# STATYSTYKA OPISOWA

Dr Alina Gleska

Instytut Matematyki WE PP

28 września 2018

1. Modele tendencji rozwojowej

**Metoda analityczna** – przyjmujemy założenie, że zmiany zjawiska w czasie można przedstawić jako funkcję zmiennej czasowej z dokładnością do składnika losowego.

Zmienna czasowa $t$ jest to zmienna, której realizacje będące kolejnymi liczbami całkowitymi, są przyporządkowane poszczególnym jednostkom czasowym według zasady następstwa czasowego. Najczęściej są to liczby naturalne $t = 1, 2, \ldots, n$, gdzie $n$ – liczba obserwacji w szeregu czasowym.

Model trendu – konstrukcja formalna, za pomocą której przedstawia się przebieg badanego zjawiska w czasie. Zapis formalny ma postać:

$$Y_t = f(t) + U_t, \quad t = 1, 2, \ldots, n,$$

gdzie $f(t)$ – funkcja trendu, $U_t$ – składnik losowy.

Składnik losowy charakteryzuje efekty oddziaływania na badane zjawisko wahań przypadkowych. Zakłada się przy tym, że jego wartość oczekiwana wynosi 0, natomiast wariancja jest skończona.

Podstawowym problemem w metodzie analitycznej jest określenie postaci analitycznej funkcji $f(t)$. Przy jej wyborze kierujemy się zazwyczaj następującymi przesłankami:

- wybrana funkcja powinna być prosta analitycznie;
- parametry strukturalne funkcji powinny być interpretowalne;
- należy wybierać funkcje, których parametry można estymować klasyczną metodą najmniejszych kwadratów;
- wybrana funkcja powinna być zgodna z empirycznymi wynikami badań.

Powyższe przesłanki powodują, że w praktyce najczęściej wykorzystywanymi funkcjami są:

- funkcja liniowa: $Y_t = \alpha_0 + \alpha_1 t$;
- funkcja paraboliczna: $Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$;
- funkcja potęgowa: $Y_t = \alpha_0 \cdot t^{\alpha_1}$;
- funkcja wykładnicza: $Y_t = \alpha_0 \cdot \alpha_1^t$;
- funkcja hiperboliczna: $Y_t = \alpha_0 + \alpha_1 \cdot \frac{1}{t}$;
- funkcja logistyczna: $Y_t = \frac{\alpha}{1+\beta e^{-rt}}$.

Wszystkie wymienione funkcje można sprowadzić, po dokonaniu pewnych transformacji, do postaci liniowej, dlatego skupimy się na estymacji liniowej funkcji trendu.

Załóżmy więc, że funkcja trendu ma postać:

$$y_t = \alpha_0 + \alpha_1 t + u_t,$$

gdzie:
$y_t$ – poziom badanego zjawiska w jednostce czasu $t$;
$t$ – zmienna czasowa $(t = 1, 2, \ldots, n)$;
$u_t$ – realizacje składnika losowego;
$\alpha_0, \alpha_1$ – parametry liniowej funkcji trendu.

Parametr $\alpha_1$ określa, jaki jest przeciętny okresowy przyrost badanej zmiennej $Y$ w analizowanym przedziale czasowym, zaś parametr $\alpha_0$ wskazuje na teoretyczny poziom tej zmiennej w tym okresie, dla którego $t = 0$.

Parametry $\alpha_1$ i $\alpha_0$ szacuje się stosując klasyczną metodę najmniejszych kwadratów, otrzymując ich oceny w postaci:

$$a_1 = \frac{\sum\limits_{i=1}^{n} y_{t_i} \cdot t_i - n \cdot \overline{y} \cdot \overline{t}}{\sum\limits_{i=1}^{n} t_i^2 - n\overline{t}^2}, \quad a_0 = \overline{y} - a_1 \overline{t},$$

gdzie:
$\bar{y}$ – średnia arytmetyczna zmiennej $Y$ w przedziale czasowym $[1, n]$;
$\bar{t}$ – średnia arytmetyczna zmiennej czasowej,
$y_{t_i}$ – realizacje zmiennej $Y$ w okresie $t_i$ $(i = 1, 2, \ldots, n)$.

Po oszacowaniu parametrów strukturalnych funkcji trendu konieczna jest ocena jej jakości. Ocena ta dokonywana jest w oparciu o pewne kryteria.

## KRYTERIUM BŁĘDU LOSOWEGO

Obliczamy dwie wartości:

- odchylenie standardowe składnika losowego (resztowego)

$$S_u = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n-k}},$$

gdzie:
$Y_t$ – wartości empiryczne zmiennej ($t = 1, 2, \ldots, n$),
$\hat{Y}_t$ – wartości teoretyczne (wartości trendu) zmiennej objaśnianej,
$k$ – liczba szacowanych parametrów trendu (czyli $k = 2$ dla funkcji liniowej).
Parametr ten mówi, jaki - przeciętnie biorąc - błąd popełniamy szacując poziom badanego zjawiska na podstawie funkcji trendu.

- współczynnik zmienności losowej

$$V_u = \frac{S_u}{\overline{Y}} \cdot 100\%,$$

gdzie $\overline{Y}$ – średnia arytmetyczna zmiennej objaśnianej.

Miara ta określa, jaki procent średniego poziomu badanego zjawiska stanowi odchylenie standardowe składnika resztowego. Z reguły przyjmuje się, że jeżeli

$$V_u \leq (10 - 15)\%,$$

to błąd losowy modelu trendu można uznać za relatywnie mały i w konsekwencji ocenić model za dopuszczalny z punktu widzenia tego kryterium.

## KRYTERIUM DOKŁADNOŚCI OPISU BADANEGO ZJAWISKA

Obliczamy dwie wartości:

- współczynnik zgodności:

$$\phi^2 = \frac{\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2} = \varphi^2.$$

Miernik ten przyjmuje wartości z przedziału $[0, 1]$ i określa, jaką część badanego zjawiska nie opisuje wybrana funkcja trendu. Im współczynnik zgodności jest bliższy 0, tym dokładność opisu trendu przez zastosowaną funkcję jest lepsza.

- współczynnik determinacji:

$$R^2 = 1 - \phi^2 = 1 - \varphi^2.$$

Miara ta również przyjmuje wartości z przedziału $[0, 1]$ i określa, jaka część zmienności badanego zjawiska jest wyjaśniona przez funkcję trendu. Jeśli $R^2 \to 1$, tym model trendu jest lepszy. Umownie przyjmuje się, że jeśli $\phi^2 \leq 20\%$, to model trendu jest dopuszczalny z punktu widzenia tego kryterium. Warto jednak zaznaczyć, że o przyjętej normie granicznej decyduje podmiot badający.

## KRYTERIUM PRECYZJI SZACUNKU PARAMETRÓW MODELU TRENDU

W tym celu oblicza się błędy średnie szacunku ocen parametrów funkcji trendu, a następnie porównuje się je z tymi ocenami. Jeśli otrzymana relacja (umownie biorąc) jest większa od 2, to można uznać, że dany parametr funkcji trendu został oszacowany precyzyjnie.

Jeśli funkcja trendu jest liniowa, to błędy średnie szacunku ocen parametrów dane są wzorami:

$$D(a_1) = \frac{S_u}{\sqrt{\sum_{t=1}^{n}(t-\bar{t})^2}} = \frac{S_u}{\sqrt{\sum_{t=1}^{n} t^2 - n\bar{t}^2}},$$

$$D(a_0) = \sqrt{\frac{S_u^2 \cdot \sum_{t=1}^{n} t^2}{n \cdot \sum_{t=1}^{n} (t - \bar{t})^2}} = \sqrt{\frac{S_u^2 \cdot \sum_{t=1}^{n} t^2}{n \cdot (\sum_{t=1}^{n} t^2 - n\bar{t}^2)}}.$$

Oceny parametrów trendu liniowego uznamy za precyzyjne, jeśli:

$$t_1 = \frac{a_1}{D(a_1)} > 2, \quad t_2 = \frac{a_0}{D(a_0)} > 2.$$

Oszacowaną funkcję trendu można następnie wykorzystać do sporządzenia prognozy. Prognozę zmiennej $Y$ na okres $T$ otrzymuje się przez ekstrapolację funkcji trendu, czyli przez podstawienie do modelu w miejsce zmiennej czasowej $t$ jej realizacji właściwej dla okresu prognozowanego $T$.

Załóżmy, że realizacja ta wynosi $t_T$. Wtedy $y^P_{t_T} = f(t_T)$, $t_T > n$, czyli w przypadku liniowej funkcji trendu:

$$y^P_T = a_0 + a_1 t_T.$$

Tak skonstruowana prognoza jest prognozą punktową. Ocena jej jakości dokonywana jest przez obliczenie średniego błędu prognozy ex ante według wzoru:

$$D(y^P_T) = S_u \cdot \sqrt{1 + \frac{1}{n} + \frac{(t_T - \bar{t})^2}{\sum\limits_{t=1}^{n} t^2 - n\bar{t}^2}}$$

oraz względnego błędu prognozy ex ante według wzoru:

$$V_P = \frac{D(y^P_T)}{y^P_T} \cdot 100\%.$$

Z reguły przyjmuje się, że jeżeli:

- $V_P \leq 5\%$ – to prognoza jest wysoce precyzyjna,
- $5\% < V_P \leq 10\%$ – to prognoza jest dostatecznie precyzyjna,
- $V_P > 10\%$ – to prognoza ma niedostateczną precyzję.

Należy jednak wyraźnie podkreślić, że o tym, czy prognozę przyjąć, czy też odrzucić decyduje jej odbiorca.

Przykład.