

Příklad 1

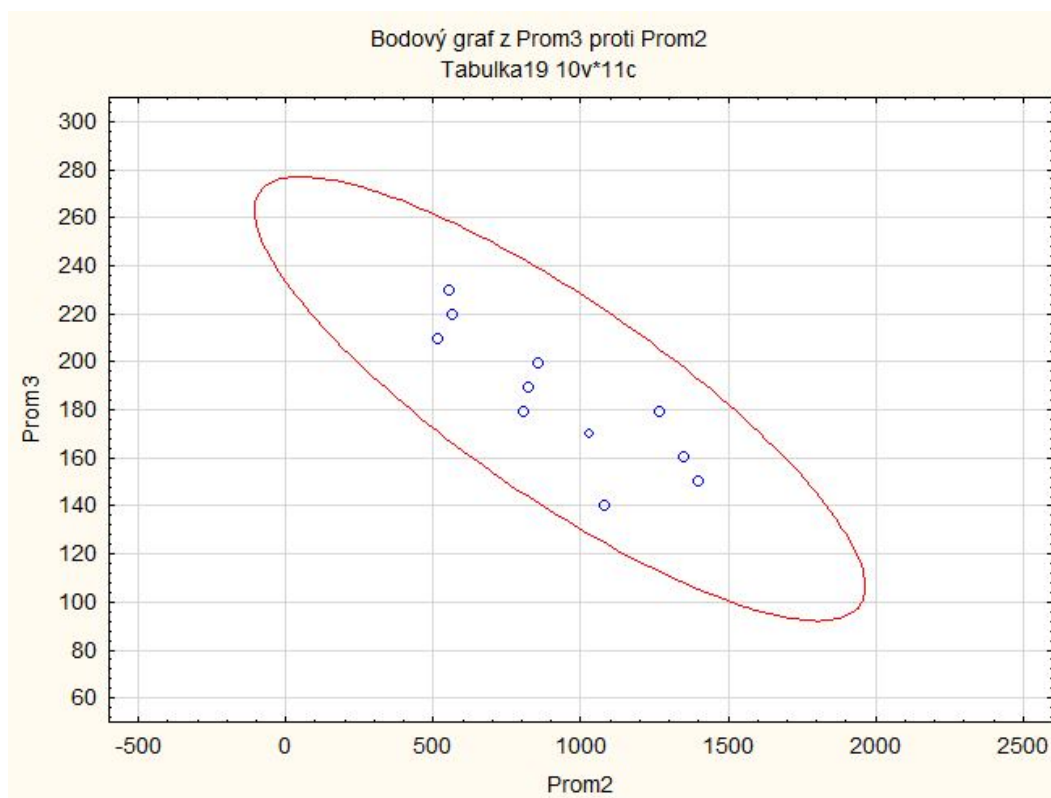
Marketérka Penzionu v Parku ve vinařské oblasti Čejkovic u Hodonína nám poskytla interview, ve kterém uvedla, že poskytovatelé penzionů obecně upřednostňují delší pobyty před kratšími. Důvodem pro ně je větší ziskovost vlivem snížení výše denních nákladů. Na podporu svého tvrzení nám poskytla data o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

Dny	1	1	1	2	2	2	3	3	3	4	4
Tržby za pobyt	310	350	360	800	820	850	1230	1280	1470	1550	1600
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. Výsledek je zachycen na následujícím obrázku:



KORELACE

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. V dalším kroku je tedy nutné vypočítat realizace výběrové průměry, rozptyly a kovarianci:

$$m_1 = \frac{1}{n} \sum_{i=1}^n x_i = 929,09, \quad m_2 = \frac{1}{n} \sum_{i=1}^n y_i = 184,54$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_1)^2 = 94408,26, \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_2)^2 = 752,07$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_2)(x_i - m_1) = -7141,32$$

Jelikož již máme veškeré mezivýpočty hotové můžeme přejít k samotnému výpočtu výběrové korelace r_{12} dle vzorce:

$$r_{12} = \frac{s_{12}}{\sqrt{s_1^2} \sqrt{s_2^2}} = \frac{-7141,32}{\sqrt{752,07} \sqrt{94408,26}} = -0,8475$$

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 pivotovú 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 prvním kroku si stanovíme nulovou hypotézu $\rho = 0$ proti alternativní hypotéze $\rho \neq 0$. Realizace statistiky T_0 získáme dosazením do výše zmíněného vzorce:

$$T_0 = \frac{R_{12} \sqrt{n-2}}{\sqrt{1-R_{12}^2}} = -4,790212651.$$

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 . Můžeme tedy s rizikem omylu 5% tvrdit, že záporný vztah mezi Tržbami a Denními Náklady byl prokázán.