

Prognozowanie- wiadomości wstępne

Prognozowanie to racjonalne wnioskowanie o zdarzeniach nieznanych na podstawie zdarzeń znanych.

Celem prognozy jest dostarczenie obiektywnych informacji potrzebnych do podejmowania decyzji.

Prognozy a symulacje.

Prognoza – *co będzie w momencie t ,*

Symulacja – *co by było gdyby*

Przykład

Z rozpatrywanego modelu wynika, że wydatki na prasę i książki stanowią 5% miesięcznych dochodów rodziny. Ustalono, że miesięczne dochody rodziny wyniosą 4000 zł. Możemy zatem postawić **prognozę**, że wydatki na prasę i książki wyniosą 200 zł.

Jeśli jednak wyznaczalibyśmy wydatki na prasę i książki dla różnych wariantów dochodu, np.
wydatki 190 zł dla dochodu 3800,
wydatki 200 zł dla dochodu 4000,
wydatki 220 zł dla dochodu 4400,
to byłyby **symulacje**.

Procedury prognozowania

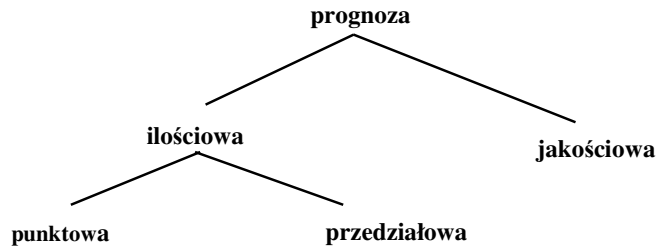
- Proste i intuicyjne (na podstawie prostych charakterystyk liczbowych),
- Ekonometryczne,
- Poprzez analogię,
- Prognozy ekspertów (heurystyczne),
- Wyznaczanie różnych scenariuszy rozwoju.

Prognozowane zmiany wartości badanego zjawiska mogą być:

- ilościowe (zgodne z dotychczasową prawidłowością np. trendem lub funkcją regresji),
- jakościowe (odejście od dotychczasowych prawidłowości)

Uproszczona klasyfikacja prognoz.

Ze względu na wartości prognozy:



Ze względu na okres prognozy:

- **Krótkookresowa** (na taki okres w którym mogą zachodzić tylko zmiany ilościowe),
- **Średniookresowa** (na taki okres w którym mogą zachodzić zmiany ilościowe i niewielkie zmiany jakościowe),
- **Długookresowa** (na taki okres w którym mogą zachodzić zarówno zmiany ilościowe jak i jakościowe).

W praktyce niekiedy podział ten odnosi się do zasięgu ekstrapolacji (liczba jednostek czasu wyjścia z prognoza w przyszłość) w porównaniu z liczbą danych:

do 10% - prognoza **krótkookresowa**,

od 10 do 20% - prognoza **średniookresowa**,

powyżej 20% - prognoza **długookresowa**,

Ponieważ wartości prognoz wyznaczamy w oparciu o dane, to muszą być one dobrej jakości.

Cechy danych decydujące o ich jakości:

- rzetelność,
- jednoznaczność,
- identyfikowalność,
- kompletność,
- aktualność,
- koszt (zbierania i opracowania),
- porównywalność (np. w zakresie: czasowym, terytorialnym, pojęciowym).

Etapy prognozowania:

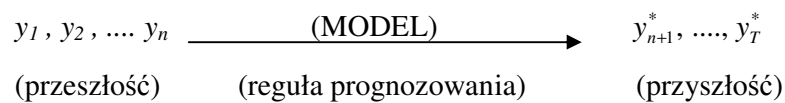
- Sformułowanie zadania prognostycznego
 - Określenie zmiennych prognozowanych,
 - Ustalenie celu prognozy,
 - Ustalenie horyzontu prognozy i warunków jej dopuszczalności
- Określenie przesłanek prognostycznych
 - Określenie czynników kształtujących badane zjawisko,
 - Zbieranie danych,
- Wybór metody prognozowania
- Wyznaczanie prognoz
- Ocena dopuszczalności prognoz
- Wykorzystanie prognozy
- Weryfikacja i monitorowanie (przy powtarzalności) prognozy.

Podstawowy schemat prognozowania.

Y - badane zjawisko,

y_t - obserwacje badanego zjawiska,

y_t^* - prognozowane wartości badanego zjawiska.



Bezwzględny błąd prognozy jest równy $y_\tau^* - y_\tau$,

Względny błąd prognozy jest równy $\frac{|y_\tau^* - y_\tau|}{y_\tau}$ (ma zwykle sens dla zjawisk o wartościach

dodatnich), można go wyrażać w procentach.

gdzie y_τ to prawdziwa wartość zjawiska w okresie prognozy.

Uwaga

Bezwzględny błąd prognozy niekiedy definiuje się jako $y_\tau - y_\tau^*$.

Względny błąd prognozy niekiedy definiuje się jako $\frac{|y_\tau^* - y_\tau|}{y_\tau^*}$.

Prawdziwą wartość błędu prognozy można wyznaczyć dopiero po ustaleniu prawdziwej wartości badanego zjawiska, wcześniej błąd można tylko **oszacować**.

Szacowanie błędu prognozy.

1. Na podstawie prognoz wygaśłych (ex post),
2. Metoda stochastyczna (ex ante).

Ad. 1. Wykorzystuje się informacje o trafności prognozowania w przeszłości. Przyjmuje się, że trafność prognoz przyszłych będzie podobna do trafności prognoz przeszłych. Prognozy wygaśłe używane do szacowania powinny być wyznaczone w ten sam sposób jak ostateczna prognoza. Jako oszacowanie błędu prognozy można np. przyjąć średnią z

modułów błędów bezwzględnych $\frac{1}{k} \sum_{t=1}^k |y_t^* - y_t| * 100\%$ lub względnych

$\frac{1}{k} \sum_{t=1}^k \left| \frac{y_t^* - y_t}{y_t} \right| * 100\%$ prognoz wygaśłych. Ten sposób szacowania błędu prognozy

zastosujemy przy modelach adaptacyjnych.

Ad. 2. Wykorzystuje się stochastyczne założenia o stosowanym modelu. Przyjmuje się, że błąd prognozy jest zbliżony do średniej rozbieżności między możliwymi wartościami prognozowanego zjawiska a możliwymi prognozami tego zjawiska w okresie prognozy.

Jako oszacowanie błędu prognozy można np. błąd średniokwadratowy $\sqrt{\frac{1}{k} \sum_{t=1}^k (y_t^* - y_t)^2}$

lub względny błąd średniokwadratowy $\sqrt{\frac{1}{k} \sum_{t=1}^k \left(\frac{y_t^* - y_t}{y_t} \right)^2} * 100\%$. Ten sposób szacowania

błędu prognozy zastosujemy przy modelach ekonometrycznych.

Niekiedy przyjmuje się, że prognoza jest **dopuszczalna**, gdy szacowany błąd nie przekracza 5 – 10%.

Schemat prognozowania na podstawie modelu ekonometrycznego $y = f(x)$

x_τ^* - wektor zmiennych objaśniających dla okresu prognozy.

Prognoza punktowa: $y_\tau^* = f(x_\tau^*)$.

Prognoza przedziałowa: $\langle y_\tau^* - \Delta_1, y_\tau^* + \Delta_2 \rangle$,

Zwykle $\Delta_1 = \Delta_2 = \Delta$ (= błąd bezwzględny prognozy przedziałowej).

Jakość prognozy w znacznym stopniu zależy od jakości zastosowanego modelu ekonometrycznego.

Oprócz tego

- Błąd prognozy powinien być mały,
- Przyjęte wartości zmiennych objaśniających powinny być wiarygodne,
- Okres prognozy powinien być sensowny.

Przykład

Rozpatrując model $Y = 200 - 10X$,

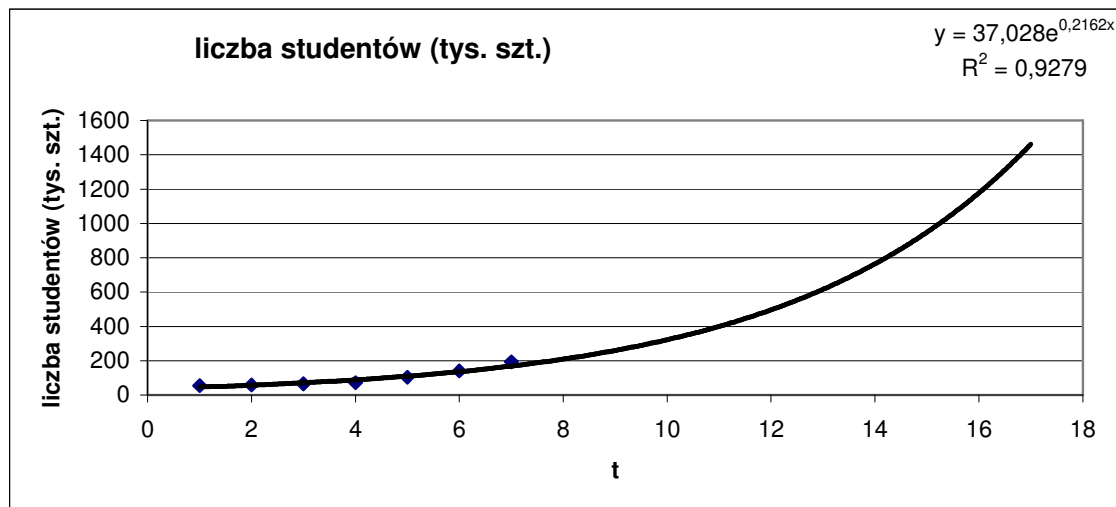
Y – jednostkowe koszty produkcji,

X – wielkość produkcji

W tym przypadku prognozy tracą sens dla $x > 20$.

Przykład

Liczba studentów kierunków ekonomicznych w Polsce (tys. osób) liczona na koniec roku akademickiego w latach 1991-97 wynosiła: 54, 58, 65, 71, 104, 140, 193.



Rozpatrując model $Y = f(t)$,

Y – liczba studentów,

t – rok

W tym przypadku prognoza np. na rok 2007 (ponad 1,46 mln osób) byłaby przesadna.

Przypomnienie informacji o jednorównaniowych liniowych modelach ekonometrycznych.

Jedna zmienna objaśniająca.

$$\hat{Y} = b_0 + b_1 X$$

Y zmienna objaśniana, X zmienna objaśniająca.

przybliżone wartości parametrów strukturalnych

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} =$$

$$= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \frac{s_Y}{s_X} r = \frac{\text{cov}(X, Y)}{s_X^2}$$

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

Wariancja resztowa.

Niech $e_i = y_i - \hat{y}_i$, gdzie $\hat{y}_i = b_0 + b_1 x_i$ wtedy

$$s_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} \quad \text{czyli} \quad s_e^2 = \frac{\sum_{i=1}^n y_i^2 - b_0 \sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i y_i}{n-2}$$

$s_e = \sqrt{s_e^2}$ oznacza średnie (standardowe) odchylenie od prostej regresji.

Współczynnik determinacji

$$R^2 \in \langle 0, 1 \rangle$$

(określa jaką część całkowitej zmienności cechy Y wyjaśnia model regresji liniowej)

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = \frac{b_0 \sum y_i + b_1 \sum x_i y_i - n \bar{y}^2}{\sum y_i^2 - n(\bar{y})^2} =$$

$$= \frac{b_1 (\sum x_i y_i - n \bar{x} \bar{y})}{\sum y_i^2 - n(\bar{y})^2} = \frac{\text{cov}^2(X, Y)}{s_X^2 s_Y^2} = r^2$$

k - zmiennych objaśniających.

X_i - zmienne objaśniające.

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} \quad Y = Xb$$

$$b = (X^T X)^{-1} X^T Y$$

Własność

$$Y^T Y = \sum_{i=1}^n y_i^2$$

$$\text{Dla } k = 1 \quad X^T X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}, \quad X^T Y = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{bmatrix},$$

$$\text{Dla } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ gdy } ad - cb \neq 0$$

$$\text{to} \quad A^{-1} = \frac{1}{ad - cb} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Wariancja resztowa.

Niech, $e = Y - \hat{Y}$, gdzie $\hat{Y} = Xb$ wtedy

$$S_e^2 = S^2 = \frac{e^T e}{n - (k+1)} = \frac{\sum_{i=1}^n e_i^2}{n - (k+1)} = \frac{1}{n - (k+1)} (Y^T Y - b^T X^T Y)$$

Współczynnik determinacji

$$R^2 = \frac{b^T X^T Y - n \bar{Y}^2}{Y^T Y - n \bar{Y}^2} \quad R^2 \in \langle 0, 1 \rangle$$