Lineární regrese řešené příklady

Příklad 1

U sta lidí jsme zjišťovali ochotu platit (OP) za různé množství statku X. Jednotlivé hodnoty byly zprůměrovány a vyneseny do tabulky.

x	1	2	3	4	5	6	7	8	9	10
OP	290	365	420	445	501	598	635	687	750	880

- Odhadněte parametry regresní přímky $OP = \beta_0 + \beta_1 \cdot x$, která vystihuje závislost OP na X a interpretujte hodnoty regresních koeficientů

  Abychom získali parametry regresní přímky budeme muset vyřešit následující soustavu rovnic:

$$\hat{\beta}_1 = \frac{n \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2} = \frac{10,35644 - 56,5571}{10,385 - 55^2} = 60,6485$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 557,1 - 60,6485 \cdot 5,6 = 223,533$$

  Výsledná rovnice regresní přímky je tvaru $\hat{Y} = 223,533 + 60,6485 \cdot x$. Při zvýšení množství statku X o jednotku se zvětší OP průměrně o $60,6485$.

- Rozhodněte o kvalitě lineárního modelu pomocí koeficientu determinace

  K výpočtu koeficientu determinace je zapotřebí postupně zjistit hodnoty součtů následujících čtverců:

$$S_T = \sum_i (y_i - \bar{y})^2 \quad S_R = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 \cdot x_i - \bar{y})^2 \quad S_e = \sum_i (y_i - \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i)^2$$

  Pomocné výpočty:

Množství statku	Očekávaná Y	Rezidua	Normovaná rezidua	S_e	S_R	S_T
1	284,18	5,8181	0,237	33,851	74484,33	71342,41
2	344,83	20,16	0,824	406,81	45058,42	36902,41
3	405,47	14,521	0,593	210,86	22988,99	18796,41
4	466,12	-21,12	-0,86	446,36	8276,037	12566,41
5	526,77	-25,77	-1,05	664,38	919,5597	3147,21
6	587,42	10,575	0,432	111,84	919,5597	1672,81
7	648,07	-13,07	-0,53	170,89	8276,037	6068,41
8	708,72	-21,72	-0,88	471,81	22988,99	16874,01
9	769,36	-19,36	-0,79	375,18	45058,42	37210,41
10	830,01	49,981	2,042	2498,1	74484,33	104264,4
				5390,206	303454,7	308844,9

  Z tabulky vyplývá, že $S_T = 308844,9$, $S_R = 303454,7$ a $S_e = 5390,206$. Dosazením získáme hodnotu koeficientu (indexu) determinace jako $ID^2 = S_R/S_T = 0,982547$, ze kterého vyplývá, že $98,25\,\%$ variability ochoty platit je vysvětleno regresním vztahem.

- Testujte závislost Y na X pomocí t – testu

  Řešení se provede testováním regresního koeficientu u proměnné X. V prvním kroku stanovíme nulovou hypotézu $H_0 : \beta_1 = 0$ a příslušnou alternativní hypotézu $H_1 : \beta_1 \neq 0$. Ve druhém kroku vypočteme hodnotu směrodatné chyby

$$s(\hat{\beta}_1) = \sqrt{\frac{S_e}{n-2}} \cdot \sqrt{\frac{n}{n \cdot \sum x_i^2 - \left(\sum x_i\right)^2}} = \sqrt{\frac{5390,206}{10-2}} \cdot \sqrt{\frac{10}{10,385 - 55^2}} = 2,85779$$

  a následně hodnotu testové statistiky $t = \hat{\beta}_1 / s(\hat{\beta}_1) = 60,648/2,8577 = 21,222$. Na základě hodnoty $97,5\%$ kvantilu studentova rozdělení s 9 stupni volnosti, který je roven 2,26216 zamítáme nulovou hypotézu o nevýznamnosti regresního parametru β_1. Lineární regrese se jeví jako vhodný model a může být použit dále například k predikcím.

- Pomocí odhadnutého regresního modelu vypočítejte průměrnou ochotu platit za 4 jednotky statku

  Pro hodnotu $x = 4$ získáváme individuální předpověď dosazením do předchozího vztahu tím získáme výsledek $\hat{OP} = 466,127$.