МИНОБРНАУКИ РОССИИ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)

Кафедра МОЭВМ

ИДЗ № 3
по дисциплине «Математическая статистика»

Студентка гр. 5381	_________________	Кочнева О.Р.
Преподаватель	_________________	Чирина А.В.

Санкт-Петербург

2017

Постановка задачи:

Вар. 8 (538117)

Результаты статистического эксперимента приведенны в таблице 1. Требуется оценить характер (случайной) зависимости переменной Y от переменной X.

1. Построить графически результаты эксперимента. Сформулировать линейную регрессионную модель переменной Y по переменной X. Построить МНК оценки параметров сдвига β_0 и масштаба β_1. Построить полученную линию регрессии. Оценить визуально соответствие полученных данных и построенной оценки.
2. Построить и интерпретировать несмещенную оценку дисперсии. На базе ошибок построить гистограмму с шагом h. Проверить гипотезу нормальности ошибок на уровне α по χ^2. Оценить расстояние полученной оценки до класса нормальных распределений по Колмогорову. Визуально оценить данный факт.
3. В предположении нормальности ошибок построить доверительные интервалы для параметров β_0 и β_1 уровня доверия $1-\alpha$. Построить доверительный эллипс уровня доверия $1-\alpha$ для (β_0, β_1) (вычислить его полуоси).
4. Сформулировать гипотезу независимости переменной Y от переменной X. Провести проверку значимости.
5. Сформулировать модель, включающую дополнительный член с X^2. Построить МНК оценки параметров $\beta_1, \beta_2, \beta_3$ в данной модели. Изобразить графически полученную регрессионную зависимость.
6. Построить несмещенную оценку дисперсии. Провести исследование нормальности ошибок как в п.3.
7. В предположении нормальности ошибок построить доверительные интервалы для параметров $\beta_1, \beta_2, \beta_3$ уровня $1-\alpha$. Написать уравнение доверительного эллипсоида уровня доверия $1-\alpha$.
8. Сформулировать гипотезу линейной регрессионной зависимости переменной Y от переменной X и проверить ее значимость на уровне α.
9. Интерпретировать полученные результаты. Написать отчет.

Таблица 1 $\alpha_1 = 0.01$; $h = 1.00$.

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Y	7.97	5.76	11.51	6.81	7.14	7.31	7.59	8.29	6.91	7.02	7.41	7.24	7.44	6.20	8.61	8.93	8.00
X	3	4	0	4	0	4	4	0	0	0	0	0	4	4	0	1	1

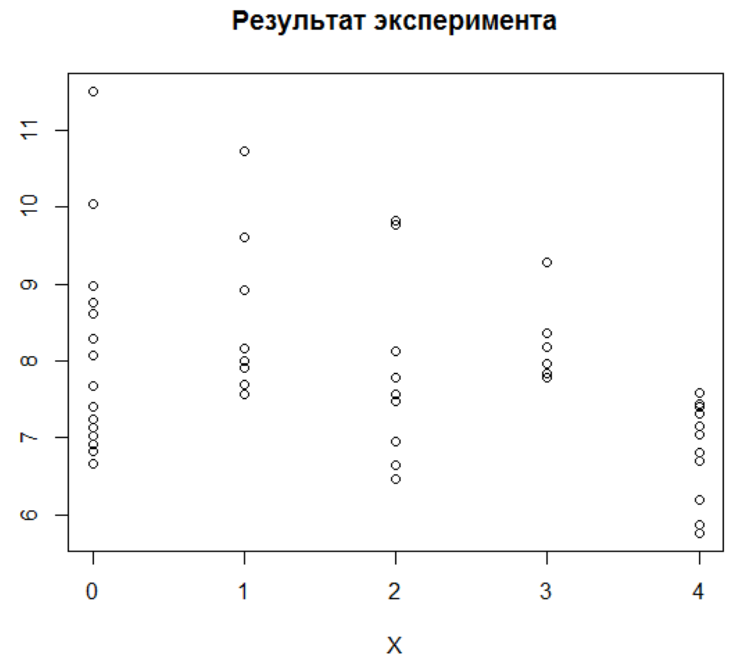
No	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Y	5.86	7.48	7.56	6.64	7.78	8.18	7.91	7.68	7.41	7.57	7.15	9.61	10.73	8.12	8.97	6.83	7.04
X	4	2	1	2	2	3	1	0	4	2	4	1	1	2	0	0	4

No	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Y	7.57	7.79	6.66	8.76	9.83	6.96	8.16	8.36	6.70	6.47	8.07	7.84	10.04	7.69	9.78	9.29
X	1	3	0	0	2	2	1	3	4	2	0	3	0	1	2	3

Ход работы.

1. **Построить графически результаты эксперимента. Сформулировать линейную регрессионную модель переменной Y по переменной X. Построить МНК оценки параметров сдвига β_0 и масштаба β_1. Построить полученную линию регрессии. Оценить визуально соответствие полученных данных и построенной оценки.**

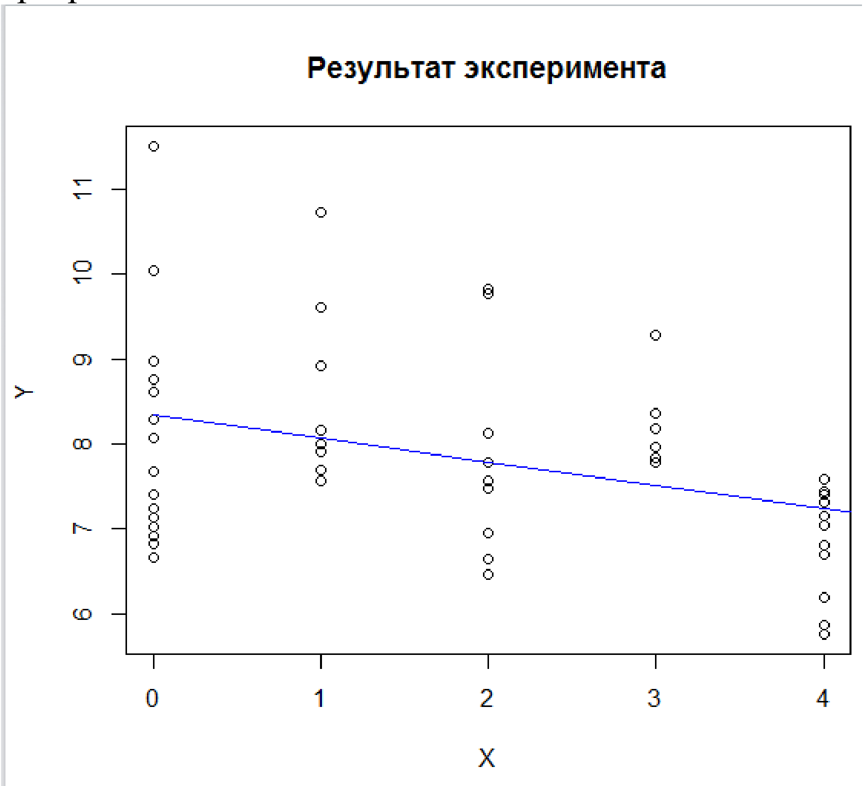
   Графический результат эксперимента:



МНК
β_0=8.3425871
β_1=-0.2752736

Функция линейной регрессии имеет вид: y=8.3425871-0.2752736 *x+E

Линия регрессии.



2. **Построить и интерпретировать несмещенную оценку дисперсии. На базе ошибок построить гистограмму с шагом h. Проверить гипотезу нормальности ошибок на уровне α по χ2. Оценить расстояние полученной оценки до класса нормальных распределений по Колмогорову. Визуально оценить данный факт.**

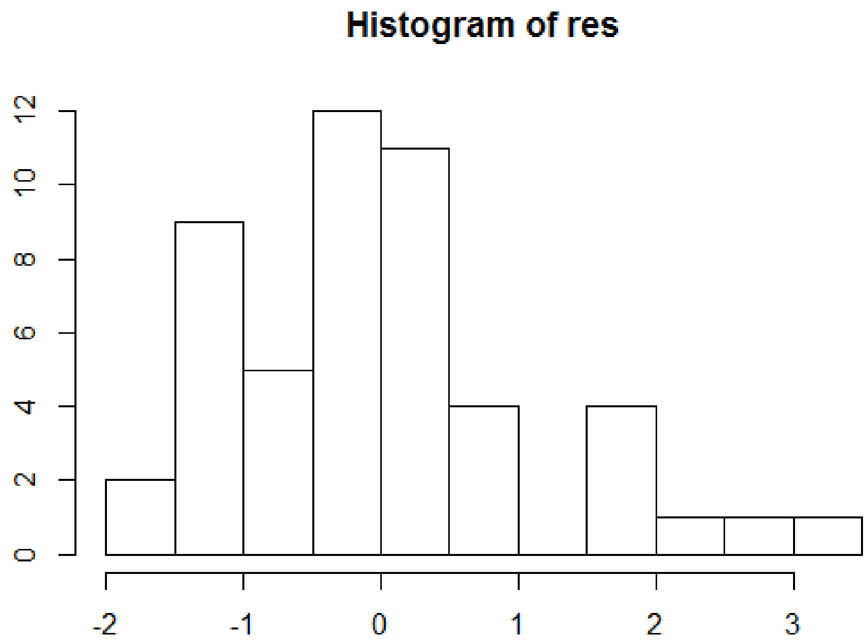
Ошибки $Y - X^T\beta$:

```
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]     [,7]      [,8]
[1,] -1.682587 -1.512587 -1.481493 -1.432587 -1.381493 -1.322587 -1.32204 -1.202587
         [,9]     [,10]     [,11]      [,12]      [,13]      [,14]      [,15]
[1,] -1.15204 -1.102587 -1.041493 -0.9325871 -0.8320398 -0.6625871 -0.5414925
          [,16]      [,17]      [,18]      [,19]      [,20]      [,21]      [,22]
[1,] -0.5073134 -0.4973134 -0.4314925 -0.3773134 -0.3120398 -0.2725871 -0.2220398
          [,23]      [,24]       [,25]       [,26]       [,27]      [,28]      [,29]
[1,] -0.2014925 -0.1573134 -0.09149254 -0.06731343 -0.05258706 -0.0120398 0.06850746
          [,30]     [,31]     [,32]     [,33]     [,34]     [,35]     [,36]     [,37]
[1,] 0.09268657 0.1685075 0.1985075 0.2674129 0.2732338 0.3232338 0.3279602 0.3485075
         [,38]     [,39]     [,40]     [,41]     [,42]     [,43]    [,44]    [,45]
[1,] 0.4174129 0.4532338 0.6274129 0.6632338 0.8432338 0.8626866 1.542687 1.697413
        [,46]   [,47]   [,48]    [,49]    [,50]
[1,] 1.773234 1.98796 2.03796 2.662687 3.167413
```

Несмещенная оценка дисперсии:

$$\hat{\sigma}^2 = \frac{\left\|Y - X^T\hat{\beta}\right\|^2}{n-2} = \frac{\sum(Y_i - \widehat{\beta_0} - \widehat{\beta_1}X_i)^2}{n-2} = 1.252863$$

Гистограмма с шагом h на базе ошибок:



Проверка гипотезы нормальности ошибок на уровне α_1=0.10 по X^2

Основная гипотеза:

$H_0: Y - X^T\tilde{\beta} \sim N(0, \sigma^2)$;

Статистика критерия: $X^2 = \sum_{i=1}^{r} \frac{(v_i - np_i)^2}{np_i} \Rightarrow \chi^2_{r-2}$

$P_{\theta_0}(X^2 < x_\alpha) = 1 - \alpha_1$

	n_k	p_k	np_k	$n_k - np_k$	$\frac{(n_k - np_k)^2}{np_k}$
[-1.68; -1,03]	10	8.936620	446.8310	-436.8310	0.12653300
[-1,03; -0.31]	10	10.608739	530.4370	-520.4370	0.03493001
[-0,31; 0,09]	10	7.056790	352.8395	-342.8395	1.22753885
[0,09; 0,63]	10	9.059337	452.9669	-442.9669	0.09767236
[0,63; 3.17]	10	14.338514	716.9257	-706.9257	1.31273731
x_α = 13.2761		$\sum p_k =$ 50		$X^2 = \sum \frac{(n_k - np_k)^2}{np_k} =$ 2.799412	

$$\boxed{\text{при } r = 5} \quad X^2 = 2.799412$$

$P_{\theta_0}(X^2 < x_\alpha) = 1 - \alpha_1$

$x_\alpha = K_{r-2}^{-1}(1-\alpha_1) = \underline{13.2761}$

Критерий:

$$\Phi(\vec{x}) = \begin{cases} 0, X^2 < x_\alpha \\ 1, X^2 > x_\alpha \end{cases}$$

$X^2 < x_\alpha$ ---2.799412 < 13.2761 - принимаем гипотезу H_0.

Оценка расстояния полученной оценки до класса нормальных распределений по Колмогорову.

Статистика критерия:

$\overline{D_n} = \sqrt{n}\max_x|F_n(x) - F_0(x)| \to K(x)$, где $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\mathrm{I}_{(-\infty,x_i)}(x_i)$ - выборочная функция распределения

$P_{\theta_0}\left(\overline{D_n} < x_\alpha\right) = 1 - \alpha_1$

$x_\alpha = K^{-1}(1-\alpha_1)$

Критерий: $\Phi(\vec{x}) = \begin{cases} 0, \overline{D_n} < x_\alpha \\ 1, \overline{D_n} > x_\alpha \end{cases}$

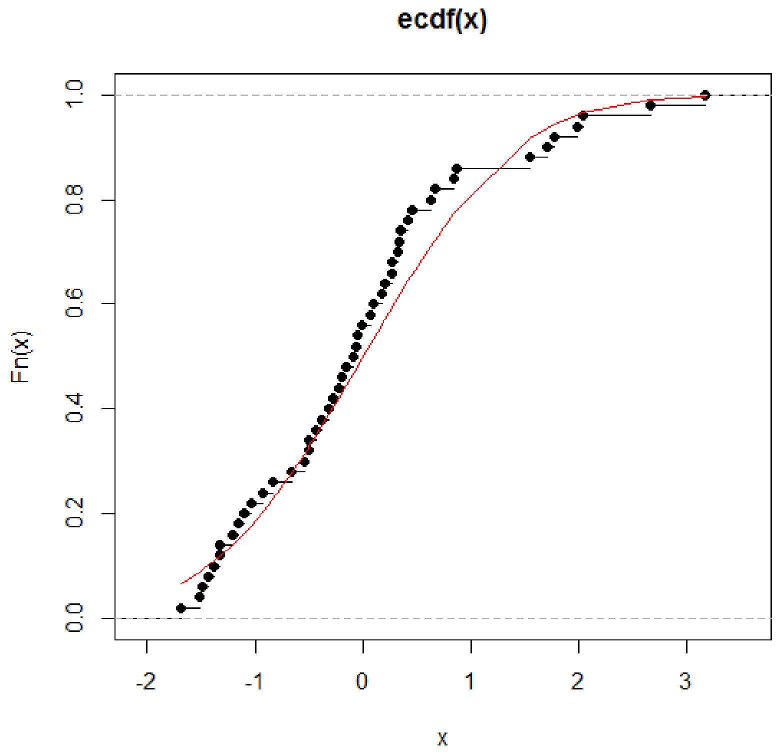
$$\overline{D_n} < x_\alpha$$

0.868099<13.2761

=>гипотезу принимаем

Статистика Колмогорова D_n имеет одно и тоже расстояние (зависящее только от n)

$$F_n \equiv F_0$$



3. **В предположении нормальности ошибок построить доверительные интервалы для параметров β_0 и β_1 уровня доверия $1 - \alpha$. Построить доверительный эллипс: уровня доверия $1 - \alpha$ для (β_0, β_1)(вычислить его полуоси)**

β_0_ β_0 *и* $\beta_1 \beta_1$ *уровня доверия 1-α_1.*

$$\widetilde{\beta_0} \sim N(\beta_0, var(\widetilde{\beta_0}))$$

$$b_0 = var(\widetilde{\beta_0}) = (XX^T)^{-1}_{11}$$

$$\widetilde{\beta_0}, \hat{\sigma}^2 - \text{ независимы } \rightarrow \frac{\widetilde{\beta_0} - \beta_0}{\hat{\sigma}\sqrt{b_0}} = S_{n-2}$$

$$P\left(-x_\alpha \le \frac{\widetilde{\beta_0} - \beta_0}{\hat{\sigma}\sqrt{b_0}} \le x_\alpha\right) = S_{n-2}(x_\alpha) - S_{n-2}(-x_\alpha) = 1 - \alpha_1$$

$$x_\alpha = S^{-1}_{n-2}\left(1 - \right. = 1.677224$$

$$-x_\alpha \le \frac{\widetilde{\beta_0} - \beta_0}{\hat{\sigma}\sqrt{b_0}} \le x_\alpha \Leftrightarrow -x_\alpha\hat{\sigma}\sqrt{b_0} \le \widetilde{\beta_0} - \beta_0 \le x_\alpha\hat{\sigma}\sqrt{b_0} \Leftrightarrow \widetilde{\beta_0} - x_\alpha\hat{\sigma}\sqrt{b_0} \le \beta_0 \le \widetilde{\beta_0} + x_\alpha\hat{\sigma}\sqrt{b_0}$$

ДИ [7.690489 : 8.994685]

$$\widetilde{\beta_1} \sim N(\beta_1, var(\widetilde{\beta_1}))$$

$$b_1 = var(\widetilde{\beta_1}) = (XX^T)^{-1}_{22}$$

$$\widetilde{\beta_1}, \hat{\sigma}^2 - \text{ независимы } \rightarrow \frac{\widetilde{\beta_1} - \beta_1}{\hat{\sigma}\sqrt{b_1}} = S_{n-2}$$

$$\widetilde{\beta_1} - x_\alpha\hat{\sigma}\sqrt{b_1} \le \beta_1 \le \widetilde{\beta_1} + x_\alpha\hat{\sigma}\sqrt{b_1}$$

ДИ: [-0.5533292 : 0.002781974]

4. **Сформулировать гипотезу независимости переменной Y от переменной X. Провести проверку значимости.**

$H_0: \beta_1 = 0$

Статистика критерия:

$$\mathcal{F} = \frac{\widehat{\beta_1}^2}{2b_1\hat{\sigma}^2} = F_{2,n-2}$$

$$x_\alpha = F^{-1}_{2,n-2}(1-\alpha_1) = 5.076664$$

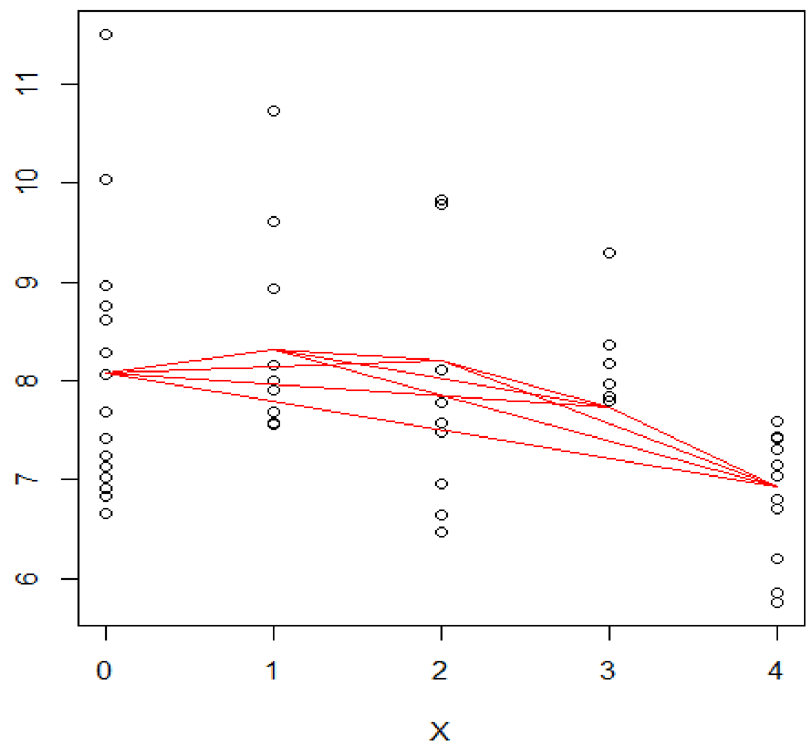
Критерий:

$$\Phi(\vec{x}) = \begin{cases} 0, \mathcal{F} < x_\alpha \\ 1, \mathcal{F} > x_\alpha \end{cases}$$

$$\mathcal{F} = 3.52549 < x_\alpha = 5.076664$$

Следовательно, гипотезу принимаем.

5. **Сформулировать модель, включающую дополнительный член с X^2. Построить МНК оценки параметров β_1 β_2 β_3 в данной модели. Изобразить графически полученную регрессионную зависимость.**
   График модели:



$$E(Y|X) = \beta_1 + \beta_2 X + \beta_3 X^2$$

$$\beta = (\beta_1, \beta_2, \beta_3)$$

Оценка по МНК параметров β:

$$\hat{\beta} = (XX^T)^{-1}XY$$

$$\widehat{\beta_1} = 8.0835365$$

$$\widehat{\beta_2} = 0.4091548$$

$$\widehat{\beta_3} = -0.1744058$$

6. **Построить несмещенную оценку дисперсии. Провести исследование нормальности ошибок, как в п.3.**

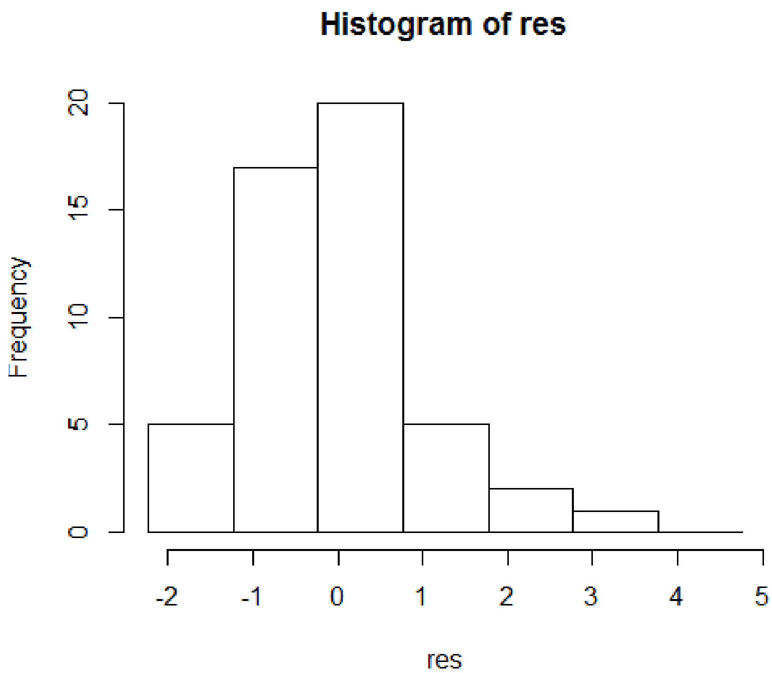
Несмещенная оценка дисперсии:

$$\hat{\sigma}^2 = \frac{\left\|Y - X^T\hat{\beta}\right\|^2}{n-3} = \frac{\sum(Y_i - \widehat{\beta_1} - \widehat{\beta_2}X_i - \widehat{\beta_3}X_i^2)^2}{n-3} = 1.184312$$

Ошибки $Y - X^T\hat{\beta}$:

```
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]      [,9]
[1,] -1.734223 -1.564223 -1.423537 -1.253537 -1.244223 -1.173537 -1.169662 -1.069662 -1.063537
          [,10]      [,11]      [,12]      [,13]      [,14]      [,15]      [,16]      [,17]
[1,] -0.9435365 -0.8435365 -0.7582855 -0.7482855 -0.7296625 -0.7242228 -0.6735365 -0.6342228
          [,18]      [,19]      [,20]      [,21]      [,22]      [,23]      [,24]      [,25]
[1,] -0.6282855 -0.4242228 -0.4082855 -0.4035365 -0.3182855 -0.2296625 -0.1582855 -0.1196625
           [,26]       [,27]      [,28]      [,29]     [,30]     [,31]     [,32]     [,33]
[1,] -0.08422278 -0.01353652 0.04865155 0.09865155 0.1103375 0.2064635 0.2203375 0.2286516
         [,34]     [,35]     [,36]     [,37]     [,38]     [,39]     [,40]     [,41]     [,42]
[1,] 0.3803375 0.4386516 0.4803375 0.5103375 0.5264635 0.6117145 0.6186516 0.6603375 0.6764635
         [,43]    [,44]    [,45]    [,46]    [,47]    [,48]    [,49]    [,50]
[1,] 0.8864635 1.291715 1.548652 1.575777 1.625777 1.956463 2.411715 3.426463
```

Гистограмма с шагом h на базе ошибок:



Основная гипотеза:

$H_0: Y - X^T\tilde{\beta} \sim N(0, \sigma^2)$;

Статистика критерия: $X^2 = \sum_{i=1}^{r} \frac{(v_i - np_i)^2}{np_i} \Rightarrow \chi^2_{r-2}$

$P_{\theta_0}(X^2 < x_\alpha) = 1 - \alpha_1$

	n_k	p_k	np_k	$n_k - np_k$	$\frac{(n_k - np_k)^2}{np_k}$
[-1.68; -1,03]	10	8.597797	429.8899	-419.8899	0.22868324
[-1,03; -0.31]	10	10.796023	539.8012	-529.8012	0.05369318
[-0,31; 0,09]	10	7.253942	362.6971	-352.6971	1.03954968
[0,09; 0,63]	10	9.285910	464.2955	-454.2955	0.05491384
[0,63; 3.17]	10	14.066328	703.3164	-693.3164	1.17550365
$x_\alpha =$ 13.2767		$\sum p_k =$ 50		$X^2 = \sum \frac{(n_k - np_k)^2}{np_k} =$ 2.557344	

$\boxed{\text{при } r = 5}$ $\quad X^2 = 2.557344$

$P_{\theta_0}(X^2 < x_\alpha) = 1 - \alpha_1$

$x_\alpha = K_{r-2}^{-1}(1 - \alpha_1) = \underline{13.2767}$

Критерий:

$$\Phi(\vec{x}) = \begin{cases} 0, X^2 < x_\alpha \\ 1, X^2 > x_\alpha \end{cases}$$

$X^2 < x_\alpha$ ---2.557344 < 13.2767 - принимаем гипотезу H_0.

Оценка расстояния полученной оценки до класса нормальных распределений по Колмогорову.

$H_0 : Y - X^T\hat{\beta} \sim N(0, \hat{\sigma}^2)$

Статистика критерия:

$\overline{D_n} = \sqrt{n} max_x |F_n(x) - F_0(x)| \to K(x)$

$P_{\theta_0}(\overline{D_n} < x_\alpha) = 1 - \alpha_1$

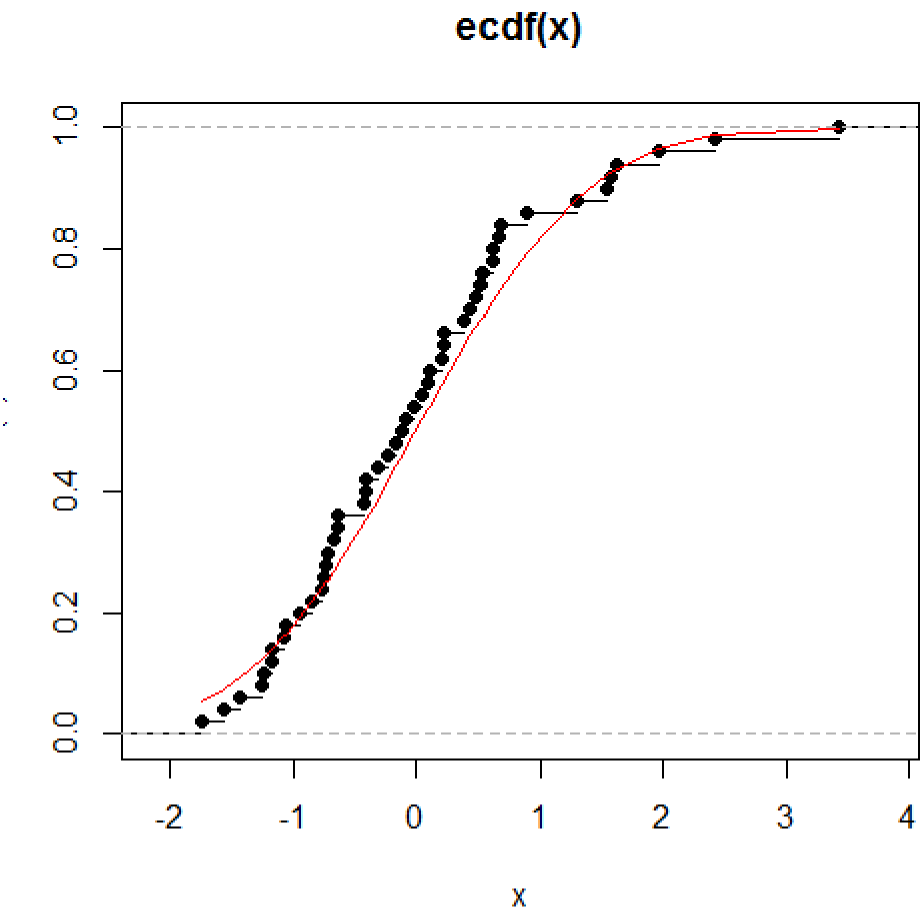
$x_\alpha = K^{-1}(1 - \alpha_1) = \underline{13.2767}$

Критерий:

$$\Phi(\vec{x}) = \begin{cases} 0, \overline{D_n} < x_\alpha \\ 1, \overline{D_n} > x_\alpha \end{cases}$$

$$\overline{D_n} = 0.7573278 < x_\alpha = \underline{13.2767}$$

Следовательно, гипотезу принимаем.



7. В предположении нормальности ошибок построить доверительные интервалы для параметров β_1 β_2 β_3 уровня доверия $1-\alpha$. Написать уравнение доверительного эллипсоида уровня доверия $1-\alpha$.

$\hat{\beta}_i \sim N(\beta_i, var(\hat{\beta}_i))$

$b_i = var(\hat{\beta}_i) = (XX^T)^{-1}_{iii}$

$\hat{\beta}_i, \hat{\sigma}^2 -$ независимы $\rightarrow \dfrac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{b_i}} = S_{n-3}$

$$P\left(-x_\alpha \le \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{b_i}} \le x_\alpha\right) = S_{n-3}(x_\alpha) - S_{n-3}(-x_\alpha) = 1 - \alpha_1$$

$$x_\alpha = S^{-1}_{n-3}\left(1 - \frac{\alpha_1}{2}\right) = 1.677927$$

$$-x_\alpha \le \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{b_i}} \le x_\alpha \Leftrightarrow -x_\alpha\hat{\sigma}\sqrt{b_i} \le \hat{\beta}_i - \beta_i \le x_\alpha\hat{\sigma}\sqrt{b_i} \Leftrightarrow \hat{\beta}_i - x_\alpha\hat{\sigma}\sqrt{b_i} \le \beta_i$$

$$\le \hat{\beta}_i + x_\alpha\,\hat{\sigma}\sqrt{b_i}$$

ДИ для β_1: $[7.355066;\ 8.812007]$

ДИ для β_2: $[-0.574061;\ 1.392371]$

ДИ для β_3: $[-0.4152745;\ 0.06646287]$

Уравнение доверительного эллипсоида:

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$B = C^T (XX^T)^{-1} C$$

$$\frac{\left(\hat{\beta} - \beta\right)^T B^{-1} (\hat{\beta} - \beta)}{3s^2} \stackrel{\text{def}}{=} F_{3,n-3} \begin{cases} \dfrac{\left(\hat{\beta} - \beta\right)^T B^{-1} (\hat{\beta} - \beta)}{\sigma^2} \stackrel{\text{def}}{=} \chi_3^2 \\ \dfrac{(n-2)s^2}{\sigma^2} \stackrel{\text{def}}{=} \chi_{n-3}^2 \end{cases}$$

$$x_\propto = F_{3,n-3}^{-1}(1-\alpha)$$

$$P\left(\frac{\left(\hat{\beta} - \beta\right)^T B^{-1} \left(\hat{\beta} - \beta\right)}{3s^2} < x_\alpha\right) = 1 - \alpha$$

$$=> P\left(\left(\hat{\beta} - \beta\right)^T B^{-1} \left(\hat{\beta} - \beta\right) < 3s^2 x_\alpha\right)$$

$\left(\hat{\beta} - \beta\right)^T B^{-1} \left(\hat{\beta} - \beta\right) < 3s^2 x_\alpha$ -уравнение доверительного эллипсоида

8. Сформулировать гипотезу линейной регрессионной зависимости переменной Y от переменной X. Провести проверку значимости на уровне α.

$H_0 : \beta_3 = 0$

Статистика критерия:

$$\mathcal{F} = \frac{\widehat{\beta_3}^2}{3 b_3 \hat{\sigma}^2} = F_{3,n-3}$$

$$x_\alpha = 3(1 - \alpha_1) = \ 1.259461$$

Критерий:

$$\Phi(\vec{x}) = \begin{cases} 0, \mathcal{F} < x_\alpha \\ 1, \mathcal{F} > x_\alpha \end{cases}$$

$$\mathcal{F} = 1.259461 < x_\alpha = \ 4.227901$$

Следовательно, гипотезу принимаем.