если $r = 0$, то

$$t_в = \frac{r_в}{\sqrt{1-r_в^2}}\sqrt{n-2} \sim t(n-2)$$

Если $r \neq 0$, то

Рассматривается статистика $z = \frac{1}{2}\ln\frac{1+r_в}{1-r_в} = Arth\, r_в$ которая при $n \geq 10$

приближенно распределена $N\left(a_z, \frac{1}{\sqrt{n-3}}\right)$, где $a_z = Arth(r) + \frac{r}{2(n-1)}$

Следовательно, $(z-a_z)\sqrt{n-3} \sim N(0,1)$

Пример 1

Дано: $n_1 = 28,\ r_1 = 0{,}71,\ n_2 = 39, r_2 = 0{,}85\quad \alpha = 0.01$

а) $H_0 : \rho_1 = \rho_2$

$H_1 : \rho_1 \neq \rho_2$

б) Для каких значений r_2 можно считать, что разность незначима?

Решение.

Используем статистику

$$Arth\, r_1 - Arth\, r_2 \sim N\left(a_{z_1} - a_{z_2}, \sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}\right) \sim N\left(0, \sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}\right)$$

$$M\left(Arth\, r_1 - Arth\, r_2\right) = M\left(Arth\, r_1\right) - M\left(Arth\, r_2\right) =$$

$$= Arth(r) + \frac{r}{2(n_1-1)} - Arth(r) + \frac{r}{2(n_2-1)} = \underbrace{\frac{r}{2(n_1-1)} - \frac{r}{2(n_2-1)}}_{\text{очень мало}} = 0$$

$$D\left(Arth\, r_1 - Arth\, r_2\right) = DArth\, r_1 + DArth\, r_2 = \frac{1}{n_1-3} + \frac{1}{n_2-3}$$

$$z_B = \frac{Arth\, r_1 - Arth\, r_2}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}} \sim N(0,1)$$

а) Тогда по определению ошибки первого рода

$$P\left(\underbrace{|z_B| > C}_{z\in S}\middle| H_0\right) = \alpha \Rightarrow P\left(\underbrace{|z_B| < u_{1-\alpha}}_{z\notin S}\right) = 1-\alpha$$

Тогда допустимые значения для статистики имеют вид (не критичсекие)

$$|z_B| < u_{0.995}$$

Находим

$$z_B = \frac{Arth\, r_1 - Arth\, r_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = \frac{Arth\, 0,71 - Arth\, 0,85}{\sqrt{\frac{1}{36} + \frac{1}{25}}} = \frac{30}{\sqrt{61}}(-0.369) = -1.417 .$$

Находим квантиль: $u_{0.995} = 2.576$

Поскольку $|z_B| < u_{0.995}$ гипотеза H_0 принимается

б) гипотеза H_0 принимается, если

$$|z_B| < 2.576 \Rightarrow \frac{|Arth\, 0,71 - Arth\, 0,85|}{\sqrt{\frac{1}{36} + \frac{1}{25}}} < 2.576 \Rightarrow$$

$$\Rightarrow |Arth\, 0,71 - Arth\, 0,85| < 2.576 \cdot \sqrt{\frac{1}{36} + \frac{1}{25}}$$

$$|Arth\, 0.71 - Arth\, r_2| < 2.576 \frac{\sqrt{61}}{30} = 0.671$$

$$0.217 = Arth\, 0.71 - 0.671 < Arth\, r_2 < Arth\, 0.71 + 0.671 = 1.558$$

$$\tanh(0.217) < r_2 < \tanh(1.558)$$

$$0.214 < r_2 < 0.915$$

Регрессия и ее свойства

Определение. Регрессией компоненты η случайного вектора (ξ, η) на компоненту ξ называется условное математическое ожидание $f(\xi) = M(\eta|\xi)$
При этом $\eta - f(\xi) = \varepsilon$ –ошибка

$$\eta = f(\xi) + \varepsilon$$

Свойства функции регрессии.

1. $D\eta = Df(\xi) + D\varepsilon,$ то есть $f(\xi)$ и ε некоррелированы, и если $D\varepsilon = \overline{\sigma}_\eta^2$, то
$$D\eta = Df(\xi) + \overline{\sigma}_\eta^2.$$
2. Регрессия - наилучшее приближение η через ξ в следующем смысле:
$$\min_g M\big(\eta - g(\xi)\big)^2 = M\big(\eta - f(\xi)\big)^2$$

Линейная регрессионная модель (общий случай).

Общий случай линейной регрессионной модели:

$(\vec{X}_1, Y_1), (\vec{X}_2, Y_2), \dots (\vec{X}_n, Y_n)$ - выборка из $(p+1)$ – мерного распределения,

Векторы $\vec{X}_k = \left(X_k^{(1)}, X_k^{(2)}, \dots, X_k^{(p)}\right), k = \overline{1,n}$ – векторы факторов (элементы факторного пространства, или входные переменные)

$Y_k = \sum_{j=0}^{m-1} \beta_j \Psi_j(\vec{X}_k) + \varepsilon_k$- отклик, или выходная переменная, измеряемая с ошибкой ε_k,

$\Psi_j(\vec{X}_k)$ – базисные функции, характеризующие модель

Примеры.

1. Простейшая линейная регрессия: $Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k$,

$p = 1$- единственный фактор X_k,

$m = 2$ – двумерный вектор параметров регрессии $\vec{\beta}^{\mathrm{T}} = (\beta_0, \beta_1)$,

$\Psi_0(X_k) = 1, \Psi_1(X_k) = X_k$ - базисные функции.

2. Простейшая квадратическая регрессия: $Y_k = \beta_0 + \beta_1 X_k + \beta_2 X_k^2 + \varepsilon_k$,

$p = 1$- единственный фактор X_k,

$m = 3$ – трехмерный вектор параметров регрессии $\vec{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, \beta_2)$,

$\Psi_0(X_k) = 1, \Psi_1(X_k) = X_k, \Psi_2(X_k) = X_k^2$ - базисные функции.

3. Простейшая линейная двухфакторная регрессия: $Y_k = \beta_0 + \beta_1 X_k^{(1)} + \beta_2 X_k^{(2)} + \varepsilon_k$,

$p = 2$- двумерный фактор $\vec{X}_k = \left(X_k^{(1)}, X_k^{(2)}\right)$,

$m = 3$ – трехмерный вектор параметров регрессии $\vec{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, \beta_2)$,

$\Psi_0(\vec{X}_k) = 1, \Psi_1(\vec{X}_k) = X_k^{(1)}, \Psi_2(\vec{X}_k) = X_k^{(2)}$ - базисные функции.

Матричная запись: $Y = \mathcal{F}\vec{\beta} + \vec{\varepsilon}$, где

$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix}$- вектор откликов,

$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_{m-1} \end{pmatrix}$ - вектор параметров регрессии,

$\vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}$ - вектор ошибок, удовлетворяющий условиям $(*)$

$\mathcal{F} = \begin{pmatrix} \Psi_0(\vec{X}_1) & \cdots & \Psi_{m-1}(\vec{X}_1) \\ \vdots & \ddots & \vdots \\ \Psi_0(\vec{X}_n) & \cdots & \Psi_{m-1}(\vec{X}_n) \end{pmatrix}$- матрица базисных функций.

Условия $(*)$: 1. $M\varepsilon_k = 0$ для всех $k = \overline{1, n}$.

2. $M\varepsilon_k\varepsilon_j = 0$ для всех $k \neq j$.

3. $D\varepsilon_k = \delta^2$ для всех $k = \overline{1, n}$.

4. Значения входных переменных $\vec{X}_k, k = \overline{1, n}$, считаются неслучайными (измеренными без ошибок), откуда в частности следует, что

$MY = \mathcal{F}\vec{\beta}$-вектор математических ожиданий

$\Sigma_Y = \Sigma_{\vec{\varepsilon}} = \delta^2 E$ – матрица ковариаций, E-единичная матрица $n \times n$

ОНК для общей линейной регрессионной модели

Теорема 1. Если выполнены условия $(*)$и матрица $\mathcal{M} = \mathcal{F}^{\mathrm{T}}\mathcal{F}$ является невырожденной, то вектор $\widehat{\vec{\beta}}(Y) = (\mathcal{F}^{\mathrm{T}}\mathcal{F})^{-1}\mathcal{F}^{\mathrm{T}}Y$ является ОНК параметра $\vec{\beta}$, обладающей свойствами несмещености и оптимальности в классе несмещенных, линейных по Y оценок.

Доказательство*.*

1) Найдем ОНК параметра $\vec{\beta}$, для чего минимизируем сумму квадратов ошибок:

$$\sum_k \left(Y_k - \sum_{j=0}^{m-1} \beta_j \Psi_j(\vec{X}_k) \right)^2 = \left(Y - \mathcal{F}\vec{\beta}\right)^T \left(Y - \mathcal{F}\vec{\beta}\right) = \sum_k {\varepsilon_k}^2 = \vec{\varepsilon}^{\mathrm{T}}\vec{\varepsilon}$$

Дифференцируем по каждому из параметров β_j критические точки (точки минимума):

$$-2 \sum_k \left(Y_k - \sum_{j=0}^{m-1} \beta_j \Psi_j(\vec{X}_k) \right) \Psi_j(\vec{X}_k) = 0, j = \overline{0, m-1}.$$

Эту систему условий можно записать в матричном виде и разрешить относительно $\vec{\beta}$:

$$\mathcal{F}^{\mathrm{T}}\left(Y - \mathcal{F}\vec{\beta}\right) = 0 \leftrightarrow \mathcal{F}^{\mathrm{T}}Y = \mathcal{F}^{\mathrm{T}}\mathcal{F}\vec{\beta} \leftrightarrow \hat{\beta} = \widehat{\vec{\beta}}(Y) = (\mathcal{F}^{\mathrm{T}}\mathcal{F})^{-1}\mathcal{F}^{\mathrm{T}}Y .$$

Для ОНК $\vec{\beta} = \widehat{\vec{\beta}}(Y) = (\mathcal{F}^{\text{т}}\mathcal{F})^{-1}\mathcal{F}^{\text{т}}Y$ ковариационная матрица имеет вид

$$\Sigma_{\widehat{\beta}} = \delta^2(\mathcal{F}^{\text{т}}\mathcal{F})^{-1}$$

Примеры. Выпишем ОНК для простейшей линейнойй регрессии

$$Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k.$$

$$\mathcal{F}^{\text{т}}\mathcal{F} = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum X_K \\ \sum X_K & \sum X_K^2 \end{pmatrix}$$

$$(\mathcal{F}^{\text{т}}\mathcal{F})^{-1} = \frac{1}{n\sum X_K^2 - (\sum X_K)^2} = \begin{pmatrix} \sum X_K^2 & -\sum X_K \\ -\sum X_K & n \end{pmatrix}$$

$$\mathcal{F}^{\text{т}}Y = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_k \\ \sum Y_k X_k \end{pmatrix}.$$

Получаем

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathcal{F}^{\text{т}}\mathcal{F})^{-1}\mathcal{F}^{\text{т}}Y = \frac{1}{n\sum X_K^2 - (\sum X_K)^2} \begin{pmatrix} \sum X_K^2 & -\sum X_K \\ -\sum X_K & n \end{pmatrix} \begin{pmatrix} \sum Y_k \\ \sum Y_k X_k \end{pmatrix} =$$

$$= \frac{1}{n\sum X_K^2 - (\sum X_K)^2} = \begin{pmatrix} \sum Y_k \sum X_K^2 - \sum X_K \sum Y_k X_k \\ n\sum Y_k X_k - \sum X_K \sum Y_k \end{pmatrix},$$

$$\hat{\beta}_1 = \frac{n\sum Y_k X_k - \sum X_K \sum Y_k}{n\sum X_K^2 - (\sum X_K)^2} = \frac{\sum Y_k X_k - n\overline{XY}}{\sum X_K^2 - n(\overline{X})^2} = \frac{\sum(Y_k - \overline{Y})(X_k - \overline{X})}{\sum(X_k - \overline{X})^2},$$

$$\hat{\beta}_0 = \frac{\sum Y_k \sum X_K^2 - \sum X_K \sum Y_k X_k}{n\sum X_K^2 - (\sum X_K)^2} = \frac{\sum Y_k\,(\sum X_K^2 - (\sum X_K)^2) - \sum X_K\,(\sum Y_k X_k - \sum X_K \sum Y_k)}{n\sum X_K^2 - (\sum X_K)^2} = \overline{Y} - \hat{\beta}_1\overline{X}.$$

В модели простейшей линейной регрессии матрица ковариаций имеет вид:

$$\Sigma\left(\widehat{\vec{\beta}}\right) = \frac{\delta^2}{n\sum X_k^2 - (\sum X_k)^2} \begin{pmatrix} \sum X_k^2 & -\sum X_k \\ -\sum X_k & n \end{pmatrix},$$

Или

$$D(\hat{\beta}_1) = \frac{n\delta^2}{n\sum X_k^2 - (\sum X_k)^2} = \frac{\delta^2}{\sum(X_k)^2 - n\overline{X}^2} = \frac{\delta^2}{\sum(X_k - \overline{X})^2},$$

$$D(\hat{\beta}_0) = \frac{\delta^2 \sum X_k^2}{n\sum X_k^2 - (\sum X_k)^2} = \frac{\delta^2 \sum X_k^2}{n\left(\sum (X_k)^2 - n\overline{X}^2\right)} = \frac{\delta^2 \sum X_k^2}{n\sum\left(X_k - \overline{X}\right)^2},$$

$$cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\delta^2 \sum X_k}{n\sum X_k^2 - (\sum X_k)^2} = -\frac{\delta^2 \overline{X}}{\sum (X_k)^2 - n\overline{X}^2} = -\frac{\delta^2 \overline{X}}{\sum\left(X_k - \overline{X}\right)^2}$$

Оценка дисперсии ошибок

Теорема 2. Если выполнены условия $(*)$ и ранг матрицы $\mathcal{F}$ максимален, то несмещенная оценка остаточной дисперсии в модели $Y = \mathcal{F}\vec{\beta} + \vec{\varepsilon}$ имеет вид:

$$\hat{\delta}^2 = \frac{1}{n-m}\left(Y - \mathcal{F}\,\widehat{\vec{\beta}}\right)^{\mathrm{T}}\left(Y - \mathcal{F}\,\widehat{\vec{\beta}}\right) = \frac{1}{n-m}\sum\left(Y_k - \sum_{j=0}^{m-1}\hat{\beta}_j \Psi_j(\vec{X}_k)\right)^2.$$

Закон распределения остаточной суммы квадратов в нормальной модели

Теорема 3. Если выполнены условия $(*)$, $\varepsilon_k \sim \mathcal{N}(0, \delta)$, ранг матрицы $\mathcal{F}$ максимален, то

$$\frac{\hat{\delta}^2(n-m)}{\delta^2} = \frac{1}{\delta^2}\left(Y - \mathcal{F}\,\widehat{\vec{\beta}}\right)^{\mathrm{T}}\left(Y - \mathcal{F}\,\widehat{\vec{\beta}}\right) = \frac{1}{\delta^2}\sum\left(Y_k - \sum_{j=0}^{m-1}\hat{\beta}_j \Psi_j(\vec{X}_k)\right)^2$$

распределена по закону $\chi^2(n-m)$ и не зависит от $\widehat{\vec{\beta}}$.

Доверительный интервал для остаточной дисперсии

$$\chi^2_{\frac{\alpha}{2}}(n-m) < \frac{\hat{\delta}^2(n-m)}{\delta^2} < \chi^2_{1-\frac{\alpha}{2}}(n-m)$$

$$\frac{\hat{\delta}^2(n-m)}{\chi^2_{1-\frac{\alpha}{2}}(n-m)} < \delta^2 < \frac{\hat{\delta}^2(n-m)}{\chi^2_{\frac{\alpha}{2}}(n-m)}$$

$$\hat{\delta}^2(n-m) = \sum\left(Y_k - \sum_{j=0}^{m-1}\hat{\beta}_j \Psi_j(\vec{X}_k)\right)^2$$

В частности **для простейшей линейной регрессии**

$$\frac{\sum\left(Y_k - \widehat{\beta_0} - \widehat{\beta_1}X_k\right)^2}{\chi^2_{1-\frac{\alpha}{2}}(n-2)} < \delta^2 < \frac{\sum\left(Y_k - \widehat{\beta_0} - \widehat{\beta_1}X_k\right)^2}{\chi^2_{\frac{\alpha}{2}}(n-2)}$$

Доверительные интервалы для параметров регрессии при неизвестной δ^2

По теореме 1 $\widehat{\beta_\iota} \sim N\left(\beta_i, \delta\sqrt{\tau_{ii}}\right)$, где τ_{ii} – диагональный элемент матрицы $(\mathcal{F}^{\mathrm{T}}\mathcal{F}\ )^{-1}$

Следовательно, $\frac{\widehat{\beta_\iota}-\beta_i}{\delta\sqrt{\tau_{ii}}} \sim N(0,1)$ и по теореме 3 не зависит от $\hat{\delta}^2$, для которого $\frac{\hat{\delta}^2(n-m)}{\delta^2} \sim \chi^2(n-m)$. Тогда

$$\frac{\widehat{\beta_\iota}-\beta_i}{\hat{\delta}\sqrt{\tau_{ii}}} = \frac{\dfrac{\widehat{\beta_\iota}-\beta_i}{\delta\sqrt{\tau_{ii}}}}{\dfrac{\hat{\delta}}{\delta}} \sim t(n-m)$$

Откуда

$$\hat{\beta}_i - t_{1-\frac{\alpha}{2}}(n-m)\cdot\hat{\delta}\sqrt{\tau_{ii}} < \beta_i < \hat{\beta}_i + t_{1-\frac{\alpha}{2}}(n-m)\cdot\hat{\delta}\sqrt{\tau_{ii}},$$

Элементы регрессионного анализа

Проверка значимости параметров регрессионной модели

Проверяется гипотеза: H_{0i}: $\beta_i = 0$

Против альтернативы: H_{1i}: $\beta_i \neq 0$

Критерий строим на основе статистики $\widehat{\beta_\iota}$, для которой уже получен доверительный интервал. При рассматриваемой альтернативе критическое множество имеет вид:

$$\mathcal{S} = \left\{\left|\widehat{\beta_\iota}\right| > C\right\}$$

При справедливости основной гипотезы $P\left(\left|\widehat{\beta_\iota}\right| > C \middle| \beta_i = 0\right) = \alpha$

Поскольку согласно доказанному

$$\frac{\widehat{\beta_\iota}-\beta_i}{\hat{\delta}\sqrt{\tau_{ii}}} = \frac{\dfrac{\widehat{\beta_\iota}-\beta_i}{\delta\sqrt{\tau_{ii}}}}{\dfrac{\hat{\delta}}{\delta}} \sim t(n-m),$$

где τ_{ii} – диагональный элемент матрицы $(\mathcal{F}^{\mathrm{T}}\mathcal{F}\ )^{-1}$, то

$$P\left(\left|\widehat{\beta_\iota}\right| > C \middle| \beta_i = 0\right) = \alpha$$

$$\alpha = P\left(\left|\frac{\widehat{\beta_\iota}}{\hat{\delta}\sqrt{\tau_{ii}}}\right| > \frac{C}{\hat{\delta}\sqrt{\tau_{ii}}} \middle| \beta_i = 0\right) = 2\left[1 - F_{t(n-m)}\left(\frac{C}{\hat{\delta}\sqrt{\tau_{ii}}}\right)\right]$$

$$F_{t(n-m)}\left(\frac{C}{\hat{\delta}\sqrt{\tau_{ii}}}\right) = 1 - \frac{\alpha}{2}$$

Откуда

$$C = \hat{\delta}\sqrt{\tau_{ii}}\cdot t_{1-\frac{\alpha}{2}}(n-m),$$

Пример 2

Результаты равноточных измерений глубины h проникновения тела в преграду при различных значениях удельной энергии E приведены в таблице.

Подберите линейную зависимость вида $h = \beta_1 E + \beta_0$, оцените дисперсии

i	0	1	2	3	4	5	6	7	8	9	10	11	12
E_i	41	50	81	104	120	139	154	180	208	241	250	269	301
h_i	4	8	10	14	16	20	19	23	26	30	31	36	37

коэффициентов

Выпишем матрицу базисных функций $\psi_0(x) = 1; \psi_1(x) = x$

$$\mathcal{F} = \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix}$$

Найдем оценки параметров регрессии

$$\mathcal{F}^T \mathcal{F} = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum X_k \\ \sum X_k & \sum X_k^2 \end{pmatrix}$$

$$\left(\mathcal{F}^T \mathcal{F}\right)^{-1} = \frac{1}{n\sum X_k^2 - \left(\sum X_k\right)^2} \begin{pmatrix} \sum X_k^2 & -\sum X_k \\ -\sum X_k & n \end{pmatrix}$$

$$\mathcal{F}^T Y = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_k \\ \sum X_k Y_k \end{pmatrix}.$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(\mathcal{F}^T \mathcal{F}\right)^{-1} \mathcal{F}^T Y = \frac{1}{n\sum X_k^2 - \left(\sum X_k\right)^2} \begin{pmatrix} \sum X_k^2 & -\sum X_k \\ -\sum X_k & n \end{pmatrix} \begin{pmatrix} \sum Y_k \\ \sum X_k Y_k \end{pmatrix} =$$

$$= \begin{pmatrix} 0.38931 & -0.0019 \\ -0.0019 & 0.00001 \end{pmatrix} \begin{pmatrix} 274 \\ 55805 \end{pmatrix} = \begin{pmatrix} 0.671882 \\ 0.124072 \end{pmatrix}$$

Найдём дисперсию точности измерений.

Для ОНК $\vec{\beta} = \widehat{\vec{\beta}}(Y) = (\mathcal{F}^\top \mathcal{F})^{-1} \mathcal{F}^\top Y$ ковариационная матрица имеет вид

$$\Sigma_{\hat{\beta}} = \hat{\delta}^2 \left(\mathcal{F}^T \mathcal{F}\right)^{-1}$$

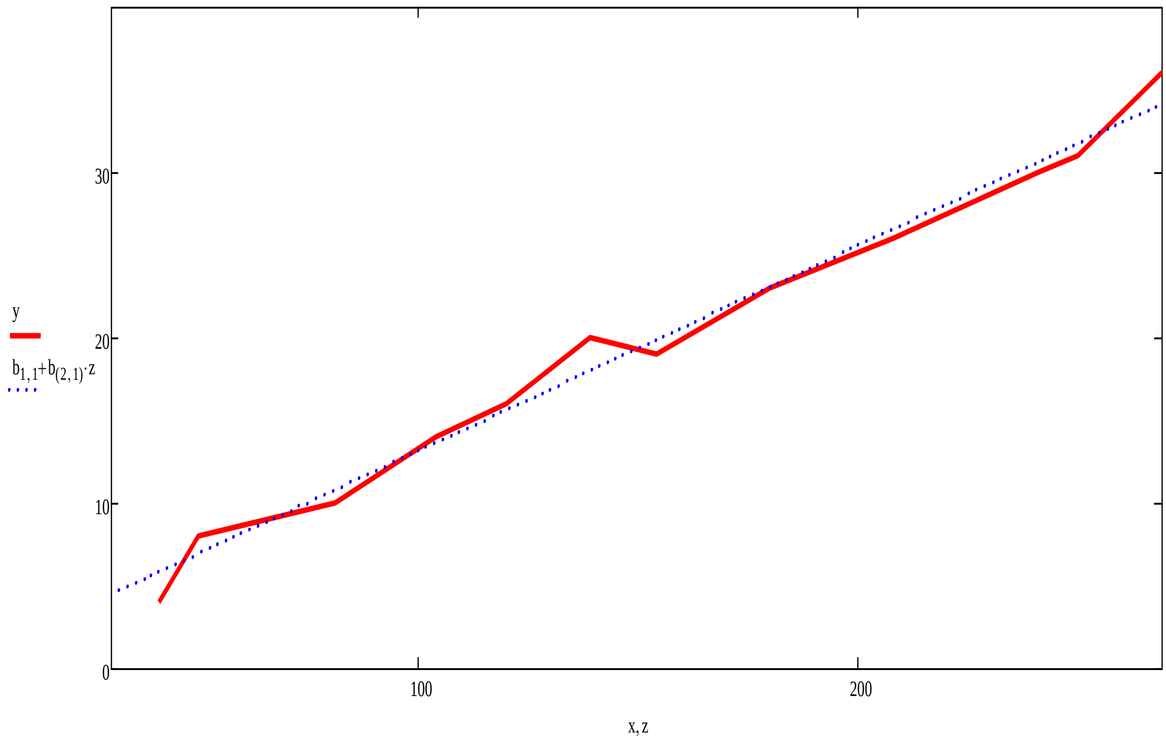
$$\hat{\delta}^2 = \frac{1}{n-m}\left(Y - \mathcal{F}\,\vec{\beta}\right)^T \left(Y - \mathcal{F}\,\vec{\beta}\right) = \frac{1}{n-m}\Sigma\left(Y_k - \sum_{j=0}^{m-1}\hat{\beta}_j \Psi_j\left(\vec{X}_k\right)\right)^2 =$$

$$= \frac{1}{n-2}\sum\left(Y_k - \hat{\beta}_0 - \hat{\beta}_1 X_k\right)^2 = 1.46153$$

$$\Sigma_{\hat{\beta}} = \hat{\delta}^2 \left(\mathcal{F}^T \mathcal{F}\right)^{-1} = \begin{pmatrix} 0.568991 & -0.002776 \\ -0.002776 & 0.000017 \end{pmatrix}$$

$$D\left(\hat{\beta}_1\right) = 0.000017$$

$$D\left(\hat{\beta}_0\right) = 0.568991$$



ПРИМЕР 3

Себестоимость y одного экземпляра книги в рублях в зависимости от тиража x (тыс. экземпляров) представлены в таблице.

x_i	1	2	3	5	10	20	30	50	100	200
y_i	10.15	5.52	4.08	2.85	2.11	1.62	1.41	1.30	1.21	1.15

Постройте на уровне доверия $1-\alpha=0,95$ доверительные интервалы для параметров гиперболической регрессии $y=\frac{a_1}{x}+a_0$

Выпишем матрицу базисных функций $\psi_0\left(x\right)=1;\psi_1\left(x\right)=\frac{1}{x}$

$$\mathcal{F}=\begin{pmatrix} 1 & \frac{1}{X_1} \\ \dots & \dots \\ 1 & \frac{1}{X_n} \end{pmatrix}$$

Найдем оценки параметров регрессии

$$\mathcal{F}^T\mathcal{F}=\begin{pmatrix} 1 & \dots & 1 \\ \frac{1}{X_1} & \dots & \frac{1}{X_n} \end{pmatrix}\begin{pmatrix} 1 & \frac{1}{X_1} \\ \dots & \dots \\ 1 & \frac{1}{X_n} \end{pmatrix}=\begin{pmatrix} n & \sum\frac{1}{X_k} \\ \sum\frac{1}{X_k} & \sum\frac{1}{X_k^2} \end{pmatrix}$$

$$\left(\mathcal{F}^T\mathcal{F}\right)^{-1}=\frac{1}{n\sum\frac{1}{X_k^2}-\left(\sum\frac{1}{X_k}\right)^2}\begin{pmatrix} \sum\frac{1}{X_k^2} & -\sum\frac{1}{X_k} \\ -\sum\frac{1}{X_k} & n \end{pmatrix}$$

$$\mathcal{F}^TY=\begin{pmatrix} 1 & \dots & 1 \\ \frac{1}{X_1} & \dots & \frac{1}{X_n} \end{pmatrix}\begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}=\begin{pmatrix} \sum Y_k \\ \sum\frac{Y_k}{X_1} \end{pmatrix}.$$

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}=\left(\mathcal{F}^T\mathcal{F}\right)^{-1}\mathcal{F}^TY=\frac{1}{n\sum\frac{1}{X_k^2}-\left(\sum\frac{1}{X_k}\right)^2}\begin{pmatrix}\sum\frac{1}{X_k^2} & -\sum\frac{1}{X_k}\\ -\sum\frac{1}{X_k} & n\end{pmatrix}\begin{pmatrix}\sum Y_k\\ \sum\frac{Y_k}{X_1}\end{pmatrix}=$$

$$=\begin{pmatrix}0.156 & -0.248\\ -0.248 & 1.101\end{pmatrix}\begin{pmatrix}31.4\\ 15.223\end{pmatrix}=\begin{pmatrix}1.118856\\ 8.976211\end{pmatrix}$$

Найдём дисперсию точности измерений.

Для ОНК $\vec{\beta}=\widehat{\vec{\beta}}(Y)=(\mathcal{F}^{\tau}\mathcal{F})^{-1}\mathcal{F}^{\tau}Y$ ковариационная матрица имеет вид

$$\Sigma_{\hat{\beta}}=\hat{\delta}^2\left(\mathcal{F}^m\mathcal{F}\right)^{-1}$$

$$\hat{\delta}^2=\frac{1}{n-m}\left(Y-\mathcal{F}\ \vec{\beta}\right)^m\left(Y-\mathcal{F}\ \vec{\beta}\right)=\frac{1}{n-m}\Sigma\left(Y_k-\sum_{j=0}^{m-1}\hat{\beta}_j\Psi_j\left(\vec{X}_k\right)\right)^2=$$

$$=\frac{1}{n-2}\sum\left(Y_k-\hat{\beta}_0-\hat{\beta}_1\frac{1}{X_k}\right)^2=0.00342$$

$$\Sigma_{\hat{\beta}}=\hat{\delta}^2\left(\mathcal{F}^T\mathcal{F}\right)^{-1}=\begin{pmatrix}0.000533 & -0.000849\\ -0.000849 & 0.003769\end{pmatrix}$$

$$D\left(\hat{\beta}_1\right)=0.003769$$

$$D\left(\hat{\beta}_0\right)=0.000533$$

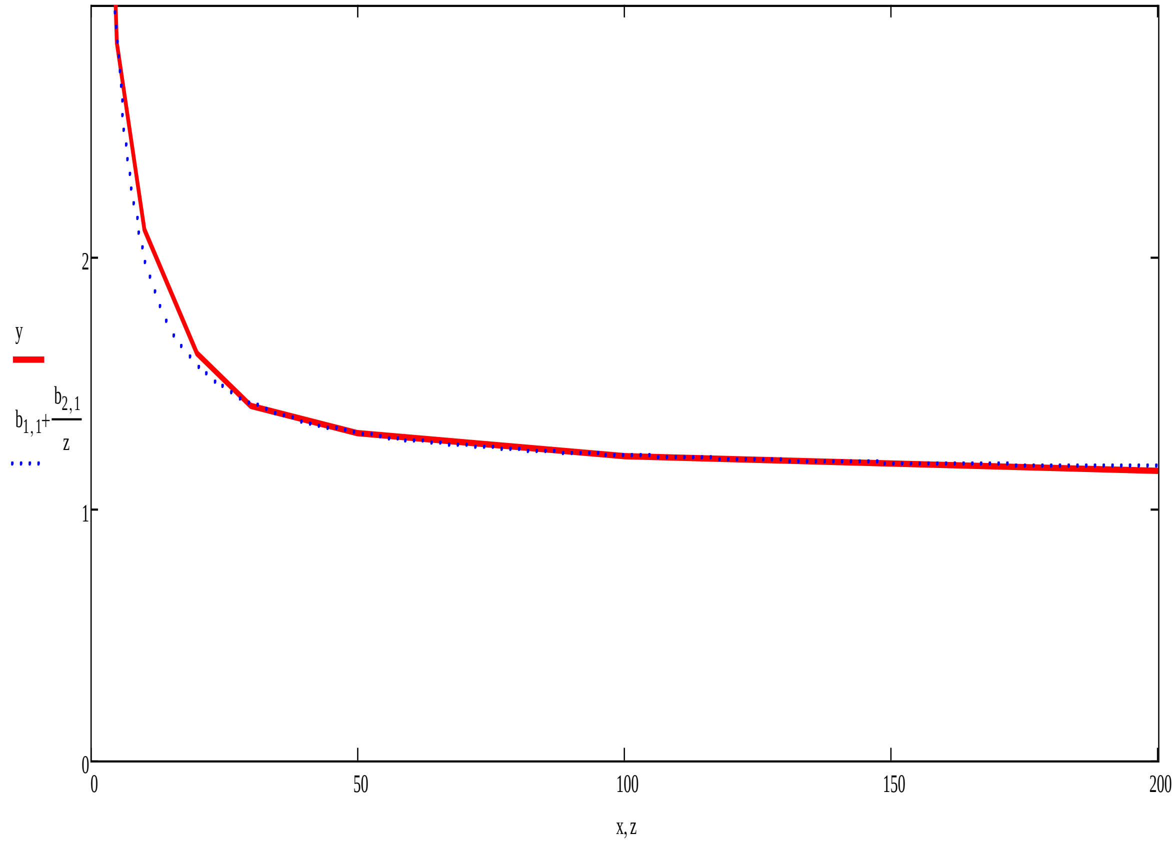
Найдём Доверительные интервалы для коэффициентов регрессии при $1-\alpha=0,95$

$$\hat{\beta}_i-t_{1-\frac{\alpha}{2}}(n-m)\cdot\hat{\delta}\sqrt{\tau_{ii}}<\beta_i<\hat{\beta}_i+t_{1-\frac{\alpha}{2}}(n-m)\cdot\hat{\delta}\sqrt{\tau_{ii}},$$

$$\hat{\beta}_0-t_{0.975}(8)\cdot\sqrt{D\hat{\beta}_0}<\beta_0<\hat{\beta}_i+t_{0.975}(8)\cdot\sqrt{D\hat{\beta}_0},$$

$$\hat{\beta}_1-t_{0.975}(8)\cdot\sqrt{D\hat{\beta}_1}<\beta_1<\hat{\beta}_1+t_{0.975}(8)\cdot\sqrt{D\hat{\beta}_1},$$

$$1,065<\beta_0<1,172,$$
$$8,931<\beta_1<9,115,$$



y
$b_{1,1}+\frac{b_{2,1}}{z}$
2
1
0
0
50
100
150
200
x, z

Пример 5

Найдите коэффициенты регрессии $y=\beta_0+\beta_1 x+\beta_2 x^2$ по наблюденным данным

X_k	0	2	4	6	8	10
Y_k	5	-1	-0.5	1.5	4.5	8.5

$$\mathcal{F}^T\mathcal{F}=\begin{pmatrix}1 & \dots & 1\\ X_1 & \dots & X_n\\ X_1^2 & \dots & X_n^2\end{pmatrix}\begin{pmatrix}1 & X_1 & X_1^2\\ \dots & \dots & \dots\\ 1 & X_n & X_n^2\end{pmatrix}=\begin{pmatrix}n & \sum X_K & \sum X_K^2\\ \sum X_K & \sum X_K^2 & \sum X_K^3\\ \sum X_K^2 & \sum X_K^3 & \sum X_K^4\end{pmatrix},$$

$$\mathcal{F}^T Y=\begin{pmatrix}1 & \dots & 1\\ X_1 & \dots & X_n\\ X_1^2 & \dots & X_n^2\end{pmatrix}\begin{pmatrix}Y_1\\ \dots\\ Y_n\end{pmatrix}=\begin{pmatrix}\sum Y_k\\ \sum Y_k X_k\\ \sum Y_k X_k^2\end{pmatrix},$$

$$\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\beta_2\end{pmatrix}=\left(\mathcal{F}^T\mathcal{F}\right)^{-1}\mathcal{F}^T Y=\begin{pmatrix}n & \sum X_K & \sum X_K^2\\ \sum X_K & \sum X_K^2 & \sum X_K^3\\ \sum X_K^2 & \sum X_K^3 & \sum X_K^4\end{pmatrix}^{-1}\begin{pmatrix}\sum Y_k\\ \sum Y_k X_k\\ \sum Y_k X_k^2\end{pmatrix}=$$

$$=\begin{pmatrix}6 & 30 & 220\\ 30 & 220 & 1800\\ 220 & 1800 & 15664\end{pmatrix}\begin{pmatrix}18\\ 126\\ 1180\end{pmatrix}=\begin{pmatrix}4\\ -2.164286\\ 0.267857\end{pmatrix}$$