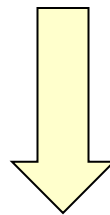


- **Обзор (продолжение)**

Обобщенная регрессионная модель

- **Дисперсионно-ковариационная матрица** (используется для оценки значимости параметров модели)



Оценка дисперсии остатков

$$s_u^2 = \frac{u'u}{n-k} = \frac{Y'Y - A'X'Y}{n-k}$$

$n - k$ — число степеней свободы

$$\hat{Y} = \begin{pmatrix} 2,43 \\ 5,13 \\ 6,799 \\ 7,017 \\ 6,881 \\ 8,671 \\ 4,362 \\ 5,789 \\ 7,434 \\ 3,372 \\ 1,75 \\ 3,406 \end{pmatrix}$$

$$u = \begin{pmatrix} 0,5 \\ 0,14 \\ 0,051 \\ -0,007 \\ 0,139 \\ -0,321 \\ -0,032 \\ -0,019 \\ 0,246 \\ -0,211 \\ -0,23 \\ -0,256 \end{pmatrix}$$

$$Y'Y = 384,666$$

$$\sum A'X'Y = 384,04664$$

$$Y'Y - \sum A'X'Y = 0,61936$$

$$\sigma_u^2 = 0,06882$$

$$\text{Var}(\hat{A}) = \hat{\sigma}_u^2 (X'X)^{-1} = 0,06882 \cdot \begin{pmatrix} 2,47452 & 0,18978 & -0,2403 \\ 0,18978 & 0,74547 & -0,0801 \\ -0,2403 & -0,0801 & 0,02925 \end{pmatrix} =$$

$$= \begin{pmatrix} 0,17029 & 0,01306 & -0,0165 \\ 0,01306 & 0,0513 & -0,0055 \\ -0,0165 & -0,0055 & 0,00201 \end{pmatrix}$$

$$\sigma_{a_0}^k = 0,17029 \quad \Rightarrow \quad \sigma_{a_0}^{\perp k} = 0,41266$$

$$\sigma_{a_1}^k = 0,0513 \quad \Rightarrow \quad \sigma_{a_1}^{\perp k} = 0,2265$$

$$\sigma_{a_2}^k = 0,00201 \quad \Rightarrow \quad \sigma_{a_2}^{\perp k} = 0,04487$$

Определим наличие смещенности оценок

$$\frac{\sigma_{a_0}}{|a_0|} = \frac{0,41233}{|-0,8319|} = 0,496$$

$$\frac{\sigma_{a_1}}{|a_1|} = \frac{0,2265}{|4,74295|} = 0,048$$

$$\frac{\sigma_{a_2}}{|a_2|} = \frac{0,04487}{|0,17499|} = 0,256$$

Вывод:

оценки \hat{a}_0 и \hat{a}_2 являются смещенными

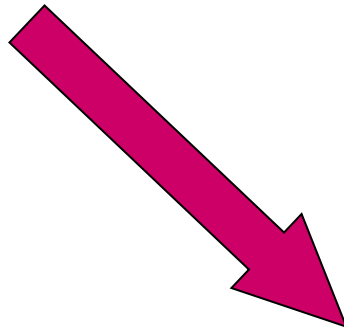
- Проверка значимости модели и ее параметров

1. Проверка значимости модели по критерию Фишера

$$F_{расч.} = \frac{MSR}{MSE}$$

$$MSR = \frac{R^2}{k-1}$$

$$MSE = \frac{1-R^2}{n-k}$$



$$F_{расч.} = \frac{R^2(n-k)}{(1-R^2)(k-1)}$$

Если

$$F_{\text{расч.}} > F_{(k-1, n-k, \alpha)}$$

для уровня значимости α

и числа степеней свободы

$$v_1 = k - 1 \quad \text{и} \quad v_2 = n - k,$$

то уравнение считается значимым

$$F_{\text{расч.}} = \frac{R^2(n-k)}{(1-R^2)(k-1)} = \frac{0,98842(12-3)}{(1-0,98842)(3-1)} = 384,1781$$

$$F_{(2,9,0.05)} = 4,26$$

Так как

$$384,1781 > 4,26,$$

то уравнение считается значимым

- **Стандартизированные
коэффициенты регрессии
(β -коэффициенты),
коэффициенты частной
детерминации**

Определение

- β -коэффициенты показывают влияние независимых переменных на зависимую в относительных единицах измерения

Стандартизированное уравнение регрессии

$$\beta_1 = 4,74295 \frac{0,39797}{2,1114} = 0,8939$$

$$\beta_2 = 0,17449 \frac{2,00903}{2,1114} = 0,166$$

Стандартизированное уравнение
множественной регрессии

$$\hat{y} = 0,89399x_1 + 0,1665x_2$$

Определение

- **Коэффициент частной детерминации**

ΔR_j^2 показывает предельный (граничный) вклад j -го регрессора (независимой переменной) в (общую вариацию) R^2

ИЛИ

- на какую величину уменьшится коэффициент множественной детерминации, если j -ю переменную исключить из модели

$$\Delta R_1^2 = 0,9840,894 \neq 0,8799$$

$$\Delta R_2^2 = 0,6510,1665 \neq 0,108$$

$$R^2 = \Delta R_1^2 + \Delta R_2^2 = 0,87996 \neq 0,1084 \neq 0,988$$

Интерпретация:

88% процентов общей вариации переменной y объясняется вариацией показателя торговой площади (x_1) и почти 11% - вариацией среднегодневного потока покупателей (x_2) .

- **Проверка значимости**

параметров модели $a_0, a_1, a_2, \dots, a_m$

и расчет доверительных

интервалов

$$t_{a_0}^{\square} = \frac{|\square_{a_0}|}{\square_{a_0}}$$

$$t_{a_1}^{\square} = \frac{|\square_{a_1}|}{\square_{a_1}}$$

$$t_{a_2}^{\square} = \frac{|\square_{a_2}|}{\square_{a_2}}$$



$$t_{a_m}^{\square} = \frac{|\square_{a_m}|}{\square_{a_m}}$$

$$a_0^* = \bar{a}_0 \pm t_{\alpha/2} \cdot \sigma_{a_0}$$

$$a_1^* = \bar{a}_1 \pm t_{\alpha/2} \cdot \sigma_{a_1}$$

$$a_2^* = \bar{a}_2 \pm t_{\alpha/2} \cdot \sigma_{a_2}$$



$$a_m^* = \bar{a}_m \pm t_{\alpha/2} \cdot \sigma_{a_m}$$

Доверительные интервалы

Оценка значимости

$$t_{a_0} = \frac{|a_0|}{o_{a_0}} = \frac{|-0,831946|}{0,412663} = 2,016$$

$$t_{a_1} = \frac{|a_1|}{o_{a_1}} = \frac{|4,742948|}{0,226498} = 20,94$$

$$t_{a_2} = \frac{|a_2|}{o_{a_2}} = \frac{|0,174988|}{0,044868} = 3,9$$

$$\alpha = 0,05$$

$$n - k = 12 - 3 = 9$$

$$t_{(\alpha, n-k)} = 2,262$$

$$\alpha = 0,1$$

$$t_{(\alpha, n-k)} = 1,833$$

Доверительные интервалы

$$a_0^* = \bar{a}_0 \pm t_{\alpha/2} \cdot \sigma_{a_0} =$$

$$= -0,8319 \pm 2,262 \cdot 0,413 = -0,8319 \pm 0,933$$

$$-1,765 \leq a_0^* \leq 0,101$$

$$a_1^* = \bar{a}_1 \pm t_{\alpha/2} \cdot \sigma_{a_1} =$$

$$= 4,7429 \pm 2,262 \cdot 0,2265 = 4,7439 \pm 0,512$$

$$4,23 \leq a_1^* \leq 5,25$$

$$a_2^* = \bar{a}_2 \pm t_{\alpha/2} \cdot \sigma_{a_2} =$$

$$= 0,175 \pm 2,262 \cdot 0,045 = 0,175 \pm 0,1018$$

$$0,0735 \leq a_2^* \leq 0,2768$$

- Прогноз значений зависимой переменной y и построение доверительных интервалов прогноза

Точечный прогноз

Оценка дисперсии прогноза

$$\sigma_{n+1}^2 = \sigma_u^2 \cdot X'_{n+1} (X'X)^{-1} X_{n+1}$$

Оценка среднеквадратического отклонения прогноза

$$\sigma_{n+1} = \sigma_u \cdot \sqrt{X'_{n+1} (X'X)^{-1} X_{n+1}}$$

Доверительные интервалы точечного прогноза

$$\bar{Y}_{n+1} - t_{\alpha/2} \sigma_{n+1} \leq M(Y_{n+1}^*) \leq \bar{Y}_{n+1} + t_{\alpha/2} \sigma_{n+1}$$

Интервальный прогноз

Оценка среднеквадратического отклонения прогноза

$$\sigma_{n+1(i)} = \sigma_u \cdot \sqrt{\mathbf{1} + X'_{n+1} (X' X)^{-1} X_{n+1}}$$

Доверительные интервалы интервального прогноза

$$\hat{Y}_{n+1} - t_{\alpha/2} \sigma_{n+1(i)} \leq Y_{n+1}^* \leq \hat{Y}_{n+1} + t_{\alpha/2} \sigma_{n+1(i)}$$

Пусть

$$X_{n+1} = X_{13} = \begin{pmatrix} 1 \\ 1,2 \\ 9 \end{pmatrix}$$

Тогда

$$y_{13} = -0,832 + 4,743 \cdot 1,2 + 0,175 \cdot 9 = 6,434$$

$$\sigma_u = \sqrt{\sigma_u^2} = \sqrt{0,0688} = 0,262$$

$$\sigma_{13} = 0,262 \cdot \sqrt{(1 \ 1,2 \ 9) \cdot \begin{pmatrix} 2,4745 & 0,1898 & -0,24 \\ 0,1898 & 0,7455 & -0,08 \\ -0,24 & -0,08 & 0,029 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1,2 \\ 9 \end{pmatrix}} =$$

$$= 0,262 \cdot \sqrt{0,317} = 0,262 \cdot 0,563 = 0,148$$

$$6,434 - 2,262 \cdot 0,148 \leq M(y_{13}^*) \leq 6,434 + 2,262 \cdot 0,148$$

$$6,1 \leq M(y_{13}^*) \leq 6,769$$

$$\sigma_{13(i)} = 0,262 \cdot \sqrt{1 + (1 \ 1,2 \ 9) \cdot \begin{pmatrix} 2,475 & 0,1898 & -0,24 \\ 0,1898 & 0,7455 & -0,08 \\ -0,24 & -0,08 & 0,029 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1,2 \\ 9 \end{pmatrix}} =$$

$$= 0,262 \cdot \sqrt{1 + 0,3169} = 0,262 \cdot 1,1476 = 0,301$$

$$6,434 - 2,262,0,301 \leq y_{13}^* \leq 6,434 + 2,262,0,301$$

$$5,754 \leq y_{13}^* \leq 7,115$$

- **Процедура многошагового регрессионного анализа**

Направления

1. по мере добавления в модель независимых переменных;
2. по мере исключения из модели многофакторной регрессии несущественно влияющих независимых переменных.

Шаги (первое направление)

- Расчет коэффициентов парной корреляции r_{y/x_j}
- Выбор среди рассчитанных коэффициентов наибольшего (по абсолютной величине). Включение в модель соответствующего показателя

- Построение модели парной регрессии, оценка значимости ее параметров
- Последовательное дополнение модели независимыми переменными по мере уменьшения значений r_{y/x_j} .
Построение моделей, оценка их значимости

Шаги (второе направление)

- Построение модели множественной регрессии с включением в нее всего набора независимых переменных;
- Оценка значимости параметров модели по критерию Стьюдента. Исключение из модели наименее значимой переменной;
- Пересчет параметров модели для оставшегося набора независимых переменных. Оценка значимости параметров и т.д.