- **Обзор (продолжение)**

  Обобщенная регрессионная модель

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

Оценка дисперсии остатков

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

$n = k$ число степеней свободы

$$\hat{Y} = \begin{pmatrix} 2{,}43 \\ 5{,}13 \\ 6{,}799 \\ 7{,}017 \\ 6{,}881 \\ 8{,}671 \\ 4{,}362 \\ 5{,}789 \\ 7{,}434 \\ 3{,}372 \\ 1{,}75 \\ 3{,}406 \end{pmatrix} \qquad u = \begin{pmatrix} 0{,}5 \\ 0{,}14 \\ 0{,}051 \\ -0{,}007 \\ 0{,}139 \\ -0{,}321 \\ -0{,}032 \\ -0{,}019 \\ 0{,}246 \\ -0{,}211 \\ -0{,}23 \\ -0{,}256 \end{pmatrix}$$

$$Y^{/}Y = 384{,}666$$

$$\hat{A}^{/}X^{/}Y = 384{,}04664$$

$$Y^{/}Y - \hat{A}^{/}X^{/}Y = 0{,}61936$$

$$\hat{\sigma}_{u}^{2} = 0{,}06882$$

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

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

$$\hat{\sigma}^2_{a_0} = 0,17029 \quad \Longrightarrow \quad \hat{\sigma}_{a_0} = 0,41266$$

$$\hat{\sigma}^2_{a_1} = 0,0513 \quad \Longrightarrow \quad \hat{\sigma}_{a_1} = 0,2265$$

$$\hat{\sigma}^2_{a_2} = 0,00201 \quad \Longrightarrow \quad \hat{\sigma}_{a_2} = 0,04487$$

Определим наличие смещенности оценок

$$\frac{\hat{\sigma}_{a_0}}{|\hat{a}_0|} = \frac{0,41233}{|-0,8319|} = 0,496$$

$$\frac{\hat{\sigma}_{a_1}}{|\hat{a}_1|} = \frac{0,2265}{|4,74295|} = 0,048$$

$$\frac{\hat{\sigma}_{a_2}}{|\hat{a}_2|} = \frac{0,04487}{|0,17499|} = 0,256$$

Вывод:

оценки $\overset{\sqcup}{a}_0$ и $\overset{\sqcup}{a}_2$ 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

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

для уровня значимости α

и числа степеней свободы

$$\nu_1 = k - 1 \quad \text{и} \quad \nu_2 = n - k,$$

то уравнение считается значимым

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

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

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

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

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

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*

$$\Delta R_1^2 = 0,984 \cdot 0,894 = 0,8799$$

$$\Delta R_2^2 = 0,651 \cdot 0,1665 = 0,108$$

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

Интерпретация:

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

- **Проверка значимости параметров модели** $\hat{a}_0, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_m$ **и расчет доверительных интервалов**

$$t_{\hat{a}_0} = \frac{\left|\hat{a}_0\right|}{\hat{\sigma}_{a_0}}$$

$$t_{\hat{a}_1} = \frac{\left|\hat{a}_1\right|}{\hat{\sigma}_{a_1}}$$

$$t_{\hat{a}_2} = \frac{\left|\hat{a}_2\right|}{\hat{\sigma}_{a_2}}$$

$$t_{\hat{a}_m} = \frac{\left|\hat{a}_m\right|}{\hat{\sigma}_{a_m}}$$

Доверительные интервалы

$$a_0^* = \hat{a}_0 \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_0}$$

$$a_1^* = \hat{a}_1 \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_1}$$

$$a_2^* = \hat{a}_2 \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_2}$$

$$a_m^* = \hat{a}_m \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_m}$$

Оценка значимости

$$t_{a_0} = \frac{|\hat{a}_0|}{\sigma_{a_0}} = \frac{|-0{,}831946|}{0{,}412663} = 2{,}016$$

$$t_{a_1} = \frac{|\hat{a}_1|}{\sigma_{a_1}} = \frac{|4{,}742948|}{0{,}226498} = 20{,}94$$

$$t_{a_2} = \frac{|\hat{a}_2|}{\sigma_{a_2}} = \frac{|0{,}174988|}{0{,}044868} = 3{,}9$$

$$\alpha = 0{,}05$$

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

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

$\alpha = 0{,}1$ $\quad$ $t_{(\alpha, n-k)} = 1{,}833$

Доверительные интервалы

$$a_0^* = \hat{a}_0 \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_0} =$$

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

$$-1{,}765 \le a_0^* \le 0{,}10$$

$$a_1^* = \hat{a}_1 \pm t_{\alpha/2} \cdot \hat{\sigma}_{a_1} =$$

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

$$4{,}23 \le a_1^* \le 5{,}255$$

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

$$= 0{,}175 \pm 2{,}262 \cdot 0{,}045 = 0{,}175 \pm 0{,}1015$$

$$0{,}0735 \le a_2^* \le 0{,}2765$$

- **Прогноз значений зависимой переменной** y **и построение доверительных интервалов прогноза**

Точечный прогноз

Оценка дисперсии прогноза

$$\hat{\sigma}^2_{n+1} = \hat{\sigma}^2_u \cdot X'_{n+1}\left(X'X\right)^{-1}X_{n+1}$$

Оценка среднеквадратического отклонения прогноза

$$\hat{\sigma}_{n+1} = \hat{\sigma}_u \cdot \sqrt{X'_{n+1}\left(X'X\right)^{-1}X_{n+1}}$$

Доверительные интервалы точечного прогноза

$$\hat{Y}_{n+1} - t_{\alpha/2}\hat{\sigma}_{n+1} \le M\left(Y^{*}_{n+1}\right) \le \hat{Y}_{n+1} + t_{\alpha/2}\hat{\sigma}_{n+1}$$

Интервальный прогноз

Оценка среднеквадратического отклонения прогноза

$$\hat{\sigma}_{n+1(i)} = \hat{\sigma}_u \cdot \sqrt{1 + X'_{n+1}\left(X'X\right)^{-1} X_{n+1}}$$

Доверительные интервалы интервального прогноза

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

Пусть

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

Тогда

$$\hat{y}_{13} = -0{,}832 + 4{,}743\;1{,}2 + 0{,}175\;9 = 6{,}434$$

$$\hat{\sigma}_{u} = \sqrt{\hat{\sigma}_{u}^{2}} = \sqrt{0{,}0688} = 0{,}262$$

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

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

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

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

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

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

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

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

- **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.