**interpretace dat: základní pojmy**



V tabulce vpravo jsou spočítané následující hodnoty:

1. *průměrné* hodnoty pro formanty F1, F2, F3, F4 zvlášť pro hlásku /r/ a zvlášť pro hlásku /l/;

funkce v Excelu:

|  |  |
| --- | --- |
| cz | en |
| =PRŮMĚR(……) | =MEAN(……) |

1. *mediány* (střední hodnoty – když bychom všechny hodnoty uspořádali postupně podle velikosti do jedné řady, medián je hodnota, kterou najdeme uprostřed této řady) pro formanty F1, F2, F3, F4 zvlášť pro hlásky /r/ a /l/; funkce v Excelu:

|  |  |
| --- | --- |
| cz | en |
| =MEDIAN(……) | =MEDIAN(……) |

1. *směrodatné odchylky* (o kolik se průměrně liší naměřené hodnoty od mediánu) pro formanty F1, F2, F3, F4 zvlášť pro hlásky /r/ a /l/, funkce v Excelu:

|  |  |
| --- | --- |
| cz | en |
| =SMODCH(……) | =SMODCH(……) |

1. *přesnost* – pokud jsou odchylky měření malé, je naše měření vyrovnané (není rozkolísané) – to je pro nás dobře – číslo v řádku „přesnost %“ udává, kolik procent z mediánové hodnoty tvoří odchylka

– čím je přesnost větší, tím se jednotlivá měření odchylují od mediánu méně, tzn. všechny hodnoty se pohybují velmi blízko mediánové hodnoty --- naše měřené hodnoty nejsou rozkolísané;

– čím je přesnost menší, tím se jednotlivá měření odchylují od mediánu více, tzn. alespoň některé hodnoty se pohybují dost daleko od mediánové hodnoty --- naše měřené hodnoty jsou rozkolísané; funkce v Excelu:

|  |
| --- |
| vlastnífunkce *=* kolik procent z mediánové hodnoty tvoří odchylka |
| = 100-(hodnotasměrodatná odchylka / hodnotamedián )\*100 |

1. na základě hodnot přesnosti z tabulky výše můžeme konstatovat, že je naše měření velmi málo rozkolísané (v této chvíli máme měření od jednoho mluvčího, u jednoho mluvčího většinou najdeme konzistentní hodnoty) – ve většině případů dosahuje hodnoty přes 95 %
2. u hodnot F1 a F3 u hlásky /r/ vidíme znatelnější rozkolísání – přesnost dosahuje 93 %

**interpretace dat II: z-score**





*z-score*:

- udává, kolikanásobek směrodatné odchylky musíme přičíst/odečíst k průměru, abychom dostali naměřenou hodnotu (pro F1 vrkat: (37 \* (+1,36)) + 530 = 580

 pro F1 hrkat: (37 \* (-1,05)) + 530 = 492)

- to znamená, že z-score nám udává (vzhledem ke směrodatné odchylce), jak moc se hodnoty přibližují (popř. vzdalují) naší průměrné hodnotě

- z-score je způsob, jak ověřit rozkolísanost datového souboru, na základě posuzování jednotlivých naměřených hodnot

PŘÍKLADY:

- pokud je z-score 0: naměřená hodnota se rovná [směrodatná odchylka\*0 + průměr] = průměr

- pokud je z-score 1: naměřená hodnota se rovná [směrodatná odchylka\*(+1) + průměr]

- pokud je z-score -1: naměřená hodnota se rovná [směrodatná odchylka\*(-1) + průměr]

INTERPRETACE HODNOT z-score:

* z-score v intervalu (-1;1) – naměřená hodnota je velmi blízko průměru
* z-score v intervalu (-2;2) – naměřená hodnota ještě ujde, není až tak vzdálená průměru
* z-score v intervalu (-3;3) – naměřená hodnota už je dost daleko od průměru, zvažujeme odstranění outlierů (velmi vychýlených hodnot)
* z-score mimo výše uvedené intervaly – hodnoty vystřelily příliš daleko od průměru, určitě by se měly považovat za outliery a z datového souboru by měly být odstraněny

!!! z-score počítáme, když máme dostatečně velký objem dat, pro malé objemy dat se nehodí !!!

to platí i pro všechny dříve uvedené hodnoty: tzn. průměrování, mediánování, měření standardní odchylky
i z-score se dělá, když mám dost dat! aplikováním statistiky na malé datové soubory zkreslujeme data!

V tabulce výše jsou červeně označeny ty z-score hodnoty, které jsou mimo interval (-2;2). To jsou hodnoty, které se oproti průměru vychylují výrazně a zvažovali bychom jejich odebrání. V našem datovém souboru se takto vychyluje celkem šest hodnot.