Obecné momenty – prosté tvary

První obecný moment: $(\Sigma y_i)/n$, $i=1\ldots n$

aritmetický průměr, „těžiště“ dat

$\bar{y}$

Druhý obecný moment: $(\Sigma y_i^2)/n$, $i=1\ldots n$

$\overline{y^2}$

Obecné momenty – prosté tvary

Příklad 1 – pokračování:

$$\bar{y} = (3+4+2+3+2+3+3+3)/8 = 23/8 = 2{,}875$$

$$\bar{y^2} = (3^2+4^2+2^2+3^2+2^2+3^2+3^2+3^2)/8 =$$

$$= (9+16+4+9+4+9+9+9)/8 = 69/8 = 8{,}625$$

Obecné momenty – prosté tvary

Příklad 1 – Poznámka:

Hodnota SOUČTU všech pozorování
(3+4+2+3+2+3+3+3) = 23
je tzv. **ÚHRN** (úhrnná hodnota).

Platí: průměr = úhrn / počet
aneb: úhrn = průměr · počet
aneb: počet = úhrn / průměr

Obecné momenty – vážené tvary

Při výpočtu lze využít četností absolutních:

$$\bar{y} = (\Sigma y_i \, n_i)/n \,, \qquad i=1\ldots K$$

$$\overline{y^2} = (\Sigma y_i^2 \, n_i)/n, \qquad i=1\ldots K$$

Obecné momenty – vážené tvary

Příklad 1 - pokračování:

$$\bar{y} = (2\cdot 2+3\cdot 5+4\cdot 1)/8 = 23/8 = 2{,}875$$

$$\bar{y^2} = (2^2\cdot 2+3^2\cdot 5+4^2\cdot 1)/8 = (4\cdot 2+9\cdot 5+16\cdot 1)/8 =$$
$$= 69/8 = 8{,}625$$

Obecné momenty – vážené tvary

Při výpočtu lze využít také četností relativních:

$$\bar{y} = \Sigma y_i \, p_i \,, \qquad i=1\ldots K$$

$$\overline{y^2} = \Sigma y_i^2 \, p_i, \qquad i=1\ldots K$$

Obecné momenty – vážené tvary

Příklad 1 - pokračování:

$$\bar{y} = 2 \cdot 0{,}250 + 3 \cdot 0{,}625 + 4 \cdot 0{,}125 = 2{,}875$$

$$\bar{y^2} = 2^2 \cdot 0{,}250 + 3^2 \cdot 0{,}625 + 4^2 \cdot 0{,}125 =$$

$$= 4 \cdot 0{,}250 + 9 \cdot 0{,}625 + 16 \cdot 0{,}125 = 8{,}625$$

Centrované momenty prosté tvary

Druhý centrovaný moment:

$$M_2(y) = \Sigma(y_i - \bar{y})^2 / n, \qquad i=1\ldots n$$

= <u>ROZPTYL</u> aneb „průměrná čtvercová odchylka od aritmetického průměru“

$$\sqrt{M_2(y)}$$

= <u>směrodatná odchylka</u>

Centrované momenty prosté tvary

Příklad 1 - pokračování:

$$M_2(y)=[(3-2{,}875)^2+(4-2{,}875)^2+(2-2{,}875)^2+$$
$$+(3-2{,}875)^2+(2-2{,}875)^2+(3-2{,}875)^2+$$
$$+(3-2{,}875)^2+(3-2{,}875)^2]/8=0{,}359$$

$$\sqrt{M_2(y)}=\sqrt{0{,}359}=0{,}599$$

Známky byly zhruba v rozmezí 2,875±0,599
aneb v rozmezí 2,276 až 3,474.

Centrované momenty
vážené tvary

K výpočtu lze užít četností absolutních:

$$M_2(y) = \Sigma(y_i - \bar{y})^2 \, n_i / n, \qquad i=1\ldots K$$

nebo relativních:

$$M_2(y) = \Sigma(y_i - \bar{y})^2 \, p_i , \qquad i=1\ldots K$$

Centrované momenty vážené tvary

Příklad 1 - pokračování:

$$M_2(y) = [(2-2{,}875)^2 \cdot 2 + (3-2{,}875)^2 \cdot 5 + (4-2{,}875)^2 \cdot 1] / 8 = 0{,}359$$

$$M_2(y) = (2-2{,}875)^2 \cdot 0{,}250 + (3-2{,}875)^2 \cdot 0{,}625 + (4-2{,}875)^2 \cdot 0{,}125 = 0{,}359$$

MOMENTY

Mají smysl jen u číselných veličin

(diskrétních či spojitých)

Vážené tvary má smysl používat jen tehdy, jsou-li k dispozici četnosti (tj. u diskrétních veličin)

Použití prostého i váženého tvaru musí dát stejný výsledek (jde o totéž, jen jinak počítáno)

MOMENTY

Chování při aditivní lineární transformaci:

Známe průměr a rozptyl.

Každé pozorování změníme o stejnou konstantu c.

Průměr se tak změní o tutéž konstantu c, rozptyl se nezmění.

MOMENTY

Příklad 1 - pokračování:

Pokud bychom žákovi všechny udělené známky o jeden stupeň zlepšili (c= -1), průměrná známka se změní z hodnoty 2,875 na hodnotu 1,875, ale rozptyl zůstane 0,359 (ověřte výpočtem).

MOMENTY

Chování při multiplikativní lineární transformaci:

Známe průměr a rozptyl.

Každé pozorování vynásobíme stejnou konstantou c.

Průměr se také musí vynásobit konstantou c, rozptyl se musí vynásobit její druhou mocninou c^2.

MOMENTY

Příklad 2:

Byly měřeny délky vyrobených trubek v centimetrech. Průměrná délka činí 49 cm a rozptyl 144 (cm^2). Nyní je nutno všechna měření převést na decimetry. Jak se změní hodnoty momentů?

Jelikož c=0,1, průměr bude činit 4,90 dm a rozptyl 1,44 (dm^2).

MOMENTY

ROZPTYL – rychlý výpočetní tvar

$$M_2(y) = \overline{y^2} - \overline{(y)}^2$$

Příklad 1 - pokračování:

$M_2(y) = 8{,}625 - 2{,}875^2 = 0{,}359$

Poznámka: Vždy musí vyjít $M_2(y) \geq 0$!

MOMENTY

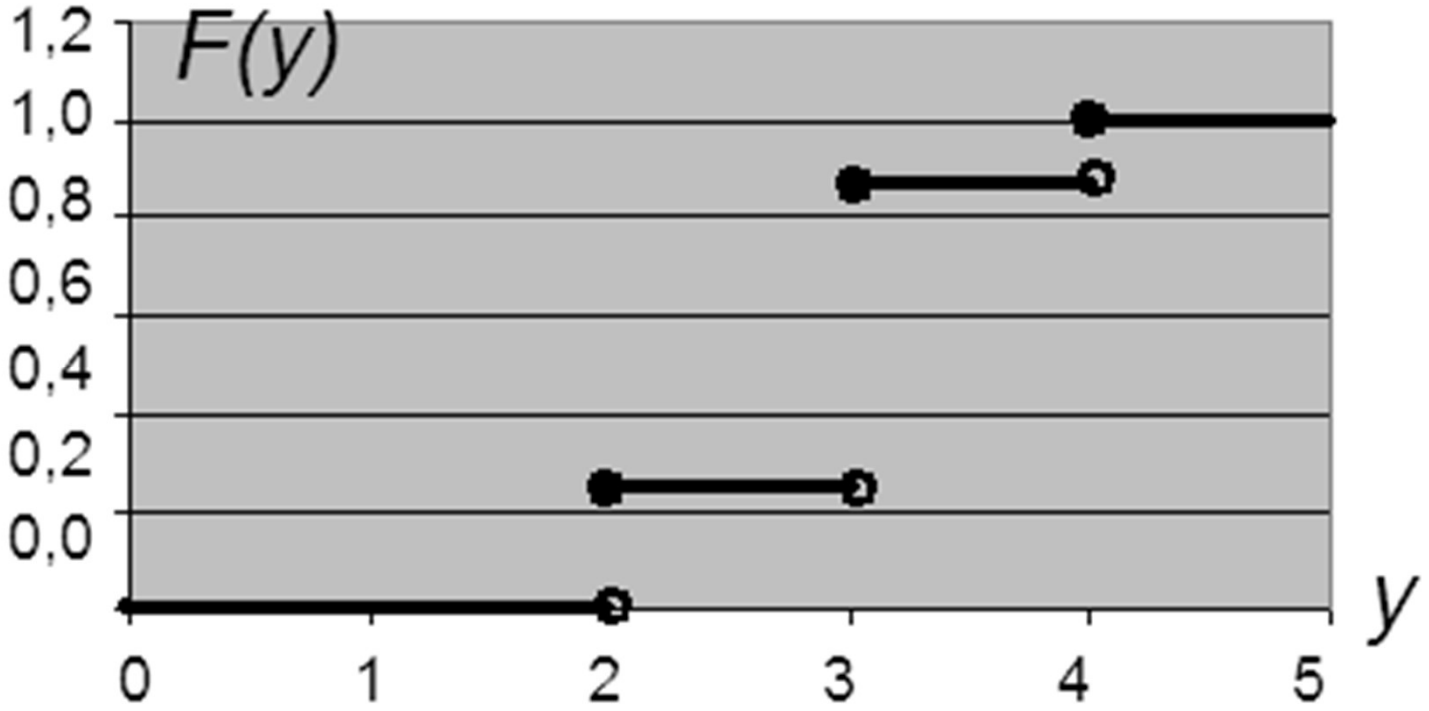
- Rozptyl – čtvercová míra variability (proměnlivosti) dat
- Relativní mírou variability je variační koeficient

$$V_y = \sqrt{M_2(y)} / \bar{y}$$

může porovnávat variabilitu souborů, v nichž je veličina zaznamenána v různých měrných jednotkách – např. platy u nás v Kč versus platy v Německu v Euro), či je na jiné úrovni (poloze)

DISTRIBUČNÍ FUNKCE

- Slouží k popisu rozdělení (distribuce) číselných dat; jakési „zobecnění“ kumul.relat.četností:
- *F*(*y*) … udává podíl pozorování s hodnotou nejvýše *y*, viz graf pro data z příkladu 1:



KVANTILY

- Udávají pro číselnou veličinu *Y* hodnotu *y*, pod níž leží požadovaný podíl pozorování
- Značíme $\tilde{y}_{100p}$, kde *p* je onen podíl (údaj mezi 0 – 1)
- *p*=0,5 … 50% kvantil – medián

(odděluje polovinu nižších od zbytku vyšších pozorování; značen obvykle jen $\tilde{y}$)

- *p*=0,25 … 25% kvantil – dolní kvartil $\tilde{y}_{25}$
- *p*=0,75 … 75% kvantil – horní kvartil $\tilde{y}_{75}$

KVANTILY

1. Určení pomocí distribuční funkce

Najdeme bod *y,* v němž poprvé *F(y)* dosáhne úroveň *p.*

Příklad 1 – pokračování:

Nalezněte medián známek žáka.

Medián…tj. p=0,5.

Kde poprvé příslušná *F(y)* dosáhla úroveň 0,5?

Podle grafu je mediánem hodnota $\tilde{y}$ =3.

Interpretace: Polovina známek měla hodnotu nejvýše 3 (tj. 3 nebo nižší)

KVANTILY

2. Určení přímo pomocí dat

a) Data uspořádáme vzestupně dle velikosti

b) Nalezneme celočíselné z vyhovující nerovnicím:

$$n \cdot p < z < n \cdot p + 1$$

c) Hledaným kvantilem je hodnota s pořadovým číslem z.

Příklad 1 – pokračování:

Nalezněte 60% kvantil známek žáka.

Známky seřadíme: 2, 2, 3, 3, 3, 3, 3, 4.

Jelikož n=8 a p=0,6, vyjde $n \cdot p$=4,8, takže hledáme z :

$$4{,}8 < z < 5{,}8$$

Tudíž z=5, aneb hledaným kvantilem je 5. známka, čili trojka: $\tilde{y}_{60}$=3.

KVANTILY

Příklad 1 – pokračování:

Interpretujte nalezený výsledek: $\tilde{y}_{60}=3$.

Znamená to, že 60 % zjištěných známek mělo hodnotu nejvýše 3 (tj. 3 nebo nižší).

Podobně interpretujte to, kdyby pro veličinu plat měl např. horní kvartil hodnotu 32 tis. Kč:

Znamenalo by to, že 75 % zjištěných platů mělo hodnotu nejvýše 32 tis. Kč (*tj. zbylých 25 % platů bylo jakých*?)

KVANTILY

2. Určení přímo pomocí dat – možný problém

a) Data uspořádáme vzestupně dle velikosti

b) Nenalezneme celočíselné z vyhovující nerovnicím:

$$n \cdot p < z < n \cdot p + 1$$

Stane se to tehdy, když obě hodnoty ($n \cdot p$; $n \cdot p+1$) vyjdou přesně celočíselně.

c) Nalezneme hodnotu s pořadovým číslem $n \cdot p$ a s následujícím pořadovým číslem $n \cdot p+1$.

d) Za hledaný kvantil prohlásíme průměr obou nalezených hodnot.

KVANTILY

2. Určení přímo pomocí dat – možný problém

Příklad 1 – pokračování:

Nalezněte medián známek žáka.

Známky seřadíme: 2, 2, 3, 3, 3, 3, 3, 4.

Jelikož n=8 a p=0,5, vyjde $n \cdot p$ =4,
takže hledáme z :

$$4 < z < 5 \qquad \text{(neexistuje)}$$

Najdeme tedy jak 4., tak 5. známku.

Jelikož jsou obě trojky, jejich průměrem je také trojka a to je hledaný medián: $\tilde{y}$=3.

KVANTILY

Poznámka 1: Není chybou, že nám pro tato data vyšly 50% a 60% kvantil shodně (=3), ale na druhou stranu to není pravidlo. Pro jiná data můžou tyto kvantily vyjít různě – v takovém případě pak ale určitě bude $\tilde{y} < \tilde{y}_{60}$.

Poznámka 2: Může se stát, že postup pomocí distribuční funkce a postup přímo z dat skončí různými výsledky (v případě problému s neexistencí z, a to kdyby se hodnoty s pořadím $n \cdot p$ a $n \cdot p+1$ lišily). Rozdíl však obvykle bývá zanedbatelný.