

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  $(\xi, \eta)$  - 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  $\xi$ ,  $r = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ .

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:

	x	y	$\varepsilon$
1	0.15914	-2.95514	-0.79599
2	2.38436	-4.41371	-0.02935
3	-0.25549	0.43573	2.18024
4	1.07014	-2.11272	0.95742
5	2.71144	-5.01649	-0.30505
6	0.39709	-2.81550	-0.41840
7	0.52783	-2.42788	0.09995
8	0.36463	-2.59444	-0.22981
9	0.71423	-4.12944	-1.41521
10	1.13811	-3.53071	-0.39260
11	2.22763	-3.28154	0.94609
12	0.19822	-1.44645	0.75177
13	-0.08039	-2.43698	-0.51738
14	0.84247	-2.03413	0.80834
15	-0.07176	-2.54278	-0.61454
16	0.86101	-1.62275	1.23826
17	0.40269	-2.74078	-0.33810
18	-1.18397	0.38033	1.19637
19	1.24082	-3.68414	-0.44332
20	0.74064	-2.55453	0.18611
21	1.90051	-6.52186	-2.62134
22	1.94187	-1.69561	2.24625
23	2.46796	-4.37453	0.09343
24	1.70676	-2.07948	1.62728
25	1.81901	-4.32993	-0.51092
26	0.70652	-3.36590	-0.65938
27	2.41859	-4.45878	-0.04019
28	2.49877	-4.61747	-0.11869
29	0.34292	-2.36257	-0.01966
30	0.14720	-2.63288	-0.48568
31	1.31592	-4.75606	-1.44015
32	2.10969	-3.96593	0.14377
33	3.21546	-6.45005	-1.23459
34	2.21710	-5.96960	-1.75250
35	2.47922	-4.51472	-0.03550
36	1.95157	-3.61954	0.33203
37	-0.00953	-0.41818	1.57229
38	-1.00047	-2.06900	-1.06947
39	-0.76219	-0.32153	0.91629
40	0.85739	-3.45238	-0.59499

41	2.55006	-2.36841	2.18165
42	0.19758	-2.88135	-0.68377
43	0.92542	-2.17536	0.75006
44	2.89567	-3.92129	0.97438
45	0.54343	-3.80790	-1.26447
46	1.56222	-3.83964	-0.27742
47	0.11299	-2.30239	-0.18940
48	0.53976	-2.92378	-0.38402
49	0.27567	-1.53508	0.74059
50	0.93079	-4.09913	-1.16834
51	2.46325	-3.79571	0.66754
52	1.18773	-2.82149	0.36624
53	2.02202	-4.53697	-0.51494
54	0.40817	-1.95760	0.45057
55	0.88780	-3.07552	-0.18772
56	0.07505	-0.73598	1.33907
57	1.75330	-2.93709	0.81622
58	0.88739	-2.80519	0.08220
59	0.93591	-3.58677	-0.65086
60	1.23328	-2.50687	0.72641
61	-0.13658	-1.97710	-0.11368
62	1.85483	-4.14993	-0.29510
63	0.42163	-1.43246	0.98917
64	1.49636	-4.27149	-0.77513
65	0.23994	-1.96404	0.27590
66	0.65861	-2.24783	0.41078
67	-1.10233	-0.28649	0.61118
68	0.69830	-1.76173	0.93657
69	-0.27238	-2.09516	-0.36754
70	0.72033	-1.97996	0.74038
71	0.79590	-1.57737	1.21853
72	0.77439	-2.14525	0.62913
73	1.34703	-2.81928	0.52775
74	1.03237	-3.50462	-0.47226
75	1.41353	-2.58982	0.82372
76	0.84465	-3.27244	-0.42779
77	1.97349	-4.11613	-0.14264
78	1.12109	-1.70231	1.41878
79	1.18917	-2.70204	0.48713
80	0.43711	-1.83367	0.60344

81	1.49842	-3.28758	0.21083
82	-0.74230	-1.29100	-0.03330
83	1.97553	-1.95033	2.02520
84	0.97592	-3.34670	-0.37079
85	1.67568	-5.25392	-1.57823
86	0.28969	-2.41126	-0.12157
87	3.38723	-7.18391	-1.79668
88	0.52657	-3.00216	-0.47559
89	0.92423	-3.80833	-0.88410
90	0.47816	-5.97622	-3.49806
91	1.92605	-4.30803	-0.38198
92	-0.06241	-0.95990	0.97769
93	1.55703	-4.11507	-0.55804
94	1.90073	-4.52719	-0.62646
95	1.98995	-4.52040	-0.53045
96	1.38361	-1.48599	1.89762
97	0.65342	-1.25788	1.39554
98	0.45981	-3.20584	-0.74603
99	0.81744	-3.12302	-0.30557
100	0.94070	-1.77102	1.16968
101	-0.99539	-0.70023	0.30439
102	2.13531	-4.25281	-0.11750
103	1.67579	-3.73588	-0.06009
104	1.20848	-1.73754	1.47094
105	0.94215	-4.42030	-1.47815
106	1.89381	-4.57742	-0.68361
107	0.77113	-2.31059	0.46054
108	-0.96565	-1.21585	-0.18150
109	0.24649	-3.40531	-1.15882
110	2.28015	-3.87113	0.40902
111	0.04710	-2.30530	-0.25821
112	2.62238	-4.88928	-0.26690
113	3.60014	-5.43599	0.16416
114	1.13965	-3.53311	-0.39346
115	-0.35072	-3.49302	-1.84374
116	1.79893	-5.34122	-1.54229
117	-0.55500	-2.03124	-0.58624
118	1.46372	-4.31586	-0.85214
119	1.05243	-2.27411	0.77832
120	0.79797	-2.82829	-0.03032

121	2.17086	-5.62651	-1.45566
122	1.88484	-3.79106	0.09378
123	-0.31789	-0.69976	0.98235
124	-0.64325	-1.95346	-0.59671
125	2.05925	-3.98445	0.07480
126	1.29008	-1.09265	2.19743
127	0.59997	-1.80494	0.79502
128	2.24310	-4.78204	-0.53894
129	-0.36641	-3.23487	-1.60128
130	-0.44141	-2.28996	-0.73137
131	2.34855	-4.70429	-0.35574
132	-0.97853	-2.00689	-0.98541
133	-0.24095	-2.49022	-0.73117
134	0.89596	-1.43064	1.46532
135	1.73297	-1.87436	1.85862
136	1.45568	-3.45218	0.00350
137	1.28808	-4.63185	-1.34378
138	-0.07369	-1.77502	0.15129
139	1.64874	-3.35873	0.29001
140	1.29916	-3.42164	-0.12248

```

> summary(df)
  x              y              eps
Min.   :-1.1840  Min.   :-7.1839  Min.   :-3.498059
1st Qu.: 0.3592  1st Qu.: -4.0131  1st Qu.: -0.565091
Median : 0.9333  Median : -2.9026  Median : -0.115588
Mean   : 1.0131  Mean   : -3.0172  Mean   : -0.004149
3rd Qu.: 1.8039  3rd Qu.: -2.0002  3rd Qu.: 0.729901
Max.   : 3.6001  Max.   : 0.4357   Max.   : 2.246255
> var(df[,1]) # Дисперсия x
[1] 1.006593
> var(df[,2]) # Дисперсия y
[1] 1.996038
> var(df[,3]) # Дисперсия eps
[1] 0.9753832

```

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,

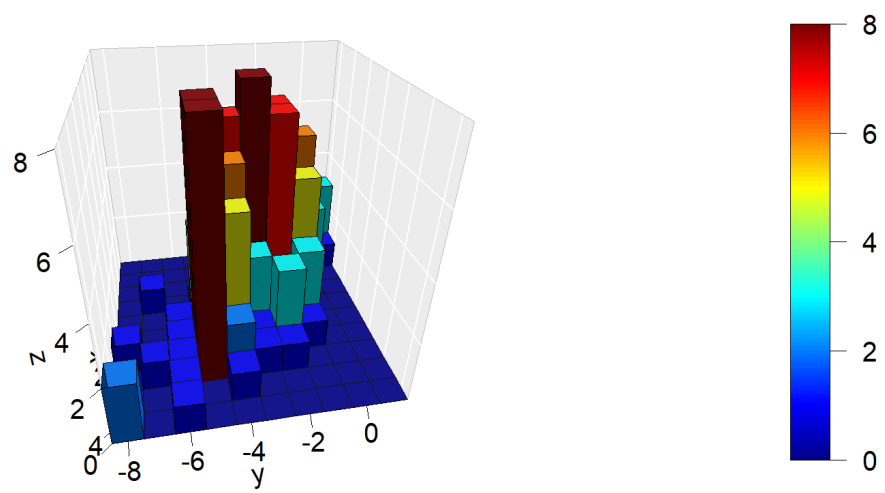
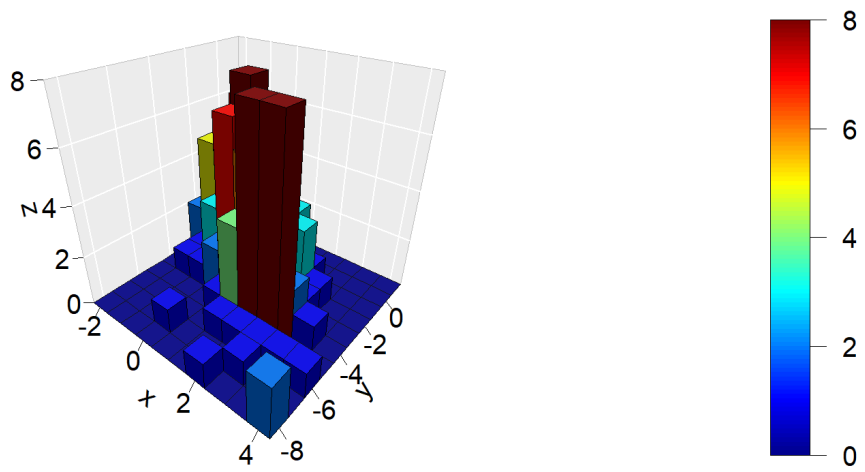
```

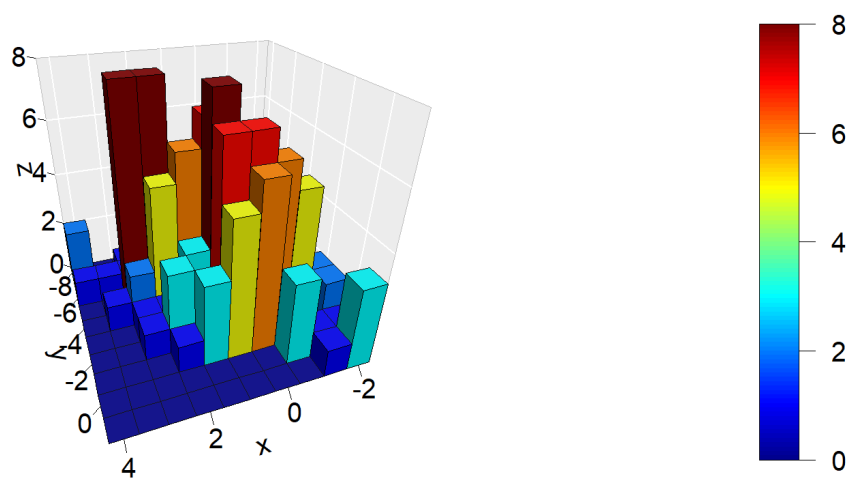
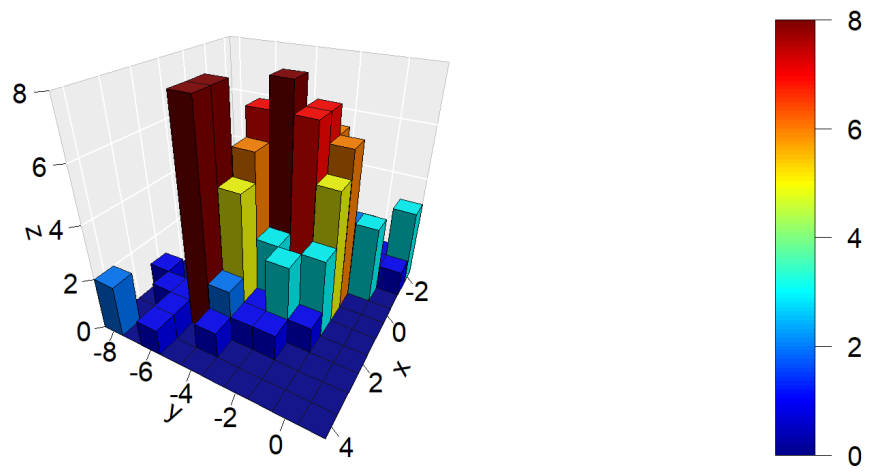
```

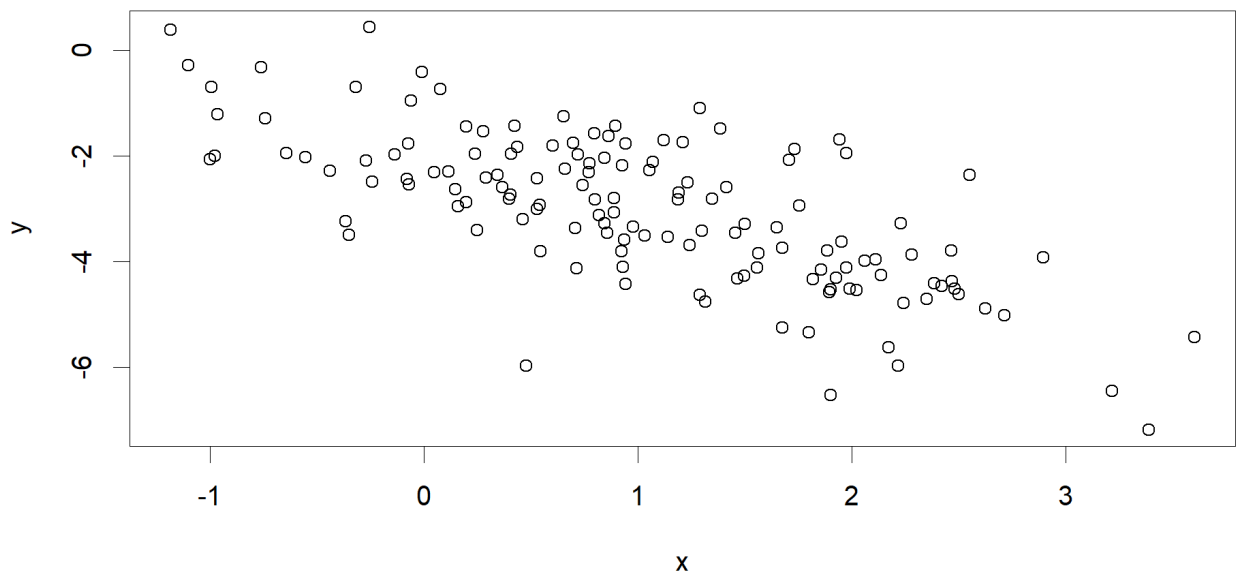
+         bty = "g", phi = 30, theta = 80,
+         lphi = 50)
>
> main_perspective3 <- hist3D(x = seq(from = floor(min(df[,1])),
+         ceiling(max(df[,1])),
+         length.out = nrow(z)),
+         y = seq(from = floor(min(df[,2])),
+         ceiling(max(df[,2])),
+         length.out = nrow(z)),
+         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)

```









## 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<sup>1</sup>:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \quad (1)$$

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_i \quad (2)$$

$$\frac{\sum_{i=1}^n x_i}{n} = \bar{x} \quad (3)$$

Из 2 получаем очевидное соотношение:

$$\sum_{i=1}^n x_i = n \cdot \bar{x} = \sum_{i=1}^n \bar{x} \implies \sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (4)$$

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

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

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

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

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

$$\begin{cases} \sum_{i=1}^n \hat{\varepsilon}_i \cdot 1 = 0 \\ \sum_{i=1}^n \hat{\varepsilon}_i \cdot x_i = 0 \end{cases} \quad (7)$$

---

<sup>1</sup>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):

$$\begin{cases} \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_1 - \sum_{i=1}^n \hat{\beta}_2 \cdot x_i = 0 \\ \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_1 x_i - \sum_{i=1}^n \hat{\beta}_2 x_i^2 = 0 \end{cases} \quad (8)$$

$$\begin{cases} \sum_{i=1}^n y_i - n\beta_1 - \hat{\beta}_2 \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i x_i - \hat{\beta}_1 \sum_{i=1}^n x_i - \hat{\beta}_2 \sum_{i=1}^n x_i^2 = 0 \end{cases} \quad (9)$$

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

$$\bar{y} - \hat{\beta}_1 - \hat{\beta}_2 \bar{x} = 0 \quad (10)$$

Или

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} \quad (11)$$

Выразим  $\hat{\beta}_1$  из (11) и подставим во второе уравнение (9):

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \cdot \bar{x} \quad (12)$$

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

$$\sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i + \hat{\beta}_2 \cdot \bar{x} \sum_{i=1}^n x_i - \hat{\beta}_2 \sum_{i=1}^n x_i^2 = 0 \quad (14)$$

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

$$\hat{\beta}_2 = \frac{\bar{y} \sum_{i=1}^n x_i - \sum_{i=1}^n y_i x_i}{\bar{x} \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2} \quad (16)$$

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

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \bar{y} x_i}{\sum_{i=1}^n x_i^2 - \sum_{i=1}^n \bar{x} x_i} = \frac{\sum_{i=1}^n (y_i - \bar{y}) x_i}{\sum_{i=1}^n (x_i - \bar{x}) x_i} \quad (17)$$

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:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \implies \bar{x} \sum_{i=1}^n (x_i - \bar{x}) = 0 \implies \sum_{i=1}^n \bar{x} (x_i - \bar{x}) = 0 \quad (18)$$

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

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})x_i - \sum_{i=1}^n \bar{x}(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})x_i - \sum_{i=1}^n \bar{x}(x_i - \bar{x})} \quad (19)$$

Внесём всё под один знак суммы:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (20)$$

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

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

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

Получаем:

```
> beta_1 <- (sum((df[,2] - meany)*(df[,1] - meanx)))/
+ (sum((df[,1] - meanx)^2))
> beta_0 <- meany - beta_1 * meanx
> beta_1
[1] -1.006984
> beta_0
[1] -1.997074
```

Или

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

Call:

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

Residuals:

Min	1Q	Median	3Q	Max
-3.4976	-0.5606	-0.1048	0.7340	2.2569

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.99707	0.11926	-16.75	<2e-16 ***
df[, 1]	-1.00698	0.08379	-12.02	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9912 on 138 degrees of freedom

Multiple R-squared: 0.5114, Adjusted R-squared: 0.5078

F-statistic: 144.4 on 1 and 138 DF, p-value: < 2.2e-16

Видим, что результаты совпадают.

Можем записать полученное уравнение линейной регрессии:

$$Y = -1.00698 - 1.99707 \cdot X$$

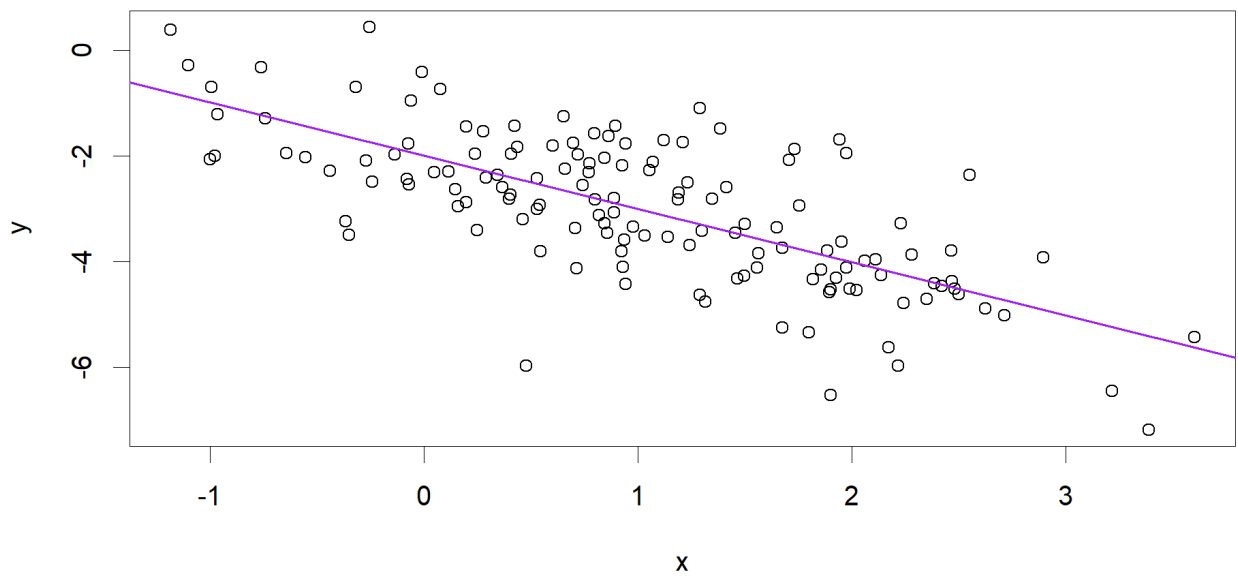
Остаточная дисперсия равна

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

```
> sigma_ost <- (sum((df[,2] - beta_1 *  
+ df[,1] - beta_0)^2))/(n-2)  
> sigma_ost  
[1] 0.9824018
```

Построим совмещённые графики диаграммы рассеяния и линии регрессии:

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





## Задание 3

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

### 1. Доверительные интервал для коэффициента корреляции компонент $r$

Рассмотрим следующие статистики:

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

```
> mu20 <- 1/n * sum((df[,1] - meanx)^2)
> mu20
[1] 0.9994035
> mu02 <- 1/n * sum((df[,2] - meany)^2)
> mu02
[1] 1.98178
> mu11 <- 1/n * sum((df[,1] - meanx)*(df[,2] - meany))
> mu11
[1] -1.006384
```

Тогда выборочный коэффициент корреляции равен:

$$r_B = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}$$

```
> rv <- mu11/sqrt(mu20*mu02)
> rv
[1] -0.7150978
```

Рассмотрим статистику  $z = \frac{1}{2} \ln \frac{1+r_B}{1-r_B} = \operatorname{arctanh} r_B$ , 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  $r$ :

$$1 - \alpha = P\left(-u_{1-\frac{\alpha}{2}} < (z - a_z)\sqrt{n-3} < u_{1-\frac{\alpha}{2}}\right) = P\left(z - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}} < a_z < z + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right) =$$

$$= P\left(\operatorname{arctanh} r_B - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}} < \operatorname{arctanh} r + \frac{r}{2(n-1)} < \operatorname{arctanh} r_B + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right)$$

Итого:

$$P\left(\tanh\left(\operatorname{arctanh} r_B - \frac{r_B}{2(n-1)} - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right) < r < \tanh\left(\operatorname{arctanh} r_B - \frac{r_B}{2(n-1)} + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right)\right)$$

Получаем:

```
> CIr_lower_bound <- tanh(atanh(rv) - rv/(2*(n-1))
+                          - qnorm(1-alpha/2)/sqrt(n-3))
> CIr_lower_bound
[1] -0.7865877
> CIr_upper_bound <- tanh(atanh(rv) - rv/(2*(n-1))
+                          + qnorm(1-alpha/2)/sqrt(n-3))
> CIr_upper_bound
[1] -0.6215435
```

Получаем доверительный интервал:

$$-0.7865877 \leq r \leq -0.6215435$$

Истинное значение коэффициента корреляции

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

попадает в полученный интервал.

## 2. Доверительный интервал для коэффициентов регрессии $\beta_0$ и $\beta_1$

Запишем матрицу базисных функций:

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

$$F^T F = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum_{k=1}^n X_k \\ \sum_{k=1}^n X_k & \sum_{k=1}^n X_k^2 \end{pmatrix} = \begin{pmatrix} 140 & 141.8275 \\ 141.8275 & 283.5953 \end{pmatrix}$$

$$(F^T F)^{-1} = \frac{1}{n \sum_{k=1}^n X_k^2 - \left( \sum_{k=1}^n X_k \right)^2} \begin{pmatrix} \sum_{k=1}^n X_k^2 & -\sum_{k=1}^n X_k \\ -\sum_{k=1}^n X_k & n \end{pmatrix} = \begin{pmatrix} 0.014477 & -0.007240 \\ -0.007240 & 0.007147 \end{pmatrix}$$

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

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

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

Доверительный интервал для коэффициента регрессии  $\beta_0$ :

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

```
> 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] -2.232888
> CIbeta0_upper_bound <- beta_0 +
+   qt(1-alpha/2, n-2) * sqrt(sigma_ost) * sqrt(M_obr[1,1])
> CIbeta0_upper_bound
[1] -1.76126
```

Получаем доверительный интервал для  $\beta_0$ :

$$-2.232888 \leq \beta_0 \leq -1.76126$$

Доверительный интервал для коэффициента регрессии  $\beta_1$ :

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

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

```
[1] -1.17267
> CIbeta1_upper_bound <- beta_1 +
+   (qt(1-alpha/2, n-2) * sqrt(sigma_ost)) * sqrt(M_obr[2,2])
> CIbeta1_upper_bound
[1] -0.8412993
```

Получаем доверительный интервал для  $\beta_1$ :

$$-1.17267 \leq \beta_1 \leq -0.8412993$$

Или

```
> confint(lm1)
                2.5 %      97.5 %
(Intercept) -2.232888 -1.7612602
df[, 1]      -1.172670 -0.8412993
```

### 3. Доверительный интервал для дисперсии $\delta^2$

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

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

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

Следовательно:

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

```
> CIdelta_ost_upper_boud<- (n-2)*delta_ost /
+   (qchisq(alpha/2, n-2))
> CIdelta_ost_upper_boud
[1] 1.26263
> CIdelta_ost_lower_boud <- (n-2)*delta_ost /
+   (qchisq(1-alpha/2, n-2))
> CIdelta_ost_lower_boud
[1] 0.7863207
```

Получаем доверительный интервал для  $\delta^2$ :

$$0.7863207 \leq \delta^2 \leq 1.26263$$