

Typy dat

Statistická data, tedy informace o jednom prvku statistického souboru, lze rozdělit do několika skupin podle jejich charakteru. Toto rozdělení má velký význam, podle typu dat lze zvolit např. vhodnou popisnou statistiku (např. míru polohy) nebo vhodný statistický test při testování hypotéz.

Data nominální

Nominální data jsou taková data, která mají pouze význam jisté kvality, proto se někdy hovoří o kvalitativních datech. Obvykle je na výběr pouze z konečné množiny možností. Jednotlivé hodnoty jsou neporovnatelné, to znamená, že data nelze seřadit, neexistuje nic jako „velikost“. Typickým příkladem nominálních dat jsou např. krevní skupina, rodinný stav nebo třeba sekvence DNA, tedy nějaké kategorie (proto se hovoří o datech kategoriálních). Zvláštní skupinou jsou data alternativní (dichotomická), ta mají pouze hodnoty ANO a NE.

V některých případech se používá kódování jednotlivých hodnot čísly, ovšem na tato čísla je nutné nahlížet pouze jako na symboly. Důvodem je to, že i když je vhodné při kódování použít nějaký systém, je v konečném důsledku lhostejné, jaké hodnotě je jaký číselný kód přiřazen.

Nominální data nelze srovnat a proto lze hodnotit pouze jejich četnost ve statistickém souboru. Použitelnou mírou polohy je modus.

Data ordinální

Ordinální data představují podobně jako data nominální výběr z nějakého počtu možností. Významným rozdílem oproti datům nominálním je to, že lze přirozeným způsobem zavést uspořádání a u každé dvojice hodnot lze snadno určit, která hodnota je větší a která menší. Právě díky této kvantifikaci se od ordinálních dat dále již hovoří o datech kvalitativních.

Typickým příkladem ordinálních dat může být například nejvyšší dosažené vzdělání. Je zřejmé, že ordinální data neumožňují posoudit „vzdálenost“ jednotlivých kategorií nebo hodnot.

Data intervalová

Intervalová data jsou taková data, ve kterých má smysl hodnotit i vzdálenosti mezi jednotlivými kategoriemi nebo hodnotami. Příkladem takových dat může být například teplota měření v Celsiově teplotní škále. Rozdíl teploty 10 °C je vždy stejný bez ohledu na konkrétní výchozí teplotu. Intervalová data však nezahrnují poměr hodnot; např. tvrzení, že nárůstem teploty z 10 °C na 30 °C se teplota zvýšila 3×, je z fyzikálního hlediska zcela chybné.

Data poměrová

Poměrová (podílová) data představují taková data, u kterých jsou již definovány i poměry jednotlivých hodnot. Patří sem vlastně všechny fyzikální veličiny definované v souladu s SI. Jejich další charakteristickou vlastností je to, že mají jasně definovanou absolutní nulu (tedy např. termodynamická absolutní nula, nulová vzdálenost, nulová hmotnost,...).

Význam rozlišování ordinálních, intervalových a poměrových dat tkví především v tom, že při manipulaci s takovými daty je třeba mít stále na mysli omezení plynoucí z jejich charakteru.

Data poměrová a intervalová jsou obvykle spojitá, jejich hodnoty se mohou plynule měnit v určitém intervalu. Naopak data ordinální a tím spíše data nominální jsou obvykle diskrétní, nabývají jen určitého konečného počtu možných hodnot.

V některých případech je třeba spojitá data převést na diskrétní. Obvyklým způsobem je prosté rozdělení možných hodnot na intervaly, přičemž každý interval představuje jednu kategorii. Typickým příkladem je rozdělení výšky na intervaly po několika centimetrech.

Odkazy

Související články

- Testování statistických hypotéz
- Míry polohy
- Míry variability

Použitá literatura

- CYHELSKÝ, Lubomír a Vladimíra VALENTOVÁ. Význam základní klasifikace ukazatelů pro korektní interpretaci vzájemných odlišností jejich hodnot. *Politická ekonomie* [online]. 2006, vol. 4, s. 542-548, dostupné také z

<<https://www.vse.cz/polek/download.php?jnl=polek&pdf=573.pdf>>. ISSN 0032-3233.

- ZVÁROVÁ, Jana. *Biomedicínská statistika I : Základy statistiky pro biomedicínské obory*. 1. vydání. 2002. ISBN 80-7184-786-0.

Externí odkazy

- ROSSITER, D.G.. *An introduction to statistical analysis* [online]. [cit. 19.8.2011]. <<https://www.itc.nl/~rossiter/teach/stats/sintro.pdf>>.