**1. Základní statistické pojmy**

Úkolem této kapitoly je zopakovat základní pojmy z oblasti aplikované statistiky, na které se budou následující kapitoly odkazovat.

**Proměnná –** znak objektu (v případě sociologické analýzy dat znak respondenta). Proměnná zastupuje jeden konkrétní znak všech zkoumaných respondentů. Např. proměnná **AGE** označuje charakteristiku věk, a může pro jednotlivé respondenty a respondentky nabývat různých hodnot. Někdo má hodnotu AGE rovnu 18, někdo jiný mí hodnotu AGE rovnu 33 apod. Obvykle rozlišujeme následující typy proměnných:

* dichotomické/binární: nabývají jen edvou hodnot, např. bydliště Praha/mimo Prahu
* nominální: lze vyjmenovat jejich hodnoty, ale nikoliv seřadit (muž, žena, výčet povolání…)
* ordinální: lze vyjmenovat jejich hodnoty a jednoznačně stanovit jejich pořadí, např. dny v týdnu nebo měsíce v roce
* kardinální/spojité: mají mnoho hodnot, které jsou seřazení, např. věk nebo výše příjmu. Dělí se na intervalové a poměrové.

**Datová matice** – shrnuje informace o všech respondentech v datovém souboru. Je uspořádána do tabulky, kde v řádcích jsou jednotlivé zkoumané případy (v sociologické analýze nejčastěji osoby – respondenti) a ve sloupcích jednotlivé proměnné, které se těchto respondentů týkají.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Věk | Vzdělání | Politická orientace | Náboženství | … |
| 1 | 22 | VŠ | L | Římkat |
| 2 | 19 | SŠ | L | . |
| 3 | 25 | VOŠ | P | . |
| … | | | | | |

**Frekvenční tabulka** – vyjádření četností jednotlivých případů konkrétní proměnné. Z důvodu lepší srovnatelnosti je vhodnější nepoužívat absolutní hodnoty četností, ale procentní vyjádření.

|  |  |
| --- | --- |
| Vzdělání | Procenta |
| ZŠ | 7 % |
| Vyučen | 21 % |
| SŚ s maturitou | 34 % |
| … | |

**Kontingenční tabulka** – frekvenční tabulka zobrazující souvislost dvou (a více) proměnných.

|  |  |  |
| --- | --- | --- |
| Vzdělání / Pohlaví | Muž | Žena |
| ZŠ | 7 % | 5 % |
| Vyučen | 30 % | 20 % |
| SŠ s maturitou | 40 % | 15 % |
| … | | |
| Total | 100 % | 100 % |

**Korelace –** ukazuje souvislost dvou proměnných. Existuje několik korelačních koeficientů pro různé typy proměnných. Obecně platí, že čím vyšší hodnoty koeficient nabývá, tím silnější je souvislost korelovaných proměnných. V sociálních vědách se obvykle považuje hodnota korelačního koeficientu 0,3 (a vyšší) za dostatečně silnou souvislost. Pozor, prokázání korelace neukazuje na kauzalitu. Ukazuje jen, že obě proměnné se mění společně, nikoliv že změna jedné způsobí změnu druhé.

Regrese – modeluje vztah mezi dvěma (případně více) proměnnými pomocí ideální křivky. V případě lineární regrese se jedná o přímku. Základní rovnice lineární regrese má tvar

y = a + bx + e

kde y je vysvětlovaná / závisle proměnná (dependent variable), x je vysvětlující / nezávisle proměnná (independend variable) a je průsečík s osou y (intercept, konstanta), b je koeficient proměnné b (sklon přímky, slope), e je chyba měření.