5. Miary współzależności

5.3. Regresja liniowa

Interpretacja parametrów prostej regresji Y względem X:

$a_1 > 0$ jeśli „x" wzrośnie o 1 jednostkę, to „y" wzrośnie średnio o „a" jednostek

$a_1 < 0$ jeśli „x" wzrośnie o 1 jednostkę, to „y" spadnie średnio o „a" jednostek.

5. Miary współzależności

5.3. Regresja liniowa

Estymatory ocen parametrów strukturalnych równania regresji Y względem X są następujące:

$$a_1 = \frac{\operatorname{cov}(x; y)}{S^2(x)} = \frac{\frac{1}{N}\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\frac{1}{N}\sum\left(x_i - \overline{x}\right)^2}$$

$$a_0 = \overline{y} - a_1 \cdot \overline{x}$$

$$a_1 = r_{xy} \cdot \frac{s(y)}{s(x)}$$

5. Miary współzależności

5.3. Regresja liniowa

Estymatory ocen parametrów strukturalnych równania regresji X względem Y są następujące:

$$b_1 = \frac{\text{cov}(x;y)}{S^2(y)} = \frac{\frac{1}{N}\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\frac{1}{N}\sum\left(y_i - \bar{y}\right)^2}$$

$$b_0 = \bar{x} - b_1 \cdot \bar{y}$$

$$b_1 = r_{xy} \cdot \frac{s(x)}{s(y)}$$

5. Miary współzależności

5.3. Regresja liniowa

Zależność korelacyjna między dwoma cechami Y i X wyrażona przy pomocy równań regresji może być wyrażona przez współczynnik korelacji liniowej Pearsona, który jest symetryczny – jest to pierwiastek z iloczynu ocen parametrów kierunkowych poszczególnych funkcji regresji

$$r_{xy} = \sqrt{a_1 \cdot b_1}$$

5. Miary współzależności

5.3. Regresja liniowa

Reszta określa niedokładność szacunku i-tej wartości cechy.

$$e_i = y_i - \hat{y}_i$$

$$z_i = x_i - \hat{x}_i$$

Syntetycznym miernikiem jakości modelu jest tzw. wariancja resztowa:

$$S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} \qquad S_z^2 = \frac{\sum_{i=1}^{n} z_i^2}{n-2}$$

5. Miary współzależności

5.3. Regresja liniowa

Wariancja z uwagi na to, że jest mianowana w „kwadracie” jednostki fizycznej badanej zmiennej objaśnianej, nie ma interpretacji – pierwiastek z wariancji jest odchyleniem standardowym reszt – informuje o przeciętnej wartości reszty – o ile średnio wartości teoretyczne zmiennej objaśnianej różnią się od wartości empirycznych

$$S_e = \sqrt{S_e^2}$$

$$S_z = \sqrt{S_z^2}$$

5. Miary współzależności

5.3. Regresja liniowa

Wariancja i odchylenie standardowe reszt to bezwzględne mierniki jakości modelu – miernikiem względnym jest współczynnik zmienności resztowej

$$V_e = \frac{S_e}{\overline{Y}} * 100\%$$

$$V_z = \frac{S_z}{\overline{X}} * 100\%$$

5. Miary współzależności

5.3. Regresja liniowa

Współczynniki zmienności resztowej wyrażają procentowo przeciętny błąd modelu regresji względem przeciętnych wartości zmiennych objaśnianych – ile procent przeciętnej wartości (średniej) zmiennej objaśnianej stanowią przeciętne (standardowe) reszty – różnice między wartościami empirycznymi i teoretycznymi zmiennej objaśnianej.

Im wariancja reszt, odchylenie standardowe reszt, współczynnik zmienności resztowej bliższy wartości zera tym lepsza funkcja (model) regresji.

5. Miary współzależności

5.3. Regresja liniowa

Współczynnik rozbieżności (indeterminacji)- przyjmuje wartości w przedziale [0,1] - [0%,100%]. Ocenia w jakiej części zmiany cechy „y" nie są wyjaśnione zmianami cechy „x". Im jego wartość bliższa 0 tym lepsza funkcja regresji (model).

Współczynnik zbieżności (determinacji) -przyjmuje wartości z przedziału [0,1] - [0%,100%] Informuje o tym jaka część zmian cechy „y" jest wyjaśniona przez funkcję regresji (model). Im jego wartość bliższa 1 tym model lepszy.

Między nimi zachodzi zależność:

$$\varphi^2 + R^2 = 1$$

5. Miary współzależności

5.3. Regresja liniowa

Współczynnik rozbieżności $$\varphi^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Współczynnik determinacji $$R^2 = 1 - \varphi^2$$