

Příklad 1

Marketérka Penzionu v Parku ve vinařské oblasti Čejkovic u Hodonína nám poskytla interview, ve kterém uvedla, že poskytovatelé penzionů obecně upřednostňují delší pobyty před kratšími. Důvodem pro ně je větší ziskovost vlivem snížení výše denních nákladů. Na podporu svého tvrzení nám poskytla data o 10 pozorováních v jednom měsíci.

Dny	1	1	1	2	2	2	3	3	3	4	4
Tržby za pobyt	510	550	560	800	820	850	1030	1080	1270	1350	1400
Denní náklady	210	230	220	180	190	200	170	140	180	160	150

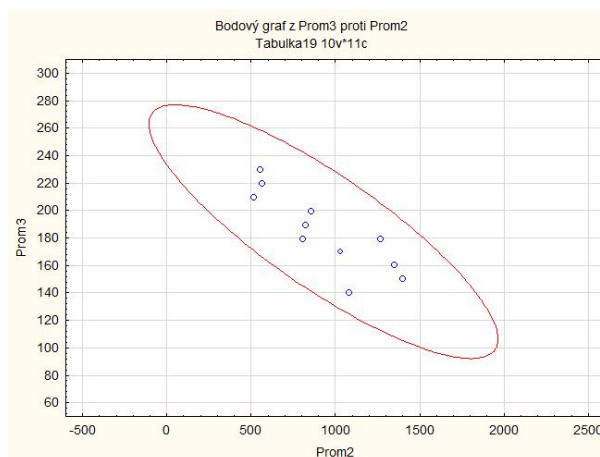
- Ověřte graficky dvourozměrnou normalitu výběru
- Vypočítejte výběrový korelační koeficient mezi Tržbami a Denními Náklady
- Otestujte statistickou významnost korelačního koeficientu
- Interpretujte výsledek, rozhodněte zda je tvrzení marketérky správné, zamyslete se zda by tato analýza šla provést jiným způsobem

Řešení

Abychom mohli využít pearsonův korelační koeficient musíme v prvním kroku ověřit dvourozměrnou normalitu výběru. Ve statistice využijeme následující postup:

Grafy → Bodové grafy → Proměnné X: Tržby; Y: naklady → OK → záložka: Detaily → Proložení vypnuto; Elipsa Normální → OK. Po vytvoření grafu upravíme měřítka na obou osách. Poklepeme na pozadí grafu - v stromové struktuře vybereme měřítka - Mód: ručně - minimum pro osu X je -600; maximum pro osu X je 2500; minimum pro osu Y je 60; maximum pro osu Y je 300 → OK.

Výsledek je zachycen na následujícím obrázku:



Jak je z obrázku patrné všechny body se nachází v elipse. Z tohoto důvodu je alespoň graficky dvourozměrná normalita potvrzena a můžeme přejít na výpočet pearsonova korelačního koeficientu.

Statistiky → Zakladní statistiky/tabulky → Korelační matice → OK → 2 seznamy → 1. seznam proměnných: Tržby; 2. seznam proměnných: Naklady → OK → záložka: Možnosti → Zobrazit detailní tabulku výsledku → Výpočet. (V sloupci $r(X,Y)$ je realizace výběrového koeficientu korelace).

Correlations (SC_KORELACE)						
Marked correlations are significant at $p < .05000$						
(Casewise deletion of missing data)						
Var. X & Var. Y	Mean	Std.Dv.	$r(X,Y)$	r^2	t	p
Tržby	929,0909	322,2563				
Naklady	184,5455	28,7623	-0,847512	0,718276	-4,79021	0,000987

Výsledná korelace je $-0,8475$. Jedná se tedy o silnou negativní korelaci. Lze tedy pro začátek tvrdit, že mezi Tržbami a Denními náklady existuje negativní vztah.

Pro ověření statistické významnosti využijeme pivotovu statistiku $T_0 = \frac{R_{12}\sqrt{n-2}}{\sqrt{1-R_{12}^2}}$ řídící se studentovým rozdělením o $n - 2$ stupních volnosti, pokud nulová hypotéza platí. V sloupci t je realizace testové statistiky a v sloupci p je odpovídající p-hodnota.

Nyní stanovíme kritický obor:

$$W = (-\infty; -t_{1-\alpha/2}(n-2)) \cup (t_{1-\alpha/2}(n-2); \infty)$$

$$W = (-\infty; /2, 262157) \cup (2, 262157; \infty)$$

Jelikož $T_0 \in W$ zamítáme H_0 (zamítnutí nulové hypotézy je v souladu s výslednou p hodnotou, která je menší než hladina významnosti). Můžeme tedy s rizikem omylu 5% tvrdit, že záporný vztah mezi Tržbami a Denními Náklady byl prokázán.