Обзор

Обобщенная регрессионная модель

Ошибки спецификации модели

- **не включение в модель существенно влияющего фактора (факторов);**
- **включение в модель несущественно влияющего фактора (факторов);**
- **использование видов зависимостей, не соответствующих истинной форме связи.**

Обозначения

Y – вектор-столбец наблюдений зависимой переменной

X – матрица наблюдений независимых переменных

$\hat{A}$ – вектор-столбец оцениваемых параметров

u – вектор-столбец остатков

Исходя из

$$Y = AX + u$$

и

$$\hat{Y} = X\hat{A}$$

$$\Downarrow$$

$$u = Y - X\hat{A}$$

Сумма квадратов остатков

$$\sum u^2 = u'u = \left(Y - X\hat{A}\right)'\left(Y - X\hat{A}\right) =$$

$$= Y'Y - 2\hat{A}'X'Y + \hat{A}X'X\hat{A}$$

$$\frac{\partial (u' u)}{\partial \hat{A}} = -2X'Y + 2X'X\hat{A} = 0$$

$$\hat{A} = (X'X)^{-1} X'Y$$

Пример

- Необходимо определить, как товарооборот сети магазинов зависит от их торговой площади и среднедневной интенсивности потока покупателей

Магазин	Товарооборот, сотен тыс. грн.	Торговая площадь, тыс. $м^2$	Среднедневная интенсивность потока покупателей, тыс. чел.
1	2,93	0,31	10,24
2	5,27	0,98	7,51
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31

$$X = \begin{pmatrix} 1 & 0{,}31 & 10{,}24 \\ 1 & 0{,}98 & 7{,}51 \\ 1 & 1{,}21 & 10{,}81 \\ 1 & 1{,}29 & 9{,}89 \\ 1 & 1{,}12 & 13{,}72 \\ 1 & 1{,}49 & 13{,}92 \\ 1 & 0{,}78 & 8{,}54 \\ 1 & 0{,}94 & 12{,}36 \\ 1 & 1{,}29 & 12{,}27 \\ 1 & 0{,}48 & 11{,}01 \\ 1 & 0{,}26 & 8{,}25 \\ 1 & 0{,}55 & 9{,}31 \end{pmatrix} \qquad Y = \begin{pmatrix} 2{,}93 \\ 5{,}27 \\ 6{,}85 \\ 7{,}01 \\ 7{,}02 \\ 8{,}35 \\ 4{,}33 \\ 5{,}77 \\ 7{,}68 \\ 3{,}16 \\ 1{,}52 \\ 3{,}15 \end{pmatrix}$$

$$X^{/} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0{,}31 & 0{,}98 & 1{,}21 & 1{,}29 & 1{,}12 & 1{,}49 & 0{,}78 & 0{,}94 & 1{,}29 & 0{,}48 & 0{,}24 & 0{,}55 \\ 10{,}24 & 7{,}51 & 10{,}81 & 9{,}89 & 13{,}72 & 13{,}92 & 8{,}54 & 12{,}36 & 12{,}27 & 11{,}01 & 8{,}25 & 9{,}31 \end{pmatrix}$$

Рассчитаем произведение матриц $X^{/}X$

$$X^{/}X = \begin{pmatrix} 12 & 10,68 & 127,83 \\ 10,68 & 11,4058 & 118,973 \\ 127,83 & 118,973 & 1410,14 \end{pmatrix}$$

Найдем обратную матрицу

$$\left(X^{/}X\right)^{-1}=\begin{pmatrix} 2,47452 & 0,18978 & -0,2403 \\ 0,18978 & 0,74547 & -0,0801 \\ -0,2403 & -0,0801 & 0,02925 \end{pmatrix}$$

Рассчитаем $X^{/}Y$

$$X^{/}Y = \begin{pmatrix} 63{,}04 \\ 66{,}0611 \\ 704{,}692 \end{pmatrix}$$

Найдем $\overset{\square}{A}$

$$\overset{\square}{A} = \begin{pmatrix} -0,8319 \\ 4,74295 \\ 0,17499 \end{pmatrix}$$

Регрессионная модель

$$\hat{y} = -0,8319 + 4,74295x_1 + 0,17499x_2$$

Интерпретация модели

- при изменении размера торговой площади на одну тыс. м^2, при прочих равных условиях, товарооборот увеличится на 4,74 сотен тыс. грн.

- при увеличении среднедневного потока покупателей на одну тыс. чел., при прочих равных условиях, товарооборот возрастет на 0,175 сотен тыс. грн.

- **Свойства оценок параметров**

1. **несмещенность;**
2. **обоснованность;**
3. **эффективность;**
4. **инвариантность.**

- *Оценки параметров модели будут несмещенными, если математическое ожидание их выборочных значений, найденных при многократном повторении выборки, не отличается от истинного значения*

$$M\left(\hat{A}\right) = A$$

1

$$Y = XA + u$$

$$\left(X'X\right)^{-1}X'$$

$$\left(X'X\right)^{-1}X'Y = \left(X'X\right)^{-1}X'XA + \left(X'X\right)^{-1}X'u$$

Так как, $\left(X'X\right)^{-1}X'X = E$ то

$$M\left[\left(X'X\right)^{-1}X'Y\right] = M\left[A + \left(X'X\right)^{-1}X'u\right]$$

$$M\left[\hat{A}\right] = A + \left(X'X\right)^{-1}X'M(u)$$

Учитывая

$$M(u) = 0$$

$$M(\tilde{A}) = A$$

Смещение

$$Q = M\left(\overset{\square}{A}\right) - A$$

Проверка смещенности

$$\frac{\overset{\square}{\sigma}_{a_i}}{\left|\overset{\square}{a}_i\right|} > 10\% \quad ????$$

- *Выборочная оценка* $\overset{\sqcup}{A}$ *называется обоснованной, если для любого сколь угодно малого числа* $\varepsilon > 0$ *выполняется соотношение*

$$\lim_{n\to\infty} P\left\{ \left| \overset{\sqcup}{A} - A \right| < \varepsilon \right\} = 1$$

- Выборочные оценки вектора $\breve{A}$ будут только тогда эффективными, когда их дисперсии будут наименьшими

3

Теорема Гаусса-Маркова

- *функция оценивания по методу 1МНК покомпонентно минимизирует дисперсию всех линейных несмещенных функций вектора оценок* $\hat{A}$

$$\sigma^2_{\hat{A}} \le \sigma^2_{\bar{A}}$$

$\sigma^2_{\breve{A}}$ – дисперсия оценок $\breve{A}$,
определенная методом 1МНК,

$\sigma^2_{\bar{A}}$ – дисперсия оценок $\bar{A}$,
определенных другими способами

!!!

- Функция оценивания 1МНК в классической линейной модели является лучшей линейно несмещенной функцией оценивания

  (с англ. ***BLUE*** *– Best Linear Unbiased Estimator*)

- *Оценка* $\hat{A}$ *параметров* A *называется инвариантной, если для произвольно заданной функции* g *оценка параметров функции* $g(A)$ *представляется в виде*

$$g(\hat{A})$$

4

Коэффициент множественной детерминации

Коэффициент множественной корреляции

Выборочная дисперсия

Среднеквадратическое отклонение

- **Коэффициенты корреляции, детерминации, эластичности**

$$r_{y/x_j} = \frac{\mathrm{cov}(x_j, y)}{\sqrt{\mathrm{var}(x_j)\mathrm{var}(y)}}$$

$$r_{x_j/x_l} = \frac{cov(x_j, x_l)}{\sqrt{var(x_j)var(x_l)}}$$

1. Нормировка

$$y_i^* = \frac{y_i - \bar{y}}{\sigma_y}$$

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}$$

2. Расчет коэффициентов парной корреляции

$$r_{y/x_j} = \frac{1}{n}\left(Y^*\right)^{/} X_j^*$$

$$r_{x_l/x_j} = \frac{1}{n}\left(X_l^*\right)^{/} X_j^*$$

Средние и СКО

$$\overline{y} = 5{,}25 \qquad \overline{x_1} = 0{,}89 \qquad \overline{x_2} = 10{,}65$$

$$\sigma_y = 2{,}11 \qquad \sigma_{x_1} = 0{,}4 \qquad \sigma_{x_2} = 2{,}01$$

Нормировка

$$y_1^* = \frac{y_1 - \overline{y}}{\sigma_y} = \frac{2{,}95 - 5{,}25}{2{,}11} = -1{,}1$$

$$y_2^* = \frac{y_2 - \overline{y}}{\sigma_y} = \frac{5{,}27 - 5{,}25}{2{,}11} = 0{,}008$$

и так далее

Нормированные значения переменных

Наблюдение	y^*	x_1^*	x_2^*
1	−1,100378	−1,457381	−0,205323
2	0,007894	0,226145	−1,564188
3	0,756214	0,804073	0,078396
4	0,831994	1,005091	−0,379536
5	0,836730	0,577927	1,526856
6	1,466645	1,507636	1,626407
7	−0,437309	−0,276400	−1,051502
8	0,244704	0,125636	0,849913
9	1,149319	1,005091	0,805115
10	−0,991445	−1,030218	0,177947
11	−1,768183	−1,633272	−1,195851
12	−0,996182	−0,854327	−0,668233

Коэффициенты парной корреляции

Матрица коэффициентов парной корреляции

$$r = \begin{pmatrix} 1 & 0{,}9843 & 0{,}65141 \\ 0{,}9843 & 1 & 0{,}5424 \\ 0{,}65141 & 0{,}5424 & 1 \end{pmatrix}$$

Коэффициент множественной детерминации

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$

или

Скорректированный коэффициент множественной детерминации (с учетом числа степеней свободы)

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-k}$$

Коэффициент множественной корреляции

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

$$R \geq 0$$

Магазин	y	$y-\bar{y}$	$(y-\bar{y})^2$	$\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$	$(y-\bar{y})(\hat{y}-\bar{y})$	$y-\hat{y}$	$(y-\hat{y})^2$
1	2,93	-2,32	5,40	2,4302	-2,82	7,97	6,55896	0,50	0,2498
2	5,27	0,02	0,00	5,1303	-0,12	0,02	-0,00205	0,14	0,0195
3	6,85	1,60	2,55	6,7986	1,55	2,39	2,46734	0,05	0,0026
4	7,01	1,76	3,09	7,017	1,76	3,11	3,09833	-0,01	0,0001
5	7,02	1,77	3,12	6,8809	1,63	2,65	2,87553	0,14	0,0193
6	8,35	3,10	9,59	8,6709	3,42	11,7	10,5830	-0,32	0,1030
7	4,33	-0,92	0,85	4,3619	-0,89	0,79	0,82303	-0,03	0,0010
8	5,77	0,52	0,27	5,7893	0,54	0,29	0,27690	-0,02	0,0004
9	7,68	2,43	5,89	7,4336	2,18	4,75	5,29068	0,25	0,0607
10	3,16	-2,09	4,38	3,3713	-1,88	3,54	3,93974	-0,21	0,0446
11	1,52	-3,73	13,94	1,750	-3,50	12,3	13,0790	-0,23	0,0529
12	3,15	-2,10	4,42	3,4058	-1,85	3,41	3,88594	-0,26	0,0654
Сумма			53,50			52,9	52,8765		0,6194

$$R^2 = 0,9884$$

$$\overline{R}^2 = 0,98585$$

$$R = \sqrt{0,98842} = 0,99419314$$

Вариация включенных в модель факторов на 98,585% объясняет вариацию зависимой переменной

Коэффициент эластичности

Средние

$$\bar{y} = 5{,}25333 \qquad \bar{x}_1 = 0{,}89$$

$$\bar{x}_2 = 10{,}6525$$

Коэффициенты эластичности

Увеличение торговой площади на 1%, при прочих равных условиях, повлечет за собой рост товарооборота на 0,80353%, а увеличение среднедневной интенсивности потока покупателей также на 1%, при прочих равных условиях, повлечет рост товарооборота на 0,35483%.