Statistický soubor

Statistický soubor je množina objektů (dat), které jsou předmětem statistického šetření. Máme-li k dispozici kompletní množinu dat (v deskriptivní statistice), mluvíme o **základním souboru.** Pokud tento soubor není dostupný a máme k dispozici pouze výběr z něj, nebo je-li předmětem zkoumání náhodná veličina, hovoříme o tzv. **výběrovém souboru.** Počet prvků souboru se nazývá **rozsah souboru.**

Základní soubor

- **Základní soubor** je zadán přesným vymezením jeho prvků. Tyto prvky jsou určeny buď výčtem, nebo stanovením jednoznačného pravidla (např. určité společné vlastnosti), které je pak kritériem příslušnosti daného prvku do daného základního souboru.
- Prvky základního souboru mohou být různé objekty – osoby, rodiny, pokusná zvířata, biologický materiál, záznam EEG apod.
- Např. základním souborem může být soubor obyvatel určitého území v daném časovém období, soubor dětí s určitou vrozenou vadou, soubor vzorků tkáně pokusných zvířat, ...
- **Rozsah** základního souboru může být *konečný* (např. demografické soubory), nebo *nekonečný,* což je spíše ideální soubor, existující pouze hypoteticky (např. soubor všech možných výsledků pokusů proveditelných za daných experimentálních podmínek nebo soubor všech osob s danou nemocí).
- **Homogenním souborem** rozumíme **soubor, který je stejnorodý,** tj. ve kterém všichni členové mají stejné vlastnosti, odpovídající předem zvolenému kritériu.[zdroj?]

Výběrový soubor

Metodami statistické inference lze vyvozovat závěry o základním souboru ze zjištěných vlastností u části jeho prvků tzv. **výběru**. Aby to bylo možné, je nutné zvolit výběr, který je **reprezentativní** tj. výběr, který odráží svým složením vlastnosti všech prvků základního souboru. Výběr, který není reprezentativní, je označován jako **selektivní.** Např. při zjišťování průměrné výšky chlapců v populaci ve věku 10 let může být výběr složený z chlapců, kteří hrají basketbal, značně selektivní.

Abychom vyčíslili stupeň nejistoty, že výběr není reprezentativní, a tak podpořili vědeckost závěrů statistické inference je nutné vytvářet výběr technikou **náhodného neboli pravděpodobnostního výběru**. Tyto metody mají zaručit, že každý prvek základního souboru má stejnou možnost, že bude zařazen do výběru a každý další prvek je vybírán nezávisle na těch, které jsme již vybrali. Podle způsobu provedení rozlišujeme různé techniky náhodného výběru např.:

- **Prostý náhodný výběr**
  - je prováděn vhodnou technikou losování. Existuje řada technik losování, včetně použití tabulek náhodných čísel. Nevýhodou tohoto postupu je však nezbytnost předchozí identifikace prvků (např. očíslování), což není v praxi u rozsáhlejších souborů proveditelné.
- **Mechanický (systematický) výběr**
  - je podmíněn určitým, předem daným uspořádáním prvků základního souboru. Do výběru zařadíme všechny prvky, které jsou od sebe vzdáleny o určitý výběrový krok k, přičemž první prvek vybereme náhodně z k prvků na začátku souboru.
- **Oblastní (stratifikovaný) výběr**
  - se provádí tehdy, když lze soubor rozdělit do takových oblastí, které jsou uvnitř homogenní (ve sledovaných znacích se příliš neliší) a mezi sebou heterogenní (mezi sebou se mohou lišit).
- **Skupinový výběr**
  - se provádí v případech, kdy základní soubor je velmi početný. Zde nevybíráme jednotlivé prvky, ale skupiny prvků, které tvoří přirozená či umělá seskupení (např. rodina). Je žádoucí, aby byly skupiny pokud možno stejně velké a prvky uvnitř skupin různorodé.
- **Vícestupňový výběr**
  - je založen na existenci určitého hierarchického uspořádání prvků základního souboru. K těmto prvkům se postupně dostáváme přes vyšší výběrové jednotky (např. města – domy – domácnosti).

Odkazy

Související články

- Popisná statistika
- Statistická inference
- Analytická studie

Použitá literatura

- BENCKO, Vladimír, et al. *Epidemiologie : výukové texty pro studenty 1. LF UK.* 2. vydání. Praha : Karolinum, 2002. s. 70-73. ISBN 80-246-0383-7.

- ŠPUNDA, Miroslav a Jaroslav DUŠEK, et al. *Zdravotnická informatika.* 1. vydání. Praha : Karolinum, 2007. 194 s. ISBN 978-80-246-1378-9.