МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

МГТУ им Н.Э.Баумана

Факультет ФН
Кафедра вычислительной математики и математической физики

Соколов Арсений Андреевич

Домашнее задание №8 по математической статистике

3 курс, группа ФН11-53Б
Вариант 9

Преподаватель
_____________ Т. В. Облакова
«___» _______ 2019 г.

Москва, 2019 г.

Задание 1

Смоделировать выборку (X_k, Y_k) из двумерного гауссовского распределения объема $n = 140$ с данными параметрами $\vec{\mu} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ и $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$. Построить двумерную гистограмму и диаграмму рассеяния полученной выборки.

Решение.

Пусть (ξ, η) – - гауссовский вектор с параметрами $\vec{\mu} = \begin{pmatrix} a \\ b \end{pmatrix}$ и $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$

Пусть

$$\eta = b + \frac{\sigma_{12}}{\sigma_{11}}(\xi - a) + \varepsilon$$

Тогда $\varepsilon \sim N(0, \sqrt{\sigma_{22}(1 - r^2)})$ и не зависит от ξ, $r = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$.

Воспользуемся этим фактом в моделировании нашей выборки:

$$r = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \frac{-1}{\sqrt{1 \cdot 2}} = -0.7071068$$

$$X \sim N(a, \sqrt{\sigma_{11}}) \Leftrightarrow X \sim N(1, 1)$$

$$\varepsilon \sim N(0, \sqrt{\sigma_{22}(1 - r^2)}) \Leftrightarrow \varepsilon \sim N(0, 1)$$

```
> alpha <- 0.05
> n <- 140
> mu <- c(1,-3)
> Sigma <- matrix(c(1,-1,-1,2), byrow = T, ncol = 2)
> x <- rnorm(n,mu[1], sqrt(Sigma[1,1]))
> r <- Sigma[1,2] / sqrt(Sigma[1,1]*Sigma[2,2])
> r
[1] -0.7071068
> eps_k <- rnorm(n,mean = 0, sqrt(Sigma[2,2]*(1-r^2)))
> y <- mu[2] + Sigma[1,2]/Sigma[1,1] * (x-mu[1]) + eps_k
> df <- matrix(c(x,y,eps_k), ncol = 3)
```

Получаемая выборка имеет вид:

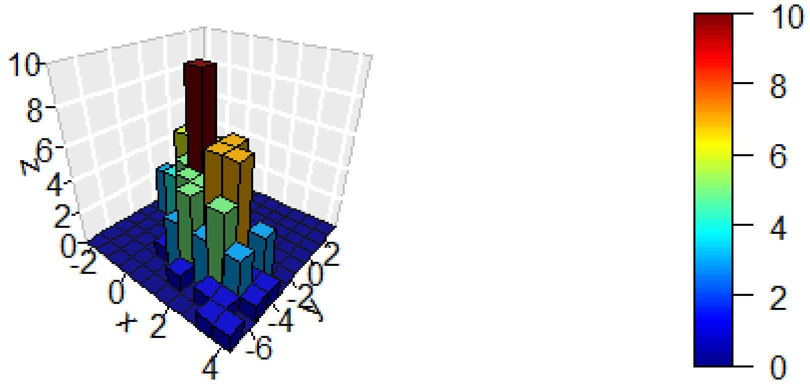
	df	y	eps
1	0.424167925	-1.00851823	1.415649695
2	-0.551432584	-1.94098479	-0.492417376
3	0.769429080	-4.42871994	-1.659290862
4	0.434355017	-2.00964707	0.424707949
5	1.321518384	-4.65436166	-1.332843279
6	0.940472747	-1.01509259	1.925380153
7	-0.015823884	-2.84507492	-0.860898806
8	1.168391726	-3.06498179	0.103409937
9	2.390894451	-6.52575031	-2.134855859
10	0.685590282	-3.20434410	-0.518753817
11	1.940803619	-5.54704973	-1.606246111
12	0.998858747	-4.48900625	-1.490147504
13	1.081047012	-3.42265924	-0.341612226
14	1.410887738	-2.68126898	0.729618763
15	0.966345072	-2.28897021	0.677374863
16	1.951856766	-4.14179975	-0.189942980
17	0.397044802	-0.17481439	2.222230417
18	0.602065793	-1.69584417	0.906221620
19	1.746899483	-2.74599118	1.000908307
20	0.772313519	-4.76139467	-1.989081151
21	-0.567206645	-1.27761691	0.155176445
22	-1.798718945	-0.56472139	-0.363440336
23	1.014804431	-1.42965786	1.585146568
24	-0.257471717	-2.87295996	-1.130431678
25	0.960351085	-1.75153864	1.208812450
26	1.166715052	-2.36325982	0.803455231
27	0.982154624	-3.88766283	-0.905508204
28	-1.807089886	2.65513739	2.848047500
29	-0.006434140	0.74325813	2.736823986
30	1.430450012	-3.00641272	0.424037292
31	-0.756980328	-0.33020225	0.912817419
32	2.360139574	-3.18187157	1.178268004
33	-0.150429883	-1.96738961	-0.117819489
34	1.821638896	-2.25613123	1.565507671
35	2.393914762	-2.80022832	1.593686442
36	-0.393541020	-1.17454012	0.431918857
37	2.030472916	-3.20983270	0.820640212
38	1.118545664	-3.90792951	-0.789383841
39	0.767480890	-2.76097974	0.006501155
40	1.022646121	-3.31377689	-0.291130766
41	-0.003502764	-2.35848219	-0.361984953
42	0.956367062	-4.19914174	-1.242774678
43	0.872905697	-3.86770089	-0.994795197
44	0.475716532	-2.99474857	-0.519032033
45	-0.845511369	-0.15019107	1.004297564

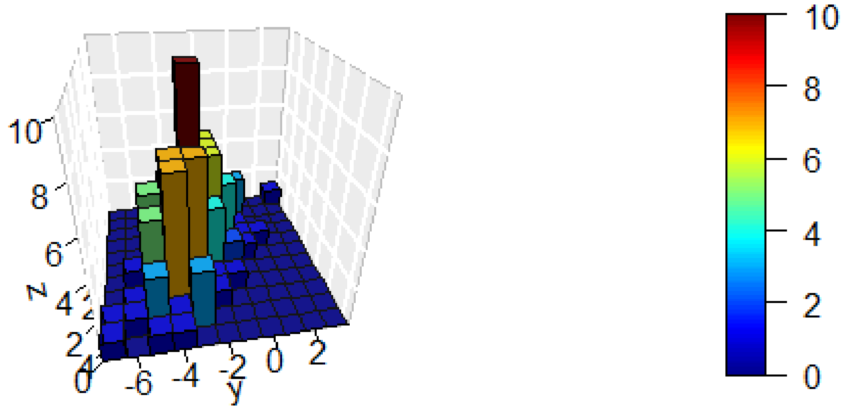
```
> summary(df)
       x                 y                  eps         
 Min.   :-1.8071   Min.   :-6.526   Min.   :-2.1349  
 1st Qu.: 0.2569   1st Qu.:-3.735   1st Qu.:-0.4840  
 Median : 0.8711   Median :-2.781   Median : 0.1044  
 Mean   : 0.8561   Mean   :-2.736   Mean   : 0.1206  
 3rd Qu.: 1.5052   3rd Qu.:-1.890   3rd Qu.: 0.8037  
 Max.   : 3.1440   Max.   : 2.655   Max.   : 2.8480  
> var(df[,1]) # Дисперсия x
[1] 0.945916
> var(df[,2]) # Дисперсия y
[1] 1.910854
> var(df[,3]) # Дисперсия eps
[1] 0.9242728
```

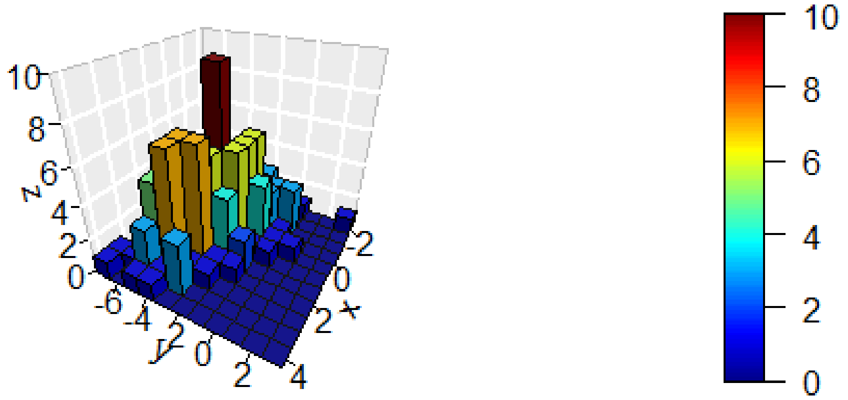
По полученной выборке построим двумерную гистограмму:

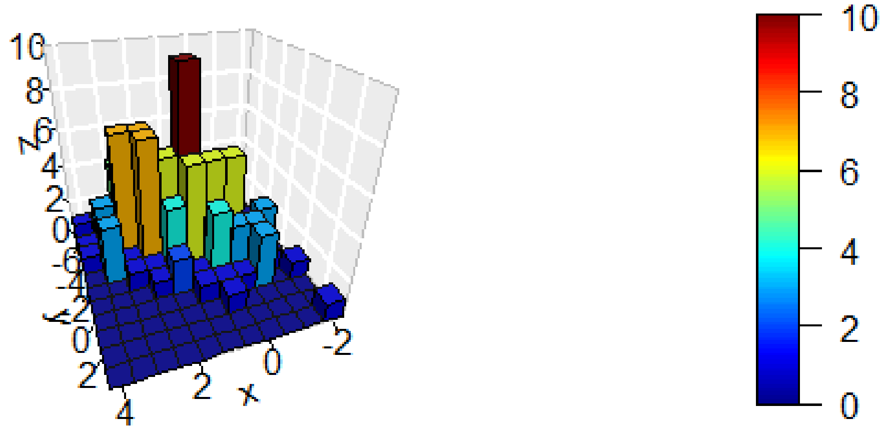
```
> num_bins <- length(hist(df[,2], breaks = "Sturges", freq = T)$breaks)
> x_c <- cut(df[,1], num_bins)
> y_c <- cut(df[,2], num_bins)
> z <- table(x_c, y_c)
> library(plot3D)
>
> main_perspective1 <- hist3D(x = seq(from = floor(min(df[,1])),
+                                     ceiling(max(df[,1])),
+                                     length.out = nrow(z)),
+                             y = seq(from = floor(min(df[,2])),
+                                     ceiling(max(df[,2])),
+                                     length.out = nrow(z)),
+                             z=z, border="black",
+                             ticktype = "detailed", lighting=T,
+                             bty = "g", phi = 30, theta = 40,
+                             lphi = 50)
>
> main_perspective2 <- hist3D(x = seq(from = floor(min(df[,1])),
+                                     ceiling(max(df[,1])),
+                                     length.out = nrow(z)),
+                             y = seq(from = floor(min(df[,2])),
+                                     ceiling(max(df[,2])),
+                                     length.out = nrow(z)),
+                             z=z, border="black",
+                             ticktype = "detailed", lighting=T,
```

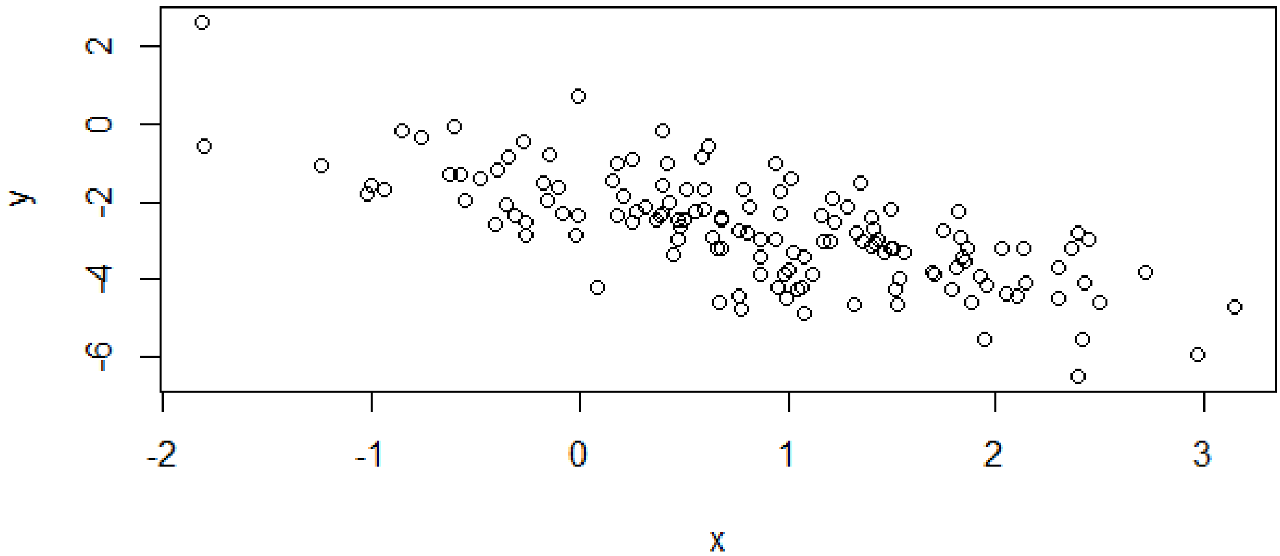
```
+                             bty = "g", phi = 30, theta = 80,
+                             lphi = 50)
>
> main_perspective3 <- hist3D(x = seq(from = floor(min(df[,1])),
+                                     ceiling(max(df[,1])),
+                                     length.out = nrow(z)),
+                             y = seq(from = floor(min(df[,2])),
+                                     ceiling(max(df[,2])),
+                                     length.out = nrow(z)),
+                             z=z, border="black",
+                             ticktype = "detailed", lighting=T,
+                             bty = "g", phi = 30, theta = 120,
+                             lphi = 50)
>
> main_perspective4 <- hist3D(x = seq(from = floor(min(df[,1])),
+                                     ceiling(max(df[,1])),
+                                     length.out = nrow(z)),
+                             y = seq(from = floor(min(df[,2])),
+                                     ceiling(max(df[,2])),
+                                     length.out = nrow(z)),
+                             z=z, border="black",
+                             ticktype = "detailed", lighting=T,
+                             bty = "g", phi = 30, theta = 160,
+                             lphi = 50)
> plot(df)
```


z
x
y
10
8
6
4
2
0


z
y
-6
-4
-2
0
2
10
8
6
4
2
0


z
x
y
10
8
6
4
2
0


z
x
y
10
8
6
4
2
0



y
2
0
-2
-4
-6
-2
-1
0
1
2
3
x

Задание 2

Найти по методу наименьших квадратов оценки коэффициентов $\hat{\beta}_1$ и $\hat{\beta}_0$ линейной регрессии $Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k$ И остаточной дисперсии $\hat{\delta^2}$, $D\varepsilon_k = \delta^2$. Построить совмещённые графики диаграммы рассеяния и линии регрессии.
Решение.
Суть МНК[1]:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1}$$
$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_i \tag{2}$$

$$\frac{\sum\limits_{i=1}^{n} x_i}{n} = \overline{x} \tag{3}$$

Из 2 получаем очевидное соотношение:

$$\sum_{i=1}^{n} x_i = n \cdot \overline{x} = \sum_{i=1}^{n} \overline{x} \Longrightarrow \sum_{i=1}^{n} (x_i - \overline{x}) = 0 \tag{4}$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{b}_1 - \hat{\beta}_2 \cdot x_i)^2 = Q(\hat{\beta}_1, \hat{\beta}_2) \tag{5}$$

$Q(\hat{\beta}_1, \hat{\beta}_2) \longrightarrow \min$

$$\begin{cases} \frac{\partial Q}{\partial \hat{\beta}_1} = \sum\limits_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i)(-1) = 0 \\ \frac{\partial Q}{\partial \hat{\beta}_2} = \sum\limits_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i)(-x_i) = 0 \end{cases} \tag{6}$$

$(6) \Longrightarrow (\hat{\beta}_1, \hat{\beta}_2)$

Заметим, что выражение в скобках в (6) является ошибкой прогноза $\hat{\varepsilon}_i$. Тогда наша система (6) перепишется в более простом виде:

$$\begin{cases} \sum\limits_{i=1}^{n} \hat{\varepsilon}_i \cdot 1 = 0 \\ \sum\limits_{i=1}^{n} \hat{\varepsilon}_i \cdot x_i = 0 \end{cases} \tag{7}$$

[1] В выводе используются более удобные (мне) обозначения. Далее в коде и в решении возобновляется использование Ваших обозначений

Для нахождения $\hat{\beta}_1, \hat{\beta}_2$ решим относительно них систему (6):

$$\begin{cases} \sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} \hat{\beta}_1 - \sum\limits_{i=1}^{n} \hat{\beta}_2 \cdot x_i = 0 \\ \sum\limits_{i=1}^{n} y_i x_i - \sum\limits_{i=1}^{n} \hat{\beta}_1 x_i - \sum\limits_{i=1}^{n} \hat{\beta}_2 x_i^2 = 0 \end{cases} \tag{8}$$

$$\begin{cases} \sum\limits_{i=1}^{n} y_i - n\beta_1 - \hat{\beta}_2 \sum\limits_{i=1}^{n} x_i = 0 \\ \sum\limits_{i=1}^{n} y_i x_i - \hat{\beta}_1 \sum\limits_{i=1}^{n} x_i - \hat{\beta}_2 \sum\limits_{i=1}^{n} x_i^2 = 0 \end{cases} \tag{9}$$

Из первого уравнения (9) при делении на n следует:

$$\overline{y} - \hat{\beta}_1 - \hat{\beta}_2 \overline{x} = 0 \tag{10}$$

Или

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{x} \tag{11}$$

Выразим β_1 из (11) и подставим во второе уравнение (9):

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \cdot \overline{x} \tag{12}$$

$$\sum_{i=1}^{n} y_i x_i - (\overline{y} - \hat{\beta}_2 \cdot \overline{x}) \sum_{i=1}^{n} x_i - \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0 \tag{13}$$

$$\sum_{i=1}^{n} y_i x_i - \overline{y} \sum_{i=1}^{n} x_i + \hat{\beta}_2 \cdot \overline{x} \sum_{i=1}^{n} x_i - \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0 \tag{14}$$

$$\hat{\beta}_2 (\overline{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2) = \overline{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i \tag{15}$$

$$\hat{\beta}_2 = \frac{\overline{y} \sum\limits_{i=1}^{n} x_i - \sum\limits_{i=1}^{n} y_i x_i}{\overline{x} \sum\limits_{i=1}^{n} x_i - \sum\limits_{i=1}^{n} x_i^2} \tag{16}$$

Уже получили правильный ответ, но приведём решение к более красивому виду:

$$\hat{\beta}_2 = \frac{\sum\limits_{i=1}^{n} y_i x_i - \sum\limits_{i=1}^{n} \overline{y} x_i}{\sum\limits_{i=1}^{n} x_i^2 - \sum\limits_{i=1}^{n} \overline{x} x_i} = \frac{\sum\limits_{i=1}^{n} (y_i - \overline{y}) x_i}{\sum\limits_{i=1}^{n} (x_i - \overline{x}) x_i} \tag{17}$$

Тут уже лучше видно, что числитель и знаменатель похожи. Но сделаем ещё одно преобразование, которое из правильного ответа сделает правильный ответ.

Вспомним про (4). Из него следует интересный факт:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0 \Longrightarrow \overline{x}\sum_{i=1}^{n}(x_i - \overline{x}) = 0 \Longrightarrow \sum_{i=1}^{n}\overline{x}(x_i - \overline{x}) = 0 \tag{18}$$

Тогда вычтем нули из числителя и знаменателя в (17):

$$\hat{\beta}_2 = \frac{\sum\limits_{i=1}^{n}(y_i - \overline{y})x_i - \sum\limits_{i=1}^{n}\overline{x}(y_i - \overline{y})}{\sum\limits_{i=1}^{n}(x_i - \overline{x})x_i - \sum\limits_{i=1}^{n}\overline{x}(x_i - \overline{x})} \tag{19}$$

Внесём всё под один знак суммы:

$$\hat{\beta}_2 = \frac{\sum\limits_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} \tag{20}$$

Эта форма записи хороша тем, что везде фигурируют отклонения наблюдений от среднего значения.

Таким образом получили выражения для расчёта $\hat{\beta}_1, \hat{\beta}_2$:

$$\begin{cases} \hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\ \hat{\beta}_1 = \overline{y} - \hat{\beta}_2\overline{x} \end{cases} \tag{21}$$

Получаем:

```
> beta_1 <- (sum((df[,2] - meany)*(df[,1] - meanx))/
+   (sum((df[,1] - meanx)^2))
> beta_1
[1] -1.021495
> beta_0  <- meany - beta_1 * meanx
> beta_0
[1] -1.861009
```

Или

```
> lm1 <- lm(df[,2] ~ df[,1])
> summary(lm1)
```

```
Call:
lm(formula = df[, 2] ~ df[, 1])

Residuals:
     Min       1Q   Median       3Q      Max
-2.26208 -0.60113 -0.02659  0.67293  2.67021

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.86101    0.10878  -17.11   <2e-16 ***
df[, 1]     -1.02150    0.08413  -12.14   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9646 on 138 degrees of freedom
Multiple R-squared:  0.5165,    Adjusted R-squared:  0.513
F-statistic: 147.4 on 1 and 138 DF,  p-value: < 2.2e-16
```

Видим, что результаты совпадают.
Можем записать полученное уравнение линейной регрессии:

$$Y = -1.02150 - 1.86101 \cdot X$$

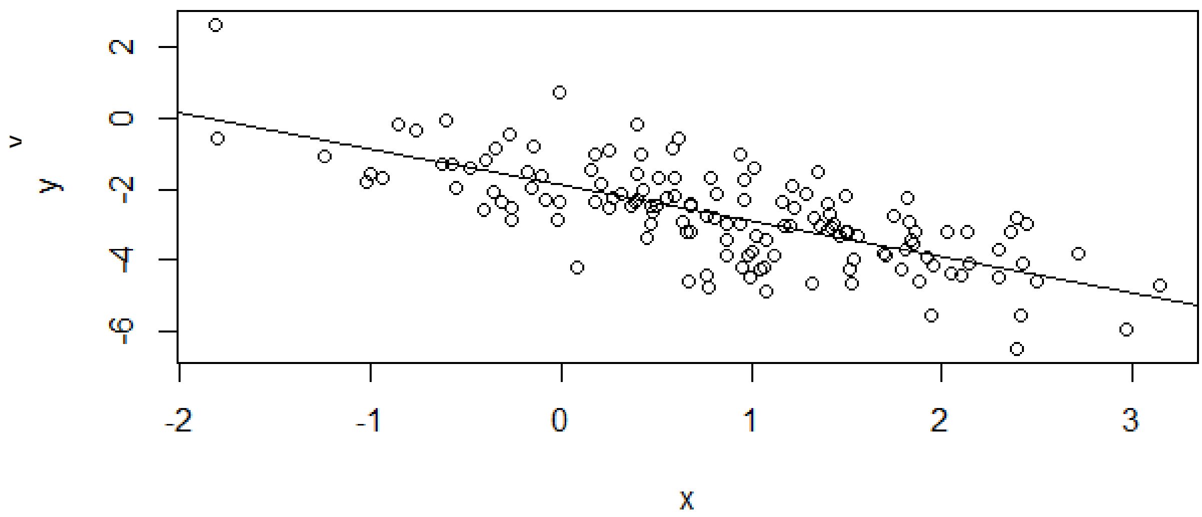
Остаточная дисперсия равна

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

```
> sigma_ost <- (sum((df[,2] - beta_1 *
+                    df[,1] - beta_0)^2))/(n-2)
> sigma_ost
[1] 0.9305302
```

Построим совмещённые графики диаграммы рассеяния и линии регрессии:

```
> plot(df)
> abline(lm1)
```



y
2
0
-2
-4
-6
-2
-1
0
1
2
3
x

Задание 3

Найти доверительные интервалы с доверительной вероятностью $1-\alpha=0.95$

1. Доверительные интервал для коэффициента корреляции компонент r

Рассмотрим следующие статистики:

$$\mu_{20} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k-\bar{X}\right)^2$$
$$\mu_{02} = \frac{1}{n}\sum_{k=1}^{n}\left(Y_k-\bar{Y}\right)^2$$
$$\mu_{11} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k-\bar{X}\right)\left(Y_k-\bar{Y}\right)$$

```
> mu20 <- 1/n * sum((df[,1] - meanx)^2)
> mu20
[1] 0.9391594
> mu02 <- 1/n * sum((df[,2] - meany)^2)
> mu02
[1] 1.897205
> mu11 <- 1/n * sum((df[,1] - meanx)*(df[,2] - meany))
> mu11
[1] -0.9593467
```

Тогда выборочный коэффициент корреляции равен:

$$r_{\text{В}} = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}$$

```
> rv <- mu11/sqrt(mu20*mu02)
> rv
[1] -0.718702
```

Рассмотрим статистику $z = \frac{1}{2}\ln\frac{1+r_{\text{В}}}{1-r_{\text{В}}} = \operatorname{arctanh} r_{\text{В}}$, которая при $n \geq 10$ приближённо распределена $\sim N(a_z, \frac{1}{\sqrt{n-3}})$, где $a_z = \operatorname{arctanh}(r) + \frac{r}{(2n-1)}$.

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

Доверительный интервал для коэффициента корреляции компонент r:

$$1-\alpha = P\Big(-u_{1-\frac{\alpha}{2}} < (z-a_z)\sqrt{n-3} < u_{1-\frac{\alpha}{2}}\Big) = P\Big(z - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}} < a_z < z + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\Big) =$$
$$= P\Big(\text{arctanh } r_{\text{в}} - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}} < \text{arctanh } r + \frac{r}{2(n-1)} < \text{arctanh } r_{\text{в}} + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\Big)$$

Итого:

$$P\Big(\tanh\Big(\text{arctanh } r_{\text{в}} - \tfrac{r_{\text{в}}}{2(n-1)} - \tfrac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\Big) < r < \tanh\Big(\text{arctanh } r_{\text{в}} - \tfrac{r_{\text{в}}}{2(n-1)} + \tfrac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\Big)\Big)$$

Получаем:

```
> CIr_lower_bound <- tanh(atanh(rv) - rv/(2*(n-1))
+                   - qnorm(1-alpha/2)/sqrt(n-3))
> CIr_lower_bound
[1] -0.7893937
> CIr_upper_bound <- tanh(atanh(rv) - rv/(2*(n-1))
+                   + qnorm(1-alpha/2)/sqrt(n-3))
> CIr_upper_bound
[1] -0.6260653
```

Получаем доверительный интервал:

$$-0.7893937 \le r \le -0.6260653$$

Истинное значение коэффициента корреляции

$$r = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = -0.7071068$$

попадает в полученный интервал.

2. Доверительный интервал для коэффициентов регресии β_0 и β_1

Запишем матрицу базисных функций:

$$F = \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix}, \qquad F^T = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix}$$

$$F^T F = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum\limits_{k=1}^{n} X_k \\ \sum\limits_{k=1}^{n} X_k & \sum\limits_{k=1}^{n} X_k^2 \end{pmatrix} = \begin{pmatrix} 140 & 119.8563 \\ 119.8563 & 234.0934 \end{pmatrix}$$

$$(F^TF)^{-1} = \frac{1}{n\sum\limits_{k=1}^{n} X_k^2 - \left(\sum\limits_{k=1}^{n} X_k^2\right)^2} \begin{pmatrix} \sum\limits_{k=1}^{n} X_k^2 & -\sum\limits_{k=1}^{n} X_k \\ -\sum\limits_{k=1}^{n} X_k & n \end{pmatrix} = \begin{pmatrix} 0.012717264 & -0.006511269 \\ -0.006511269 & 0.007605585 \end{pmatrix}$$

$\hat{\beta}_i \sim N(\beta_i, \hat{\delta}\sqrt{\tau_{ii}})$, где τ_{ii} – элемент матрицы $(F^TF)^{-1}$.

В общем случаем доверительный интервал для коэффициента регрессии β_i В случае $\varepsilon_k \sim N(0, \delta)$ при неизвестном δ:

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

Доверительный интервал для коэффициента регрессии β_0:

$$\hat{\beta}_0 - t_{1-\frac{\alpha}{2}}(n-m)\hat{\delta}\sqrt{\tau_{11}} < \beta_0 < \hat{\beta}_0 + t_{1-\frac{\alpha}{2}}(n-m)\hat{\delta}\sqrt{\tau_{11}}$$

```
> M <- matrix(c(n, sum(df[,1]),
+                 sum(df[,1]),sum((df[,1])^2)),
+               ncol = 2, byrow = T)
> M_obr <- solve(M)
>
>
>
> CIbeta0_lower_bound <- beta_0 -
+   qt(1-alpha/2, n-2) * sqrt(sigma_ost) * sqrt(M_obr[1,1])
> CIbeta0_lower_bound
[1] -2.076106
> CIbeta0_upper_bound <- beta_0 +
+   qt(1-alpha/2, n-2) * sqrt(sigma_ost) * sqrt(M_obr[1,1])
> CIbeta0_upper_bound
[1] -1.645911
```

Получаем доверительный интервал для β_0:

$$-2.076106 \le \beta_0 \le -1.645911$$

Доверительный интервал для коэффициента регрессии β_1:

$$\hat{\beta}_1 - t_{1-\frac{\alpha}{2}}(n-m)\hat{\delta}\sqrt{\tau_{11}} < \beta_1 < \hat{\beta}_1 + t_{1-\frac{\alpha}{2}}(n-m)\hat{\delta}\sqrt{\tau_{11}} \tag{22}$$

```
> CIbeta1_lower_bound <-  beta_1 -
+   (qt(1-alpha/2, n-2) * sqrt(sigma_ost)) * sqrt(M_obr[2,2])
> CIbeta1_lower_bound
```

```
[1] -1.187838
> CIbeta1_upper_bound <-  beta_1 +
+   (qt(1-alpha/2, n-2) * sqrt(sigma_ost)) * sqrt(M_obr[2,2])
> CIbeta1_upper_bound
[1] -0.8551519
```

Получаем доверительный интервал для β_1:

$$-1.187838 \leq \beta_1 \leq -0.8551519$$

Или

```
> confint(lm1)
                2.5 %     97.5 %
(Intercept) -2.076106 -1.6459110
df[, 1]     -1.187838 -0.8551519
```

3. Доверительный интервал для дисперсии δ^2

Доверительный интервал для δ^2 в случае $\varepsilon_k \sim N(0, \delta)$ при неизвестном δ:

$$\frac{\hat{\delta}^2(n-2)}{\delta^2} \sim \chi^2(n-2)$$

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

Следовательно:

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

```
> CIdelta_ost_upper_boud<- (n-2)*delta_ost /
+   (qchisq(alpha/2, n-2))
> CIdelta_ost_upper_boud
[1] 1.195962
> CIdelta_ost_lower_boud <- (n-2)*delta_ost /
+   (qchisq(1-alpha/2, n-2))
> CIdelta_ost_lower_boud
[1] 0.7448023
```

Получаем доверительный интервал для δ^2:

$$0.7448023 \leq \delta^2 \leq 1.195962$$