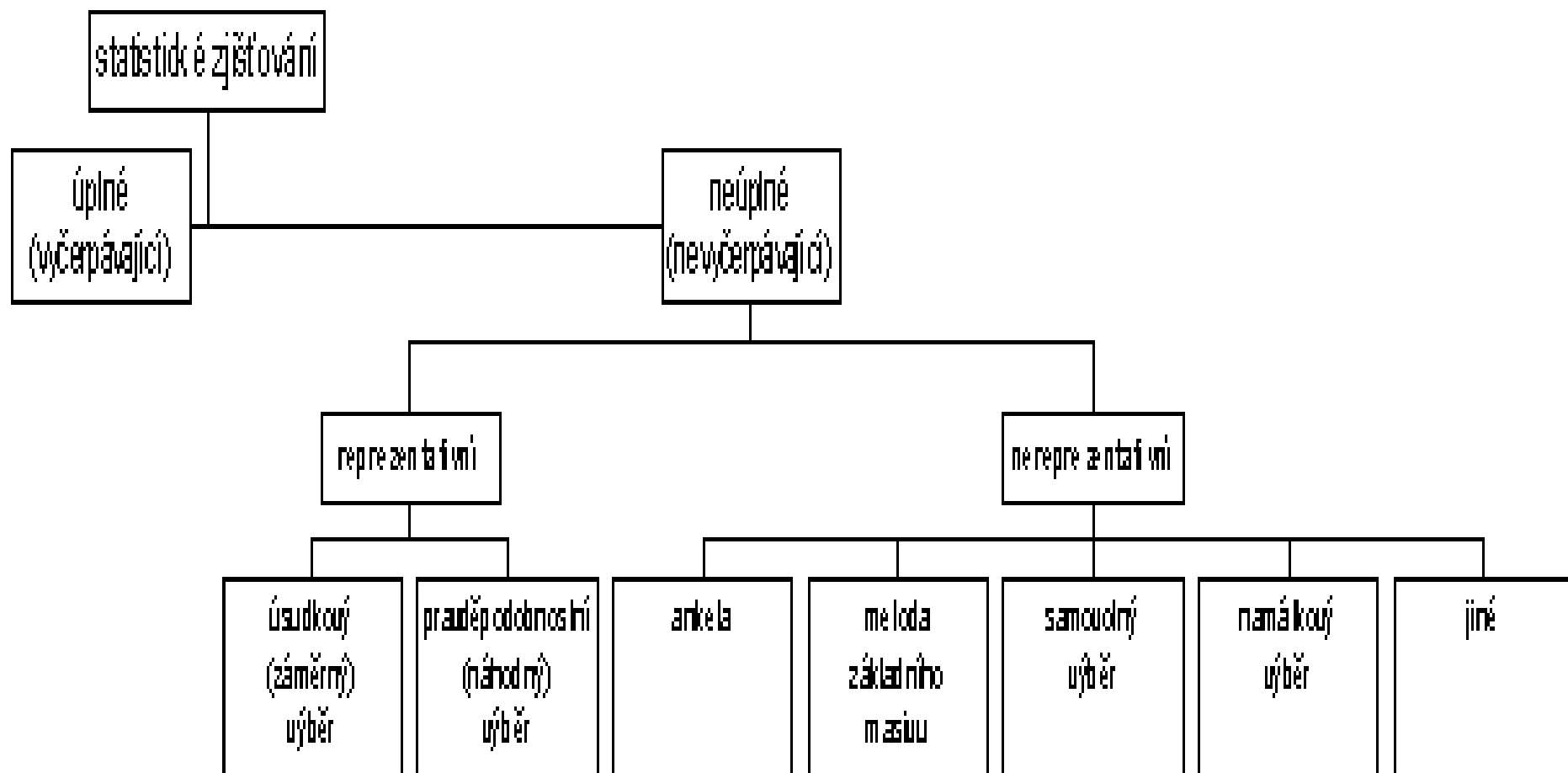
# Statistika a její pojetí

- Číselné údaje o hromadných jevech
- Praktická činnost spočívající ve sběru, zpracování, vyhodnocování jednotlivých údajů
- Teoretická disciplína zabývající se metodami sloužícími k popisu a odhadování zákonitostí, při působení podstatných a relativně stálých činitelů na hromadné jevy – masové měřítko
- Nutná podmínka – hromadný jev

# Statistické jednotky

- Elementární jednotky statistického pozorování
- Osoby - voliči, organizace, události, apod.
- Statistické znaky – vlastnosti jednotek
- Volič – věk, příjem, vzdělání, vyznání, etnicita
- Kvantitativní znaky (lze je vyjádřit číselně)
- Kvalitativní znaky (druh vlastnictví bytu, pocit spokojenosti apod.)
- Alternativní znak – nabývá jen dvou hodnot
- Množný znak – více hodnot, kvalitativní znaky

# Data a jejich získání



# Úplná vyčerpávající šetření

- Zde nepotřebujeme statistickou indukci
- Neprovádíme výběr
- Sčítání lidu, domů a bytů
- Problémy – jsou opravdu všichni v souboru?
- Volební výsledky – náš častý případ

# Výběry – neúplná šetření

- Nenáhodné (kvótní, záměrné), jen pro známé populace
- Náhodné (prostý, vícestupňový, oblastní)
- Náhodné - zde se aplikuje statistická indukce
- Pozor na záměnu, u první skupiny indukce není
- Pozor samovýběr není náhodný – anketa, žádné zobecňování
- Problém návratnosti (85% je nereálných)

# Vzorek – populace

- Kolik vran musíme pozorovat?
- Populace – základní soubor, soubor jednotek o kterém předpokládáme, že jsou pro něj naše závěry platné
- Vzorek – **výběrový soubor**, skupina jednotek, které reálně pozorujeme
- Nutná imitace složení populace tak přesně, jak je to jen maximálně možné
- S rostoucí velikostí vzorku se rozdíly snižují



# Úsudkový kvótní výběr

- Název - kvóty, údaje pro tazatele
- Charakteristiky: pohlaví, věk, vzdělání, rodinný stav, bydliště, skupina povolání, sociální skupina; počet sourozenců, počet dětí, vlastnictví např. osobního auta, národnost a jiné.
- **Pro kvóty vyhledáme v pramenech kvantitativní oporu pro rozhodnutí, jak velké mají být podíly jednotlivých kategorií dotázaných podle znaků – statistická ročenka apod.**
- kvóty pro výběr nezávisle na sobě nebo ve vzájemných vazbách - výhody a nevýhody.
- Stanovíme výběrové kvóty pro každého tazatele tak, aby pokrývaly potřebný počet a strukturu dotázaných.
- Hůře dostupné nadhodnotíme. Nepodaří-li se všem tazatelům získat potřebný počet např. žen v domácnosti nebo vysokoškoláků apod., vyšší kvóta tuto nepřesnost sníží.

# Náhodný – pravděpodobnostní výběr

- je takový výběr, ve kterém každý element populace má stejnou pravděpodobnost, že bude vybrán
- Reprezentuje všechny známé i neznámé vlastnosti populace
- Proměnné, které jsou pro nás relevantní, budou mít ve vzorku podobnou distribuci jako v celé populaci a naše závěry jsou tedy na populaci aplikovatelné

# Stratifikovaný náhodný výběr

- Základní soubor rozdělíme do podsouborů, nazýváme je „strata“
- Zde se provede prostý náhodný výběr
- Věkové rozdělení a jejich postoje k „ANO“
- Proporcionální – výběr je úměrný populaci
- Neproporcionální – pracuje se s ním v případě odlišného rozptylu u jedné skupiny – straty

# Vícestupňový náhodný výběr

- Základní soubor rozložíme do skupinek
- Jednotky jsou zastupitelné (rozdíl od strat)
- Vybereme jen některé skupinky
- Následně provádíme celostní šetření na skupince – reprezentant populace

# Metoda náhodné procházky

- Metoda náhodné procházky
- Tazatel vyrazí na cestu, první odbočka vlevo, druhá vpravo, třetí vchod na levé straně, první podlaží, první osoba (narozeniny apod.)
- Metoda sněhové koule
- První respondent, doporučí druhého, třetího atd. pořád dokola než se začnou opakovat, pozor na zkreslení (doporučím osobu blízkou)

# Statistická indukce

- Základní práce s náhodným výběrem z velkého základního celku
- Výběr má dostatečný počet jednotek
- Výběr sestaven náhodně
- Musí jít o výběr a rozhoduje náhoda!
- Desítky jednotek 30 – 50 (300 – 500 – 1000)
- Základní soubor minimálně 100krát větší než zamýšlený výběr - ČR (opravdu 70 000?)

# Malé výběry

- Výběr je složen minimálně z cca 30 – 50 případů
- Za málo peněz „hodně muziky“
- Speciální testové statistiky pro malé výběry a neparametrické metody

# Výběr z malých populací

- Populace cca stovky případů
- Ideální úplné zjišťování
- Důležitý je výběr náhodný (ne anketa)
- Problém reprezentativnosti (150N z 300)
- Problém s vrácením? Nedoporučuji



# Nereprezentativní výběry

- Anketa
- Metoda základního masivu – největší jednotka
- Samovolný výběr – provádí odborník (vězeň)
- Namátkový výběr
- Jiné, další

# Nominální proměnné

- Známé také jako kvalitativní proměnné
- Kategorie – jména bez určení více, méně
- příklady – pohlaví, barva vlasů, místo narození
- Omezené možnosti, pouze Nominální operace
- Modus – kategorie s nejvyšší četností, nejčastější barva očí v učebně
- Modální kategorie – charakteristika populace

# Pořadové proměnné

- Ordinální – můžeme hodnoty seřazovat do určité hierarchie
- Lze sledovat u jednotek vlastnosti, které jsou vyšší nižší, silnější apod.
- Bohužel ještě nevíme o kolik (vzdělání)
- Příklad – medaile (Zlatá – stříbrná – bronzová)
- Medián – hodnota ležící uprostřed všeho pozorování seřazených podle velikosti
- TV na ZŠ a nástup

# Intervalové proměnné

- Víme, zda je znak vyšší, nižší a také o kolik!
- Věk, příjem, počet dětí, apod.
- Široká škála možných operací a technik
- (korelace, regrese, apod.)
- Omezená skupina tohoto typu proměnných
- Zjišťování aritmetického průměru – intervalový popis střední hodnoty
- Součet dělený počtem sledovaných jednotek

# Kardinální (poměrové) proměnné

- Mimo výše uvedené lze u této proměnné zjistit kolikrát je jedna hodnota vyšší než druhá
- Nabývá pouze kladných hodnot
- Např. počet členů domácnosti apod.

# Rozeznání

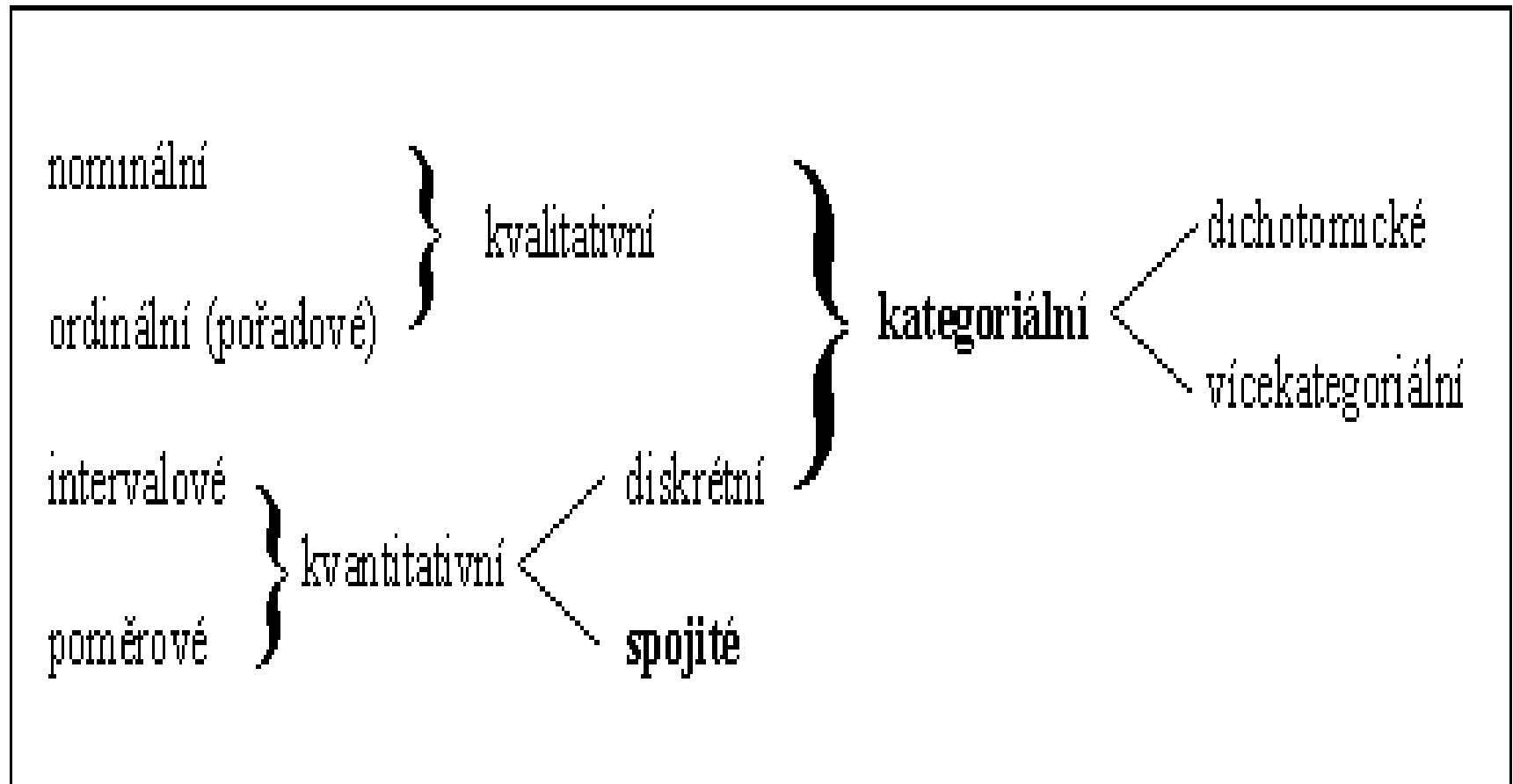
## Kritické otázky:

- (A) Je určitá kategorie proměnné větší (menší) než jiná kategorie?
  - (B) Kolikrát je větší (menší)?
- 

Jsou tyto otázky smysluplné?

A	B	
ne	ne	nominální proměnná
ano	ne	pořadová proměnná
ano	ano	intervalová proměnná

# Druhy proměnných



# Charakteristiky variability

## Variační rozpětí, rozptyl

- Variační rozpětí zobrazuje rozsah souboru
- Je to rozdíl mezi nejvyšší a nejnižší hodnotou
- Střed rozpětí – součet/2
- A) 12,15,40,15,18    B) 5,15,25,50,10
- Rozptyl: Jak se pozorování liší od průměru

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - E(x))^2$$



# Směrodatná odchylka

- Normální rozložení – počet a velikost lístků na stromě
- 68% pozorování
- 95% pozorování
- Měří homogenitu souboru
- Výběrová chyba
- Statistická významnost

# Nástrahy

- Extrémní hodnoty, zkreslení (průměrná mzda)
- Nebránit se užívání nižších technik pro vyšší „kastu“ znaků
- Nezapomenout přitom na ostatní postupy
- Ideál shodný modus, medián i arit. průměr
- Příklad – počet piv ve skupině

# Statistická významnost

- Blahuš 2000 (dle Soukup, Rabušic 2007)
- Výsledky jsou statisticky významné na hladině 0,05 znamená, že z náhodného reprezentativního výběru je riziko zobecnění na celý základní soubor nejvýše 5%.