WYKŁAD 3 20.10.2014

OCENA ESTYMACJI MODELU LINIOWEGO

1. Regresja ~~odwrotna~~ prosta

$$X = \beta_0 + B_0 Y + n$$

Dla rozpatrywanego przykładu relacji wielkości produkcji y i przepracowanych roboczogodzin x i regresja odwrotną jest postać

$$\hat{x} = 1,37 + 0,69y$$

Która z regresji jest poprawna?

Mogą być obie poprawne. Są sytuacje, że tylko jedna jest właściwa. Jeżeli wiemy, co jest przyczyną, a co skutkiem, to stosujemy regresję skutków względem przyczyny.

2. Miara dopasowania

Po oszacowaniu parametrów modelu należy zbadać stopień zgodności modelu z danymi empirycznymi w wybranej mierze dopasowania.

Wszystkie miary dopasowania bazują na wartościach reszt $e_i = y_i - \hat{y}$... na ~~wykładzie~~.

Zmienności zmiennej objaśnianej y określa się przez jej wariancję empiryczną

$$S_y^2 = \left(1/n \sum_{i=1}^{n} \ldots (y_i - \bar{y})\right)^2$$

Ponieważ 1/n jest stałą, to w badaniu zmienności zmiennej objaśnianej y będziemy brali pod uwagę wyrażenie $\sum \ldots (y_i - \bar{y})^2$.

3. Dekompozycja wzoru

Wyrażenie rozbijamy na dwa składniki

$$\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{\text{całkowita zmienność zmiennej objaśnianej Y}} = \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{\text{część zmienności zmiennej objaśnianej Y objaśnionej przez model}} + \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{część zmienności zmiennej objaśnianej Y nie objaśnionej przez model}} \qquad (1)$$

4. Współczynnik determinacji

Dzieląc obustronnie równanie (1) przez $\sum_{i=1}^{n} (y_i - \bar{y})$ otrzymamy:

$$1 = \frac{\sum^{n} (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} + \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

R^2 – współczynnik determinacji $\frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$

φ^2 – -||- zbieżności $\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$

5. Interpretacja R^2

$$1 = R^2 + \varphi^2, \qquad R^2 = 1 - \frac{\sum_i^n e_i^2}{\sum_i^n (y_i - \bar{y})^2}$$

R^2 Informuje, jaka część całkowitej zmienności zmiennej objaśnianej Y stanowi zmienność objaśniana przez model

Współczynnik determinacji przyjmuje wartości z odcinka $[0, 1]$.
Jeżeli $R^2 = 1$, to reszty $e_i = 0$ i model jest doskonale dopasowany do danych empirycznych, tzn. zmienne objaśniające całkowicie wyjaśniają zmienność zmiennej objaśnianej.

UWAGA:
Jeżeli zmienne objaśniające są mocno ze sobą skorelowane i mocno skorelowane ze ze zmienną objaśnianą, to wartość współczynnika R^2 jest wysoka i wnioski płynące z tego faktu są fałszywe. Takie sztuczne podwyższenie współczynnika determinacji nazywa się efektem KATALIZY. Usunięcie efektu katalizy polega na usunięciu z modelu zmiennych objaśniających mocno skorelowanych za...