

## Téma 3 – Příklad 1

### Zadání příkladu

Z korelační tabulky určete podmíněné průměry a rozptyly. Sestavte také datové vektory, které odpovídají oběma proměnným.

	$y_j$	
$x_i$	6	9
2	2	-
5	4	1
7	-	3
11	2	1

### Vypracování příkladu

Korelační tabulka zobrazuje dvě číselné proměnné. Podmíněnou charakteristikou rozumíme určitou hodnotu deskriptivní statistiky pro proměnnou  $y$ , která platí za předpokladu určité hodnoty proměnné  $x$ . V našem případě budeme počítat podmíněný průměr pomocí vzorce:

$$\bar{y}_i = \frac{\sum_{j=1}^l y_j n_{ij}}{n_{i.}}$$

kde  $n_{ij}$  jsou příslušné sdružené absolutní četnosti a  $n_{i.}$  je okrajová absolutní četnost, která je součtem počtu hodnot  $y$  za předpokladu určité hodnoty  $x$ .

V našem případě pro hodnotu  $x=2$  platí tento průměr:

$$\bar{y} = \frac{2 * 6}{2} = 6$$

Zbylé průměry jsou vypočteny a zobrazeny v následující souhrnné tabulce.

Podmíněný rozptyl vypočtený dle vzorce

$$s_i^2 = \frac{\sum_{j=1}^l (y_j - \bar{y}_i)^2 n_{ij}}{n_{i.}}, \text{ je pro hodnotu } x, \text{ která se rovná dvěma } s^2 = \frac{0+0}{2} = 0$$

	y <sub>j</sub>		Suma	Podmíněný průměr	Podmíněný rozptyl
x <sub>i</sub>	6	9			
2	2	-	2	6	0
5	4	1	5	6,6	1,44
7	-	3	3	9	0
11	2	1	3	7	2
<b>Suma</b>	<b>8</b>	<b>5</b>	<b>13</b>	<b>X</b>	<b>X</b>

### Řešení v SGP

Pro program Statgraphics je nejprve nutné vytvořit z korelační tabulky dva datové vektory odpovídající proměnným  $x$  a  $y$ . Pro jednu proměnnou lze využít operátoru REP a vytvořit vektor z okrajových absolutních četností, např. REP ( $x_i; n_i$ ). Druhou proměnnou ovšem musím párově přiřadit tak, aby správně odpovídaly jednotlivé hodnoty  $x$  a  $y$ . Výsledné vektory pak vypadají takto.

	x	y
1	2	6
2	2	6
3	5	6
4	5	6
5	5	6
6	5	6
7	5	9
8	7	9
9	7	9
10	7	9
11	11	6
12	11	6
13	11	9
14		

### Procedura v SGP: Describe - Numeric Data – Subset Analysis (Summary Statistics)

Při vstupním dialogu zadáme jako *Data* proměnnou, z jejíž hodnot budeme počítat výběrové charakteristiky, proměnnou  $y$ , do políčka *Codes* zadáme proměnnou podmíněnou, tedy  $x$ .

V okně Summary Statistics vidíme jednotlivé podmíněné charakteristiky. Jejich zobrazení můžeme upravit v doplňkových možnostech Pane Options. Nezapomeňme, že jde o výběrové charakteristiky, takže v případě, že nás zajímají charakteristiky základní, musíme upravit

výrazem  $(n-1/n)$ . U charakteristik tvaru rozdělení je odlišnost výraznější, jelikož program Statgraphics užívá vzorců odlišných od klasických momentových měr.

**Summary Statistics**

Data variable: y

						<i>Standard</i>	<i>Coefficient</i>	<i>Mean</i>	
x	<i>Count</i>	<i>Average</i>	<i>Median</i>	<i>Mode</i>	<i>Variance</i>	<i>Deviation</i>	<i>of variation</i>	<i>absolute deviation</i>	<i>Range</i>
2	2	6,0	6,0	6,0	0,0	0,0	0,0%	0,0	0,0
5	5	6,6	6,0	6,0	1,8	1,34164	20,3279%	0,96	3,0
7	3	9,0	9,0	9,0	0,0	0,0	0,0%	0,0	0,0
11	3	7,0	6,0	6,0	3,0	1,73205	24,7436%	1,33333	3,0
Total	13	7,15385	6,0	6,0	2,30769	1,51911	21,2349%	1,42012	3,0

## Interpratace

Za předpokladu, že proměnná x nabývá hodnoty dva, je průměrná hodnota proměnné y rovna šesti. Hodnoty y jsou v tomto případě absolutně nevariabilní. V případě, že x je rovno pěti, je průměr proměnné y roven 6,6. Průměrně se hodnoty od tohoto průměru liší o 1,2 (výběrově 1,34). Pro  $x=7$  je průměrná hodnota proměnné y devět. I zde jsou hodnoty bez jakékoliv variability. Pro x rovné dvanácti je průměrná hodnota proměnné y rovna 7, v tomto případě se hodnoty průměrně odchylují od průměru o přibližně 1,41 (výběrově 1,73).

## Řešení v MS Excel

Dle verze MS Excel je nutné využít specifických statistických funkcí pro výpočty podmíněných charakteristik.