Свойства коэффициентов регрессии и проверка гипотез

1. Случайные составляющие коэффициентов регрессии

- Модель :

$$y = \alpha + \beta x + u$$

- x – неслучайная экзогенная переменная

- Уравнение регрессии :

$$\hat{y} = a + bx$$

- Коэффициенты регрессии – случайные величины
- <u>Теорема</u>

$$b = \frac{Cov(x, y)}{Var(x)} = \beta + \frac{Cov(x, u)}{Var(x)}$$

2. Условия Гаусса-Маркова

(предположения о случайном члене)

- 1. $E(u) = 0$
- 2. $Var(u)$ постоянна для всех наблюдений
- 3. $Cov(u_i, u_j) = 0 \qquad i \neq j$
- 4. $Cov(x_i, u_i) = 0$

или $E(x_i, u_i) = 0$

Дополнительно: u распределено по нормальному закону

Необходимо понимать

- Если условия не выполнимы, вы должны это сознавать
- Если корректирующие действия возможны, то аналитик должен быть в состоянии их выполнить
- Если ситуацию исправить невозможно, вы должны быть способны оценить, насколько серьезно это может влиять на результаты

3. Несмещенность коэффициентов регрессии

- 1. $$E(b) = \beta + E\left\{\frac{Cov(x,u)}{Var(x)}\right\}$$
  - И если x – неслучайная величина, то
  $$E(b) = \beta$$
- 2. $$E(a) = \alpha$$

4.Точность коэффициентов регрессии

- Теоретические дисперсии оценок a и b

$$Var(a) = \frac{\sigma_u^2}{n}\left\{1 + \frac{\bar{x}^2}{Var(x)}\right\}$$

$$Var(b) = \frac{\sigma_u^2}{nVar(x)}$$

- Выводы:
  - a и b прямо пропорциональны дисперсии остаточного члена
  - чем больше число наблюдений, тем меньше дисперсии
  - чем больше дисперсия x, тем меньше дисперсии коэффициентов

Оценки стандартных отклонений коэффициентов

- Термин «стандартная ошибка»

$$m_b = \sqrt{\frac{\sum\left(y-\hat{y}_x\right)^2/(n-2)}{\sum\left(x-\bar{x}\right)^2}} = \sqrt{\frac{S^2}{\sum\left(x-\bar{x}\right)^2}}$$

$$m_a = \sqrt{\frac{\sum\left(y-\hat{y}_x\right)^2}{n-2}\cdot\frac{\sum x^2}{n\cdot\sum\left(x-\bar{x}\right)^2}} = \sqrt{S^2\cdot\frac{\sum x^2}{n\cdot\sum\left(x-\bar{x}\right)^2}}$$

- Применение: 1) проверка существенности коэффициента регрессии
- 2) построение доверительных интервалов

5. Оценка существенности коэффициента регрессии и свободного члена

- Фактическое значение t-критерия сравнивается с
- табличным значением при определенном уровне
- значимости α и числе степеней свободы (n-2)
- $t_b = \dfrac{b}{m_b}, \qquad t_a = \dfrac{b}{m_a}$ - фактические значения
- <u>Теорема</u> $t_b^2 = F$
- <u>Доказательство:</u>

$$t_b^2 = \frac{b^2}{m_b^2} \quad = b^2 / \frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}$$

Продолжение доказательства

$$= \frac{b^2 \sum (x-\bar{x})^2}{\sum (y-\hat{y}_x)^2 /(n-2)} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\frac{\sum (y-\hat{y}_x)^2}{(n-2)}}$$

$$= \frac{D_{факт}}{D_{ост}} = F$$

Дисперсионный анализ результатов регрессии

Источники вариации	Число степеней свободы	Сумма квадратов отклонений	Дисперсия на одну степень свободы	F-отношение	
				фактическое	табличное при α=0,05
Общая	6	15000	--	--	--
Объясненная	1	14735	14735	278	6,61
Остаточная	5	265	53	1	--

$$m_b = \sqrt{\frac{53}{10,857}} = 2,21$$

$$t_b = \frac{36,84}{2,21} = 16,67 > t_{табл} = 2,57$$

Значимость коэффициента корреляции

- Стандартная ошибка коэффициента корреляции

$$m_r = \sqrt{\frac{1-r^2}{n-2}}$$

- Фактическое значение t-критерия Стьюдента

$$t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$$

Теорема $t_r^2 = t_b^2 = F$

- Доказательство:
  - Смотри выражения для t_r и F.

- **Вывод:** проверка гипотез о значимости коэффициента регрессии b и коэффициента корреляции r равносильна проверке гипотезы о существенности линейного уравнения в целом.

6. Доверительные интервалы параметров регрессии

- Доверительный интервал для коэффициента регрессии

$$(b - t_{табл} * m_b, b + t_{табл} * m_b)$$

- Доверительный интервал для свободного члена регрессии

$$(a - t_{табл} * m_a, a + t_{табл} * m_a)$$

- Доверительный интервал для коэффициента корреляции
-

$$(r - t_{табл} * m_r, r + t_{табл} * m_r)$$

7. Интервалы прогноза по линейному уравнению регрессии

- Точечный прогноз y_p дополняется интервальной оценкой прогнозного значения

$$\widehat{y}_x - m_{\widehat{y}_x} \le y^* \le \widehat{y}_x + m_{\widehat{y}_x}$$

- <u>Теорема</u>

$$m_{\widehat{y}_x} = S \cdot \sqrt{\frac{1}{n} + \frac{\left(x_k - \bar{x}\right)^2}{\sum\left(x - \bar{x}\right)^2}}$$

Доказательство.

$$\widehat{y}_x = \bar{y} - b \cdot \bar{x} + b \cdot x = \bar{y} + b \cdot (x - \bar{x})$$

- Тогда

$$m_{\hat{y}_x}^2 = m_{\bar{y}}^2 + m_b^2 (x - \bar{x})^2$$

- т.к. $Var(\bar{y}) = \dfrac{\sigma^2}{n}$,то $m_{\bar{y}}^2 = \dfrac{S^2}{n}$

- и $m_b^2 = \dfrac{S^2}{\sum (x - \bar{x})^2}$ $\Rightarrow$

$$m_{\hat{y}_x}^2 = \frac{S^2}{n} + \frac{S^2}{\sum (x - \bar{x})^2} \cdot (x_k - \bar{x})^2 = S^2 \cdot \left(\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2} \right)$$

Доверительные границы для $\hat{y}_x$:

$$(\widehat{y}_{x_k} - t_\alpha \cdot m_{\widehat{y}_x}, \widehat{y}_{x_k} + t_\alpha \cdot m_{\widehat{y}_x})$$

- Средняя ошибка прогнозируемого индивидуального значения y составит

$$m_{\widehat{y}(x_k)} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{\sum\left(x - \bar{x}\right)^2}}$$