STATYSTYKA OPISOWA

Dr Alina Gleska

Instytut Matematyki WE PP

12 maja 2019

1. Regresja liniowa
2. Regresja krzywoliniowa

Przypomnijmy - w przypadku regresji liniowej dla jednej zmiennej objaśniającej $Y = bX + a$ współczynniki b i a obliczamy zgodnie z wzorami:

$$b = \frac{cov(X,Y)}{S_X^2} = \frac{\sum_{i=1}^{n} xy - n\overline{xy}}{\sum_{i=1}^{n} x^2 - n(\overline{x})^2} = \frac{\sum_{i=1}^{n} xy - (\sum_{i=1}^{n} x \sum_{i=1}^{n} y)/n}{\sum_{i=1}^{n} x^2 - (\sum_{i=1}^{n} x)^2/n},$$

$$a = \overline{y} - b\overline{x} = (\sum_{i=1}^{n} y - b\sum_{i=1}^{n} x)/n.$$

Ogólny model liniowy $Y = b_1X_1 + b_2X_2 + \cdots + b_kX_k$
Jeśli w modelu występuje wyraz wolny przyjmujemy, że występuje przy nim sztuczna zmienna objaśniająca , mająca zawsze wartość 1.

Przy wyznaczaniu modelu ogólnego korzystamy z zapisu macierzowego. Oznaczenia:

- b - pionowy wektor parametrów, tzn. $b = [b_k]$, $k = 1, 2, \ldots, K$;
- X - macierz wyników obserwacji zmiennych objaśniających (dane wejściowe);
- y - pionowy wektor wyników obserwacji zmiennej objaśnianej.

Wyznaczony klasyczną MNK wektor parametrów modelu liniowego jest rozwiązaniem układu tzw. równań normalnych

$$(X'X)b = X'y.$$

W postaci jawnej wektor b wyraża się wzorem:

$$b = (X'X)^{-1}X'y.$$

Wzór ten jest jednoznaczny, o ile macierz $X'X$ jest nieosobliwa. By tak było, rząd macierzy X musi być równy liczbie parametrów. W szczególności, koniecznie

- liczba nie powtarzających się obserwacji musi być większa od liczby parametrów (czyli również $n > K$);
- żadna kolumna macierzy X nie może być liniowo zależna od innych (w szczególności dwie kolumny nie mogą być identyczne).

Jak rozumieć macierz $X'X$ i wektor $X'y$?

W przypadku modelu $Y = bX + a = b_1X_1 + b_2X_2$, gdzie $b_2 = a$, $X_2 = 1$ mamy:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1^2 \end{bmatrix} \quad X'y = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}.$$

Z kolei dla modelu $Y = b_1X_1 + b_2X_2 + b_3 1$ mamy:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} 1^2 \end{bmatrix} \quad X'y = \begin{bmatrix} \sum_{i=1}^{n} x_{i1} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}.$$

EKONOMETRIA I BADANIA OPERACYJNE
Zagadnienia podstawowe
Redakcja naukowa Bogusław GUZIK
Wydawnictwo Akademii Ekonomicznej w Poznaniu

WERYFIKACJA MODELU LINIOWEGO WYZNACZONEGO KLASYCZNĄ MNK

Minimalny zestaw postulatów:

- model ekonometryczny nie może budzić zastrzeżeń merytorycznych,
- model powinien być bardzo dobrze dopasowany do danych empirycznych,
- wszystkie zmienne objaśniające modelu muszą być istotne.

Ogół działań w zakresie weryfikacji modelu można podzielić na:

- weryfikację merytoryczną (postulat 1),
- weryfikację statystyczną (postulat 2 i 3).

WERYFIKACJA MERYTORYCZNA polega na sprawdzeniu, czy model ekonometryczny jest zgodny z wiedzą ekonomiczną na temat badanego zjawiska , teorią ekonomii, wreszcie - ze zdrowym rozsądkiem. Podczas weryfikacji merytorycznej badamy np.

- czy sensowne są znaki parametrów modelu;
- czy skala parametrów jest do przyjęcia;
- czy model można sensownie ekstrapolować;
- czy z modelu dla zmiennej Y wynikają sensowne modele dla zmiennych związanych z badaną zmienną.

Dopasowanie modelu. Współczynnik determinacji

WSPÓŁCZYNNIK DETERMINACJI R^2 wskazuje, jaka część ogólnej zaobserwowanej zmienności zmiennej objaśnianej została wyjaśniona przez model ekonometryczny.

WSPÓŁCZYNNIK INDETERMINACJI φ^2 (inaczej: współczynnik rozbieżności wr) mierzy natomiast tę część zaobserwowanej zmienności zmiennej objaśnianej, która nie została przez model wyjaśniona.

Można przyjąć, że dobry model to taki, w którym $R^2 > 90\%$, a $\varphi^2 < 10\%$.

Są różne definicje tych współczynników. Standardowo przyjmuje się, że:

- miarą ogólnej zaobserwowanej zmienności zmiennej Y jest suma kwadratów odchyleń wartości tej zmiennej od jej średniej, czyli tzw. ogólna suma kwadratów OSK (po angielsku Sum of Squares Total SST):

$$OSK = \sum (y - \overline{y})^2; \tag{1}$$

- miarą obserwowanej zmienności Y, nie wyjaśnionej przez model jest suma kwadratów odchyleń zmiennej Y od modelu, czyli suma kwadratów reszt (ang. Sum of Squares for Errors SSE):

$$SKR = \sum (y - \hat{y})^2; \tag{2}$$

Współczynnik rozbieżności wr (lub indeterminacji φ^2) określony jest wzorem:

$$wr = \frac{SKR}{OSK} = \frac{SSE}{SST}. \tag{3}$$

Współczynnik determinacji R^2 określany jest jako

$$R^2 = 1 - wr = 1 - \varphi^2, \tag{4}$$

o ile $wr \leq 1$. Zauważmy, że

$$R^2 = \frac{RSK}{OSK},$$

gdzie $RSK = OSK - SKR$ to tzw. regresyjna suma kwadratów (ang. Sum of Squares for Regression SSR).

Obliczanie ogólnej sumy kwadratów OSK oraz sumy kwadratów reszt bezpośrednio z definicji może być kłopotliwe. OSK łatwiej obliczyć jako:

$$OSK = \sum y^2 - 2\sum y\bar{y} + \sum \bar{y}^2 = \sum y^2 - \frac{(\sum y)^2}{n}, \tag{5}$$

zaś SKR - w przypadku modelu liniowego wyznaczanego klasyczną MNK ze wzoru

$$SKR = \sum y^2 - b'X'y. \tag{6}$$

Dla modelu z jedną zmienną objaśniającą $Y = bX + a$ będzie to

$$SKR = \sum y^2 - (b\sum xy + a\sum y). \tag{7}$$

Z definicji (3) wynika, że współczynnik rozbieżności φ^2 przyrównuje zmienność zmiennej objaśnianej Y wokół modelu (co reprezentowane jest przez SKR) do zmiennej Y wokół średniej (co reprezentowane jest przez OSK). Współczynnik rozbieżności porównuje więc dwa opisy zmiennej: opis poprzez model do opisu poprzez średnią.

Iloraz $\frac{1}{\varphi^2}$ można przyjąć za wskaźnik tego, ile razy opis zmiennej Y poprzez model jest lepszy od opisu poprzez średnią.

ISTOTNOŚĆ ZMIENNYCH OBJAŚNIAJĄCYCH

Zmienna objaśniająca jest istotna, gdy w zauważalny (wyraźny) sposób wpływa na zmienną objaśnianą.

W przypadku modelu liniowego $Y = b_1 X_1 + \cdots + b_k X_k$ zmienna jest istotna, gdy parametr przy niej stojący jest istotnie różny od zera.

Parametry modelu ekonometrycznego praktycznie zawsze są niezerowe, dlatego trzeba sprawdzić ich istotność inaczej. Pamiętajmy, że materiał statystyczny użyty do wyznaczenia parametrów modelu jest na ogół niewielkim fragmentem zbioru wszystkich możliwych wyników obserwacji zmiennej objaśnianej i zmiennych objaśniających. W zbiorze możliwych wyników obserwacji zależność zmiennej Y od zmiennej X może w ogóle nie mieć miejsca, a to, że dla danego materiału statystycznego uzyskaliśmy parametr niezerowy, mogło być tylko przypadkowe.

Wszystkie zmienne objaśniające modelu ekonometrycznego muszą być istotne

Badanie istotności może odbywać się w różny sposób. Standardowy polega na wykorzystaniu założeń stochastycznych (probabilistycznych) o sposobie powstawania wyników obserwacji zmiennej objaśnianej. Założenia takie można znaleźć w podręcznikach ekonometrii.

W przypadku modeli liniowych, spośród wielu układów założeń stochastycznych, najczęściej przyjmuje się założenia tzw. klasycznej normalnej regresji liniowej. Przy tych założeniach - i gdy model wyznaczono klasyczną MNK - badanie istotności zmiennej przebiega następująco:

1. obliczamy tzw. empiryczną statystykę Studenta, t_K, dotyczącą badanej zmiennej objaśniającej;
2. ustalamy krytyczną wartość statystyki Studenta, t_{KR};
3. porównujemy moduł empirycznej statystyki Studenta z wartością krytyczną.

- Zmienną objaśniającą uznaje się za istotną, jeśli związana z nią statystyka empiryczna jest co do modułu większa od wartości krytycznej, tzn. $|t_K| > t_{KR}$.
- Jeśli jest odwrotnie, tzn. $|t_K| \leq t_{KR}$, to badaną zmienną objaśniającą uznajemy za nieistotną.
- Wartość krytyczną t_{KR} odczytujemy z tablic jako wartość rozkładu t-Studenta przy przyjętym poziomie istotności α (zazwyczaj $\alpha = 5\%$) oraz dotyczącej badanego modelu liczbie stopni swobody Q:

$$Q = n - K, \tag{8}$$

gdzie n jest liczbą wyników obserwacji, a K liczbą szacowanych parametrów.

- Programy informatyczne zazwyczaj podają wartości empirycznych statystyk t-Studenta. Niekiedy jednak nie wyświetla się ich, lecz podaje tzw. szacunkowe błędy średnie d_k (związane z poszczególnymi parametrami modelu). Ta informacja wystarcza do obliczenia statystyki Studenta, gdyż

$$\text{empiryczna statystyka Studenta} = \frac{\text{parametr modelu}}{\text{szacunkowy błąd średni}},$$

$$t_k = \frac{b_k}{d_k} \quad (k = 1, 2, \ldots, K). \tag{9}$$

Jeśli natomiast musimy sami obliczyć empiryczne statystyki t-Studenta, postępujemy następująco:

- najpierw liczymy oszacowanie odchylenia standardowego składników losowych s:

$$s = \sqrt{\frac{SKR}{Q}}; \tag{10}$$

- następnie liczymy tzw. szacunkowe błędy średnie d_k:

$$d_k = s\sqrt{c_k}, \tag{11}$$

  gdzie c_k jest k-tym przekątniowym elementem macierzy odwrotnej $(X'X)^{-1}$;
- na koniec, na podstawie (9), liczymy empiryczną statystykę Studenta.

Interpretacja szacunkowego błędu średniego

Założenia stochastyczne regresji liniowej m.in. orzekają, że

- istnieje prawdziwy model $\phi = \beta_1 X_1 + \cdots + \beta_k X_k$ zmiennej Y, z prawdziwymi parametrami $\beta_1, \ldots, \beta_k$. Jest to tzw. model hipotetyczny zmiennej Y;
- wartość y zmiennej Y w danej obserwacji jest jedną z wielu możliwych realizacji zmiennej Y, jakie mogły się w danej obserwacji ujawnić. Możliwość innych wartości zmiennej Y wynika z działania tzw. składnika losowego, reprezentującego np. wpływ okoliczności losowych.

W tej sytuacji użyty do wyznaczenia parametrów modelu materiał statystyczny jest jednym z wielu możliwych materiałów statystycznych, a wyznaczone przez nas parametry $b_1, \ldots, b_k$ modelu ekonometrycznego są jedną z wielu możliwych ocen parametrów $\beta_1, \ldots, \beta_k$ modelu hipotetycznego. W szczególności parametr b_k jest jedną z wielu możliwych ocen parametru hipotetycznego β_k.

Szacunkowy błąd średni d_k jest oceną rozbieżności możliwych ocen parametru β_k wokół tego parametru.

Jest oczywiście tym lepiej, im szacunkowy błąd średni jest mniejszy. Wówczas możliwe oceny parametru β_k mniej się odchylają od tego parametru. A ponieważ jedną z tych możliwych ocen jest nasza ocena b_k dlatego sądzimy, że im szacunkowy błąd średni d_k jest mniejszy, tym przypuszczalna precyzja estymacji β_k przez b_k jest większa.

Interpretacja oszacowania odchylenia standardowego składników losowych

Oszacowanie odchylenia standardowego składników losowych s jest oceną rozbieżności możliwych wartości zmiennej objaśnianej wokół modelu hipotetycznego.

Przykład.

Na podstawie 20 obserwacji otrzymano następujący model ekonometryczny dla popytu na samochody krajowe (PSK) względem dochodów (DO), indeksu cen samochodów krajowych (CSK) oraz indeksu cen samochodów importowanych (CSI):

$$\begin{array}{ccccccccc} PSK & = & 1,2DO & - & 0,23CSK & + & 0,12CSI & + & 3,4 \\ & & (4,9) & & (3,2) & & (1,9) & & (7,3) \end{array}$$

$$R^2 = 0,987.$$

Pod parametrami modelu ekonometrycznego podano moduły empirycznych statystyk Studenta. Przy 5% poziomie istotności sprawdzić istotność zmiennych objaśniających.

Ponieważ $n = 20$, $K = 4$, więc liczba stopni swobody $Q = 20 - 4 = 16$. Wartość krytyczną odczytujemy z tablic rozkładu t-Studenta dla 5% poziomu istotności i 16 stopni swobody. Jest to liczba $t_{KR} = 2,1199$.

Weryfikacja modelu

Wartości krytyczne rozkładu t-Studenta

$X \sim t_v$ - X zmienna losowa o rozkładzie t-Studenta z liczbą stopni swobody v,
α - poziom istotności,
$t_{v,\alpha}$ - wartość krytyczna - liczba taka, że $P(|X| > t_{v,\alpha}) = \alpha$

v \ a	0,400	0,300	0,200	0,100	0,050	0,025	0,025	0,010	0,005	0,001
1	1,3764	1,9626	3,0777	6,3137	12,7062	25,4519	25,4519	63,6559	127,3211	636,5776
2	1,0607	1,3862	1,8856	2,9200	4,3027	6,2054	6,2054	9,9250	14,0892	31,5998
3	0,9785	1,2498	1,6377	2,3534	3,1824	4,1765	4,1765	5,8408	7,4532	12,9244
4	0,9410	1,1896	1,5332	2,1318	2,7765	3,4954	3,4954	4,6041	5,5975	8,6101
5	0,9195	1,1558	1,4759	2,0150	2,5706	3,1634	3,1634	4,0321	4,7733	6,8685
6	0,9057	1,1342	1,4398	1,9432	2,4469	2,9687	2,9687	3,7074	4,3168	5,9587
7	0,8960	1,1192	1,4149	1,8946	2,3646	2,8412	2,8412	3,4995	4,0294	5,4081
8	0,8889	1,1081	1,3968	1,8595	2,3060	2,7515	2,7515	3,3554	3,8325	5,0414
9	0,8834	1,0997	1,3830	1,8331	2,2622	2,6850	2,6850	3,2498	3,6896	4,7809
10	0,8791	1,0931	1,3722	1,8125	2,2281	2,6338	2,6338	3,1693	3,5814	4,5868
11	0,8755	1,0877	1,3634	1,7959	2,2010	2,5931	2,5931	3,1058	3,4966	4,4369
12	0,8726	1,0832	1,3562	1,7823	2,1788	2,5600	2,5600	3,0545	3,4284	4,3178
13	0,8702	1,0795	1,3502	1,7709	2,1604	2,5326	2,5326	3,0123	3,3725	4,2209
14	0,8681	1,0763	1,3450	1,7613	2,1448	2,5096	2,5096	2,9768	3,3257	4,1403
15	0,8662	1,0735	1,3406	1,7531	2,1315	2,4899	2,4899	2,9467	3,2860	4,0728
16	0,8647	1,0711	1,3368	1,7459	2,1199	2,4729	2,4729	2,9208	3,2520	4,0149
17	0,8633	1,0690	1,3334	1,7396	2,1098	2,4581	2,4581	2,8982	3,2224	3,9651
18	0,8620	1,0672	1,3304	1,7341	2,1009	2,4450	2,4450	2,8784	3,1966	3,9217
19	0,8610	1,0655	1,3277	1,7291	2,0930	2,4334	2,4334	2,8609	3,1737	3,8833
20	0,8600	1,0640	1,3253	1,7247	2,0860	2,4231	2,4231	2,8453	3,1534	3,8496
21	0,8591	1,0627	1,3232	1,7207	2,0796	2,4138	2,4138	2,8314	3,1352	3,8193
22	0,8583	1,0614	1,3212	1,7171	2,0739	2,4055	2,4055	2,8188	3,1188	3,7922
23	0,8575	1,0603	1,3195	1,7139	2,0687	2,3979	2,3979	2,8073	3,1040	3,7676
24	0,8569	1,0593	1,3178	1,7109	2,0639	2,3910	2,3910	2,7970	3,0905	3,7454
25	0,8562	1,0584	1,3163	1,7081	2,0595	2,3846	2,3846	2,7874	3,0782	3,7251
26	0,8557	1,0575	1,3150	1,7056	2,0555	2,3788	2,3788	2,7787	3,0669	3,7067
27	0,8551	1,0567	1,3137	1,7033	2,0518	2,3734	2,3734	2,7707	3,0565	3,6895
28	0,8546	1,0560	1,3125	1,7011	2,0484	2,3685	2,3685	2,7633	3,0470	3,6739
29	0,8542	1,0553	1,3114	1,6991	2,0452	2,3638	2,3638	2,7564	3,0380	3,6595
30	0,8538	1,0547	1,3104	1,6973	2,0423	2,3596	2,3596	2,7500	3,0298	3,6460
31	0,8534	1,0541	1,3095	1,6955	2,0395	2,3556	2,3556	2,7440	3,0221	3,6335
32	0,8530	1,0535	1,3086	1,6939	2,0369	2,3518	2,3518	2,7385	3,0149	3,6218
33	0,8526	1,0530	1,3077	1,6924	2,0345	2,3483	2,3483	2,7333	3,0082	3,6109
34	0,8523	1,0525	1,3070	1,6909	2,0322	2,3451	2,3451	2,7284	3,0020	3,6007
35	0,8520	1,0520	1,3062	1,6896	2,0301	2,3420	2,3420	2,7238	2,9961	3,5911
40	0,8507	1,0500	1,3031	1,6839	2,0211	2,3289	2,3289	2,7045	2,9712	3,5510
45	0,8497	1,0485	1,3007	1,6794	2,0141	2,3189	2,3189	2,6896	2,9521	3,5203
50	0,8489	1,0473	1,2987	1,6759	2,0086	2,3109	2,3109	2,6778	2,9370	3,4960
55	0,8482	1,0463	1,2971	1,6730	2,0040	2,3044	2,3044	2,6682	2,9247	3,4765
60	0,8477	1,0455	1,2958	1,6706	2,0003	2,2990	2,2990	2,6603	2,9146	3,4602
65	0,8472	1,0448	1,2947	1,6686	1,9971	2,2945	2,2945	2,6536	2,9060	3,4466
70	0,8468	1,0442	1,2938	1,6669	1,9944	2,2906	2,2906	2,6479	2,8987	3,4350
75	0,8464	1,0436	1,2929	1,6654	1,9921	2,2873	2,2873	2,6430	2,8924	3,4249
80	0,8461	1,0432	1,2922	1,6641	1,9901	2,2844	2,2844	2,6387	2,8870	3,4164
85	0,8459	1,0428	1,2916	1,6630	1,9883	2,2818	2,2818	2,6349	2,8822	3,4086
90	0,8456	1,0424	1,2910	1,6620	1,9867	2,2795	2,2795	2,6316	2,8779	3,4019
95	0,8454	1,0421	1,2905	1,6611	1,9852	2,2775	2,2775	2,6286	2,8741	3,3958
100	0,8452	1,0418	1,2901	1,6602	1,9840	2,2757	2,2757	2,6259	2,8707	3,3905

Zróbmy zatem analizę wiedząc, że $t_{KR} = 2,12$.

- Empiryczna statystyka Studenta dla dochodu (równa 4,9) jest większa od wartości krytycznej; dochód jest więc istotną zmienną objaśniającą.
- W przypadku cen samochodów krajowych, empiryczna statystyka Studenta wynosi 3,2 i jest większa od wartości krytycznej. Zatem ceny samochodów krajowych wpływają w sposób istotny na popyt na samochody krajowe.
- Dla cen samochodów importowanych moduł empirycznej statystyki Studenta (1,9) jest mniejszy od wartości krytycznej 2,12. Dlatego uznajemy, że dla tych danych ceny samochodów importowanych są nieistotne.

Wniosek końcowy: model - mimo bardzo dobrego dopasowania - nie może być uznany za końcowy, gdyż zawiera nieistotną zmienną objaśniającą (CSI). Dalsze działania polegałyby na wykluczeniu tej zmiennej i powtórnym oszacowaniu modelu, tym razem jako zależności popytu od dochodów i cen krajowych. Z uwagi na zmianę listy zmiennych objaśniających "nowe" parametry przy dochodzie i cenach krajowych na ogół nie bedą takie, jak poprzednio.

Wśród modeli nieliniowych wyróżnia się modele:

- liniowe względem parametrów,
- linearyzowalne,
- nieliniowe w ścisłym sensie (pozostałe).

W tej chwili zajmiemy się pierwszymi z nich.

Model $\hat{Y} = f(X, b)$ jest liniowy względem parametrów, jeśli można go przedstawić jako liniową funkcję jednoznacznych przekształceń zmiennych objaśniających X, przy czym współczynniki tych przekształceń są znane z dokładnością liczbową.

Model jest więc liniowy względem parametrów, gdy ma formę

$$\hat{Y} = \sum_k b_k Z_k, \tag{12}$$

gdzie

$$Z_k = h_k(X), \tag{13}$$

przy tym wszystkie przekształcenia h_k są jednoznaczne, a ich współczynniki są znane. Zmienne Z_k, będące przekształceniami oryginalnych zmiennych objaśniających X, nazywać będziemy pomocniczymi zmiennymi objaśniającymi, a sam model (12) - pomocniczym modelem liniowym.

Wyznaczanie i weryfikacja modelu liniowego względem parametrów sprowadza się do wyznaczenia i weryfikacji pomocniczego modelu liniowego (12), w którym:

- zmienną objaśniającą jest zmienna oryginalna Y,
- zmiennymi objaśniającymi są zmienne pomocnicze Z_k.

Uzyskany z klasycznej MNK wektor parametrów wyraża się zatem wzorem:

$$b = (Z'Z)^{-1}Z'y. \tag{14}$$

PARABOLA

Przebieg paraboli zależy od znaku współczynnika stojącego przy kwadracie zmiennej objaśniającej b_2. Parabola nadaje się do opisu przebiegów:

- z jednym minimum, najpierw malejących coraz wolniej, a potem rosnących coraz szybciej ($b_2 > 0$);
- z jednym maksimum, najpierw rosnących coraz wolniej, a potem malejących coraz szybciej ($b_2 < 0$).

Model opisuje równanie:

$$Y = b_0 + b_1 X + b_2 X^2 \quad (b_2 \neq 0). \tag{15}$$

Kolejne potęgi zmiennej X, czyli X oraz X^2, traktujemy jako pomocnicze zmienne objaśniające i w odniesieniu do modelu pomocniczego stosujemy metody estymacji i weryfikacji modeli liniowych.

Zamiast więc modelu oryginalnego

$$Y = b_0 + b_1 X + b_2 X^2$$

szacujemy i weryfikujemy model pomocniczy

$$Y = b_0 + b_1 Z_1 + b_2 Z_2, \tag{16}$$

gdzie

$$Z_1 = X, \quad Z_2 = X^2. \tag{17}$$

WIELOMIAN K-TEGO STOPNIA

Model opisuje równanie:

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_k X_k \quad (b_k \neq 0). \tag{18}$$

Kolejne potęgi zmiennej X traktujemy jako kolejne pomocnicze zmienne objaśniające i szacujemy oraz weryfikujemy model liniowy względem tych pomocniczych zmiennych objaśniających. Zamiast więc modelu oryginalnego

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_k X^k$$

szacujemy i weryfikujemy model pomocniczy

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots b_k Z_k, \tag{19}$$

gdzie

$$Z_k = X^k, \quad k = 1, \ldots, K. \tag{20}$$

HIPERBOLA

Model opisuje równanie:

$$Y = b_0 + \frac{b_1}{X} \quad (X > 0;\ \ b_1 \neq 0). \tag{21}$$

Własności hiperboli:

- hiperbola ma asymptotę poziomą $Y = b_0$;
- przebieg zależy od znaku współczynnika kierunkowego b_1 - jeśli jest on dodatni, to hiperbola jest malejąca, a jeśli jest ujemny, to jest ona rosnąca;
- dla $X = 0$ hiperbola nie jest oznaczona.

Hiperbola nadaje się do opisu takich zjawisk, które względem $X > 0$:

- rosną coraz wolniej i zbliżają się asymptotycznie do poziomu nasycenia b_0 (w tym przypadku współczynnik kierunkowy $b_1 < 0$);
- maleją coraz wolniej do poziomu dolnego b_0 (w przypadku $b_1 > 0$).

Przypadek ogólny

Hiperbola wielu zmiennych objaśniających:

$$Y = b_0 + \frac{b_1}{X_1} + \frac{b_2}{X_2} + \cdots + \frac{b_k}{X_k} \quad (\text{wszystkie } X_k > 0). \tag{22}$$

Odwrotności kolejnych zmiennych traktujemy jako kolejne pomocnicze zmienne objaśniające i wyznaczamy oraz weryfikujemy odpowiedni pomocniczy model liniowy.
Zamiast modelu (22) rozpatrujemy zatem pomocniczy model liniowy:

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots + b_k Z_k, \tag{23}$$

gdzie

$$Z_1 = \frac{1}{X_1}, \; Z_2 = \frac{1}{X_2}, \; \ldots, \; Z_k = \frac{1}{X_k}. \tag{24}$$

FUNKCJA LOGARYTMICZNA

Funkcja logarytmiczna z jedną zmienną objaśniającą to

$$Y = b_0 + b_1 \log(X), \quad (X > 0, b_1 \neq 0). \tag{25}$$

Przebieg funkcji logarytmicznej zależy od znaku współczynnika kierunkowego b_1. Dla X dodatnich, funkcja logarytmiczna może opisywać dwa przebiegi:

- rosnący nieograniczenie coraz wolniej (gdy współczynnik $b_1 > 0$);
- malejący nieograniczenie coraz wolniej (gdy współczynnik $b_1 < 0$).

Przypadek ogólny

Przypadek ogólny funkcji logarytmicznej to:

$$Y = b_0 + b_1 \log(X_1) + \cdots + b_k \log(X_k), \quad \text{wszystkie } X_k > 0. \quad (26)$$

Logarytmy kolejnych zmiennych objaśniających traktujemy jako kolejne pomocnicze zmienne objaśniające i szacujemy oraz weryfikujemy odpowiedni pomocniczy model liniowy.
W przypadku modelu (26) modelem pomocniczym jest

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots + b_k Z_k, \quad (27)$$

gdzie

$$Z_1 = \log(X_1), \; Z_2 = \log(X_2), \; \ldots, \; Z_k = \log(X_k). \quad (28)$$

MODELE LINEARYZOWALNE

Model nazywamy linearyzowalnym, jeśli istnieje jednoznaczne przekształcenie obu jego stron takie, że w wyniku otrzymujemy model liniowy lub liniowy względem parametrów.

Model liniowy otrzymany w wyniku linearyzacji nazywać będziemy pomocniczym modelem liniowym, a jego zmienne - zmiennymi pomocniczymi. Jeśli będzie trzeba, pomocniczą zmienną objaśnianą oznaczać będziemy symbolem V, a pomocnicze zmienne objaśniające symbolami $Z_1, Z_2, \ldots, Z_k$.

MODEL WYKŁADNICZY

Model wykładniczy z jedną zmienną objaśniającą to

$$Y = Ap^{bX} \quad (A > 0), \tag{29}$$

gdzie p - stała dodatnia; A, b - parametry. Najczęściej używa się funkcji wykładniczej przy podstawie logarytmu naturalnego ($p = e$) lub przy podstawie logarytmu dziesiętnego ($p = 10$). Parametr A określa teoretyczny poziom zmiennej zależnej Y w okresie bezpośrednio poprzedzającym pierwszy badany okres (czyli dla $X = 0$). Przebieg funkcji wykładniczej zależy od znaku współczynnika kierunkowego (wykładnika) b. Przy $A > 0$ jest to funkcja:

- rosnąca coraz szybciej (i nieograniczenie), jeśli wykładnik $b > 0$;
- malejąca coraz wolniej do zera, jeśli wykładnik $b < 0$.

Przypadek ogólny

Model wykładniczy z wieloma zmiennymi to

$$Y = Ap^{b_1 X_1 + b_2 X_2 + \cdots + b_K X_K} \quad (A, p > 0). \tag{30}$$

Podstawowa własność funkcji wykładniczej - stałe stopy wzrostu

Stopa wzrostu pokazuje, jaki jest oczekiwany względny (relatywny) przyrost Y, gdy dana zmienna objaśniająca wzrasta o jednostkę, a pozostałe zmienne objaśniające nie ulegają zmianie. Tylko funkcje wykładnicze charakteryzują się stałymi stopami wzrostu.

W przypadku funkcji wykładniczej, stopa wzrostu (r_i) względem zmiennej objaśniającej X_i wyznaczana jest przez współczynnik kierunkowy stojący przy tej zmiennej, mianowicie

$$r_i = b_i \cdot \ln(p) \quad (1 \leq i \leq K). \tag{31}$$

W szczególności dla funkcji wykładniczej przy podstawie e jest

$$r_i = b_i. \tag{32}$$

Model potęgowy

$$Y = AX^{\beta} \tag{33}$$

można sprowadzić poprzez zlogarytmowanie obu stron równania do postaci liniowej

$$ln(Y) = ln(A) + \beta ln(X), \tag{34}$$

co możemy zapisać jako

$$Y^* = A^* + \beta X^*.$$

Otrzymaliśmy model regresji liniowej logarytmów z wartości zmiennej Y względem logarytmów z wartości zmiennej X.

Po oszacowaniu MNK parametrów α^* i β oceny parametrów wyjściowego modelu są takie:

$$A = \exp(A^*), \quad \beta = \beta^*.$$

Parametr α określa teoretyczny poziom zmiennej zależnej Y przy jednostkowej wartości zmiennej niezależnej X, natomiast parametr β określa procentową zmianę wartości zmiennej zależnej spowodowaną zmianą wartości zmiennej niezależnej o 1% (właściwość tę nazywa się elastycznością zmiennej zależnej względem zmiennej niezależnej).