

В таблице 1 представлена выборка из N=200 независимых наблюдений преобразованного к числовому виду напряжения выхода генератора теплового шума, изменение распределения которого позволяет определять некоторые особенности пациента.

1. Проверить выход генератора шума на нормальность с помощью критерия  $\chi^2$  с уровнем значимости  $\alpha = 0,03$ ;
2. Построить доверительный интервал для среднего значения с уровнем доверия  $1-\alpha = 0,97$ ;
3. Найти размер выборки, позволяющий построить критерий проверки гипотезы  $\mu_x = 0,262$  с уровнем значимости  $\alpha = 0,03$  и ошибкой второго рода  $\alpha = \beta$  для выявления 10% отклонений от гипотетического значения. Построить область принятия гипотезы.

**Таблица 1** – Выборка значений напряжения выхода генератора теплового шума

-7,6	-3,8	-2,5	-1,6	-0,7	0,2	1,1	2,0	3,4	4,6
-6,9	-3,8	-2,5	-1,6	-0,7	0,2	1,1	2,1	3,5	4,8
-6,6	-3,7	-2,4	-1,6	-0,6	0,2	1,2	2,3	3,5	4,8
-6,4	-3,6	-2,3	-1,5	-0,6	0,3	1,2	2,3	3,6	4,9
-6,2	-3,5	-2,3	-1,5	-0,6	0,3	1,3	2,3	3,6	5,0
-6,1	-3,4	-2,3	-1,4	-0,5	0,3	1,3	2,4	3,6	5,2
-6,0	-3,4	-2,2	-1,4	-0,4	0,4	1,3	2,4	3,7	5,3
-5,7	-3,4	-2,2	-1,2	-0,4	0,4	1,4	2,5	3,7	5,4
-5,6	-3,3	-2,1	-1,2	-0,4	0,5	1,5	2,5	3,7	5,6
-5,5	-3,2	-2,1	-1,2	-0,3	0,5	1,5	2,6	3,7	5,8
-5,4	-3,2	-2,0	-1,1	-0,3	0,6	1,6	2,6	3,8	6,1
-5,2	-3,1	-2,0	-1,1	-0,2	0,6	1,6	2,6	3,8	6,3
-4,8	-3,0	-1,9	-1,0	-0,2	0,7	1,6	2,7	3,9	6,3
-4,6	-3,0	-1,9	-1,0	-0,2	0,8	1,7	2,8	4,0	6,5
-4,4	-2,9	-1,8	-1,0	-0,1	0,9	1,8	2,8	4,2	6,9
-4,4	-2,9	-1,8	-0,9	0,0	0,9	1,8	2,9	4,2	7,1
-4,3	-2,9	-1,8	-0,9	0,0	1,0	1,8	3,1	4,3	7,2
-4,1	-2,7	-1,7	-0,8	0,1	1,0	1,9	3,2	4,3	7,4
-4,0	-2,6	-1,7	-0,8	0,1	1,1	1,9	3,2	4,4	7,9
-3,8	-2,6	-1,6	-0,7	0,2	1,1	2,0	3,3	4,4	9,0

## Решение

1) На уровне значимости 0,03 проверим нулевую гипотезу  $H_0$  о нормальном распределении данной выборки при помощи критерия согласия Пирсона. Для этого разобьем имеющуюся выборку на интервалы с шагом, вычисляемым по формуле:

$$h = 0,4 * Sx,$$

где  $Sx$  – среднее квадратичное отклонение.

Формулы для выборочного среднего значения и для выборочной дисперсии соответственно:

$$\bar{x} = \sum \frac{x_i}{N},$$

$$S_x^2 = \frac{1}{N-1} \sum (\bar{x} - x_i)^2,$$

где  $x_i$  – элементы данной выборки,  $N$  – объем данной выборки

Таким образом имеем, что при объеме выборки в 200 значений,  $\bar{x} = 0.263$ ,  $S_x^2 = 3.3128$ , тогда  $h = 1.3251$ . Зная шаг разбиения, подсчитаем количество интервалов:

$$K = \frac{x_{max} - x_{min}}{h} \approx 13,$$

Данные по разделению выборки на интервалы приведены в таблице 2

**Таблица 2** – Данные после разбиения на интервалы

Номер интервала	Середина интервала ( $x_i$ )	Экспериментальная частота ( $n_i$ )
1	-6,94	4
2	-5,61	8
3	-4,29	11
4	-2,96	23
5	-1,64	29
6	-0,31	31
7	1,01	27
8	2,34	23
9	3,66	22
10	4,99	11
11	6,31	6
12	7,64	4
13	8,96	1
Сумма:		200

Поскольку последний интервал содержит в себе одну СВ, необходимо объединить его с предыдущим столбцом, чтобы все интервалы содержали в себе не меньше трех СВ, тогда значение К принимаем равным 9.

Соответственно, зная количество интервалов, определим число степеней свободы для нормального распределения:  $g = K - 1 - 2 = 9$ . Исходя из этого можем подобрать табличное значение распределения хи-квадрат:  $\chi^2_{0,03\ 9} = 18,48$ . Гипотеза о нормальном распределении будет принята в случае, если экспериментальное значение окажется меньше представленного табличного.

Промежуточные данные для подсчета теоретических частот и наблюдаемых значений хи-квадрат-значений представлены в таблице 3:

**Таблица 3 – Промежуточные данные для пересчета**

Номер интервала	Середина интервала ( $x_i$ )	Экспериментальная частота ( $n_i$ )	$x_i * n_i$	$x_i^2 * n_i$
1	-6,94	4	-27,75	192,51
2	-5,61	8	-44,90	251,98
3	-4,29	11	-47,16	202,18
4	-2,96	23	-68,13	201,79
5	-1,64	29	-47,47	77,70
6	-0,31	31	-9,66	3,01
7	1,01	27	27,36	27,73
8	2,34	23	53,79	125,78
9	3,66	22	80,60	295,29
10	4,99	11	54,88	273,77
11	6,31	6	37,88	239,19
12	8,30	5	41,51	344,85
<b>Сумма:</b>		<b>200</b>	<b>50,95</b>	<b>2235,5</b>

Для проверки гипотезы необходимо найти теоретические частоты, для чего определим выборочное среднее:

$$\bar{X}_{в.ср} = \frac{\sum x_i n_i}{n} = \frac{50,95}{200} = 0,2547$$

и выборочное стандартное отклонение:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum (\bar{x}_{в.ср} - x_i)^2} = 3,3419$$

Теоретические частоты рассчитываются по формуле:

$$n'_i = \frac{h * N}{\sigma_x} f(z_i),$$

где  $f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$  – функция распределения Гаусса,  $z_i = \frac{x_i - \bar{x}_{в.ср.}}{\sigma_x}$  – z-оценка.

Значения хи-квадрат рассчитываются по следующей формуле:

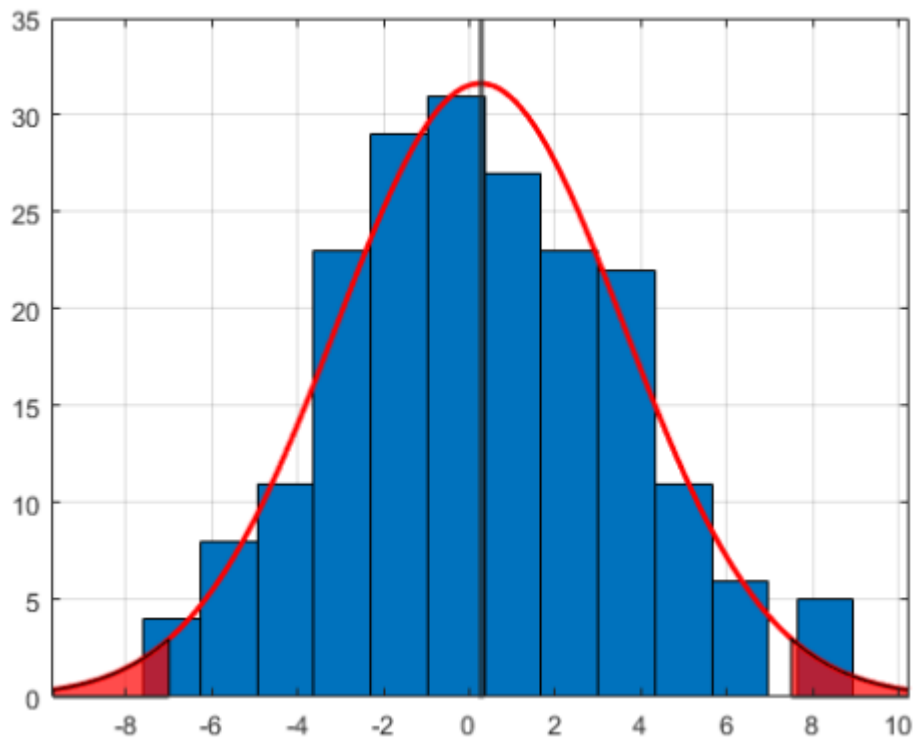
$$\chi^2 = \sum \frac{(n_i - n'_i)^2}{n'_i}$$

Результат вычислений приведен в таблице 4:

**Таблица 4 – Результаты вычисления теоретических частот и значений хи-квадрат**

i	Середина интервала (xi)	Экспериментальная частота (ni)	xi*ni	zi	f(zi)	Теоретические частоты (n1)	Хи-квадрат
1	-6,94	4,00	-27,75	-2,15	0,04	3,12	0,25
2	-5,61	8,00	-44,90	-1,76	0,09	6,78	0,22
3	-4,29	11,00	-47,16	-1,36	0,16	12,56	0,19
4	-2,96	23,00	-68,13	-0,96	0,25	19,91	0,48
5	-1,64	29,00	-47,47	-0,57	0,34	26,95	0,16
6	-0,31	31,00	-9,66	-0,17	0,39	31,19	0,00
7	1,01	27,00	27,36	0,23	0,39	30,83	0,48
8	2,34	23,00	53,79	0,62	0,33	26,05	0,36
9	3,66	22,00	80,60	1,02	0,24	18,80	0,54
10	4,99	11,00	54,88	1,42	0,15	11,60	0,03
11	6,31	6,00	37,88	1,81	0,08	6,11	0,00
12	8,30	5,00	41,51	2,41	0,02	1,74	6,09
<b>Сумма:</b>							<b>8,80</b>

Гистограмма экспериментальных частот и теоретическая кривая нормального распределения с уровнем значимости 0,03 представлены на рисунке 1:



**Рисунок 1** – Гистограмма экспериментальных частот и теоретическая кривая

Итак поскольку экспериментальное значение  $\chi^2 = 8.80 < 18.48 = \chi^2$  теоретического, то нулевая гипотеза  $H_0$  принимается.

2) Построим доверительный интервал для среднего значения ( $\alpha = 0.03$ ,  $\alpha/2 = 0.015$ ,  $1 - \alpha = 0.97$ ):

$$Prob \left[ \mu_x - \frac{\sigma_x * Z_{\alpha/2}}{\sqrt{N}} \leq \bar{X} \leq \mu_x + \frac{\sigma_x * Z_{\alpha/2}}{\sqrt{N}} \right] = 1 - \alpha$$

Подставим вычисленные и найденные по таблице z-распределения ( $Z_{0.015} = 2.17$ ) значения и получим искомый доверительный интервал:

$$Prob[-0.258 \leq \bar{X} \leq 0.767] = 0.97$$

3) Найдем размер выборки, позволяющий построить критерий проверки гипотезы  $\mu_x = 0.262$  ( $\alpha = 0.03$ ,  $\alpha = \beta$ ) для выявления 10% отклонений от гипотетического значения в соответствии с формулой:

$$N = \left( \frac{\sigma_x (z_{\alpha} + z_{\beta})}{d} \right)^2,$$

где  $d = 0.1 * \mu_x$ , а  $Z_{\beta} = 1.88$ .

Тогда размером выборки, позволяющим построить критерий проверки поставленной гипотезы, является  $N = 267289$ .

Определим область принятия гипотезы:

$$\mu_x - \frac{\sigma_x * Z_{\frac{\alpha}{2}}}{\sqrt{N}} \leq \bar{X} \leq \mu_x + \frac{\sigma_x * Z_{\frac{\alpha}{2}}}{\sqrt{N}},$$

Подставив имеющиеся значения, имеем:

$$0.248 \leq \bar{X} \leq 0.276$$

Вычисления проводились в Matlab.

```

Variant = 3; alph = 0.01*Variant;
z_alph = icdf('Normal',1-alph/2,0,1); % вызов табличного значения для z-распределения половины
x = [-7.6 -6.9 -6.6 -6.4 -6.2 -6.1 -6 -5.7 -5.6 -5.5 -5.4 -5.2 -4.8 ...
    -4.6 -4.4 -4.4 -4.3 -4.1 -4.0 -3.8 -3.8 -3.8 -3.7 -3.6 -3.5 ...
    -3.4 -3.4 -3.4 -3.3 -3.2 -3.2 -3.1 -3 -3 -2.9 -2.9 -2.9 -2.7 ...
    -2.6 -2.6 -2.5 -2.5 -2.4 -2.3 -2.3 -2.3 -2.2 -2.2 -2.1 -2.1 ...
    -2.0 -2.0 -1.9 -1.9 -1.8 -1.8 -1.8 -1.7 -1.7 -1.6 -1.6 -1.6 ...
    -1.6 -1.5 -1.5 -1.4 -1.4 -1.2 -1.2 -1.2 -1.1 -1.1 -1 -1 -1 ...
    -0.9 -0.9 -0.8 -0.8 -0.7 -0.7 -0.7 -0.6 -0.6 -0.6 -0.5 -0.4 ...
    -0.4 -0.4 -0.3 -0.3 -0.2 -0.2 -0.2 -0.1 0 0 0.1 0.1 0.2 0.2 ...
    0.2 0.2 0.3 0.3 0.3 0.4 0.4 0.5 0.5 0.6 0.6 0.7 0.8 0.9 0.9 ...
    1 1 1.1 1.1 1.1 1.1 1.2 1.2 1.3 1.3 1.3 1.4 1.5 1.5 1.6 1.6 ...
    1.6 1.7 1.8 1.8 1.8 1.9 1.9 2 2 2.1 2.3 2.3 2.3 2.4 2.4 2.5 ...
    2.5 2.6 2.6 2.6 2.7 2.8 2.8 2.9 3.1 3.2 3.2 3.3 3.4 3.5 3.5 ...
    3.6 3.6 3.6 3.7 3.7 3.7 3.7 3.8 3.8 3.9 4 4.2 4.2 4.3 4.3 ...
    4.4 4.4 4.6 4.8 4.8 4.9 5 5.2 5.3 5.4 5.6 5.8 6.1 6.3 6.3 ...
    6.5 6.9 7.1 7.2 7.4 7.9 9];

m = mean(x);
s = std(x);
step = 0.4*std(x); % шаг интервалов
K = round((max(x) - min(x))/step); % количество интервалов
column_left = min(x):step:min(x)+(K-1)*step; % левые границы интервалов
column_right = column_left + step; % правые границы интервалов
column_centers = column_left + step/2; % центры интервалов
rows = 1:K;
% подсчет экспериментальных частот для каждого интервала
pls = 0;
for i = 1:K
    if i == K
        rows(i) = length(x(x >= min(x)+(i-1)*step & x <= min(x)+i*step));
    else
        rows(i) = length(x(x >= min(x)+(i-1)*step & x < min(x)+i*step));
        rows(i) = rows(i) + pls;
        pls = 0;
    end
    if i < K/2 && rows(i) < 3
        column_centers(i+1) = (column_centers(i) + column_centers(i+1))/2;
        pls = rows(i)
    elseif i >= K/2 && rows(i) < 3
        column_centers(i-1) = (column_centers(i) + column_centers(i-1))/2;
        rows(i-1) = rows(i-1) + rows(i);
    end
end
column_centers = column_centers(rows >= 3);
rows = rows(rows >= 3);
tbl1 = table(column_centers', rows', (column_centers.*rows)', ((column_centers.^2).*rows)');
tbl1 = [tbl1; table(NaN, sum(tbl1.Var2), sum(tbl1.Var3), sum(tbl1.Var4))];
tbl1.Properties.VariableNames = {'Середина интервала (xi)', 'Экспериментальная частота (ni)', 'xi*ni', 'xi^2*ni'};

```

tbl1 = 13x4 table

	Середина интервала (xi)	Экспериментальная частота (ni)	xi*ni	xi^2*ni
1	-6.9374	4	-27.7497	192.5119

	Середина интервала (xi)	Экспериментальная частота (ni)	xi*ni	xi^2*ni
2	-5.6123	8	-44.8984	251.9830
3	-4.2872	11	-47.1588	202.1772
4	-2.9620	23	-68.1266	201.7926
5	-1.6369	29	-47.4698	77.7028
6	-0.3118	31	-9.6643	3.0129
7	1.0134	27	27.3613	27.7275
8	2.3385	23	53.7859	125.7794
9	3.6637	22	80.6004	295.2920
10	4.9888	11	54.8767	273.7683
11	6.3139	6	37.8836	239.1940
12	8.3016	5	41.5081	344.5853
13	NaN	200	50.9485	2.2355e+03

```

mu = sum(column_centers.*rows)/length(x); % среднее значение интервальной выборки
sigma = sqrt(sum(((mu - column_centers).^2).*rows)/(length(x)-1)); % подсчет выборочного стандарта
z = (column_centers-mu)/sigma; % z-оценки для интервальной выборки
f = exp(-z.^2/2)/sqrt(2*pi); % функция Гаусса
n = f*step*length(x)/sigma; % подсчет теоретических частот

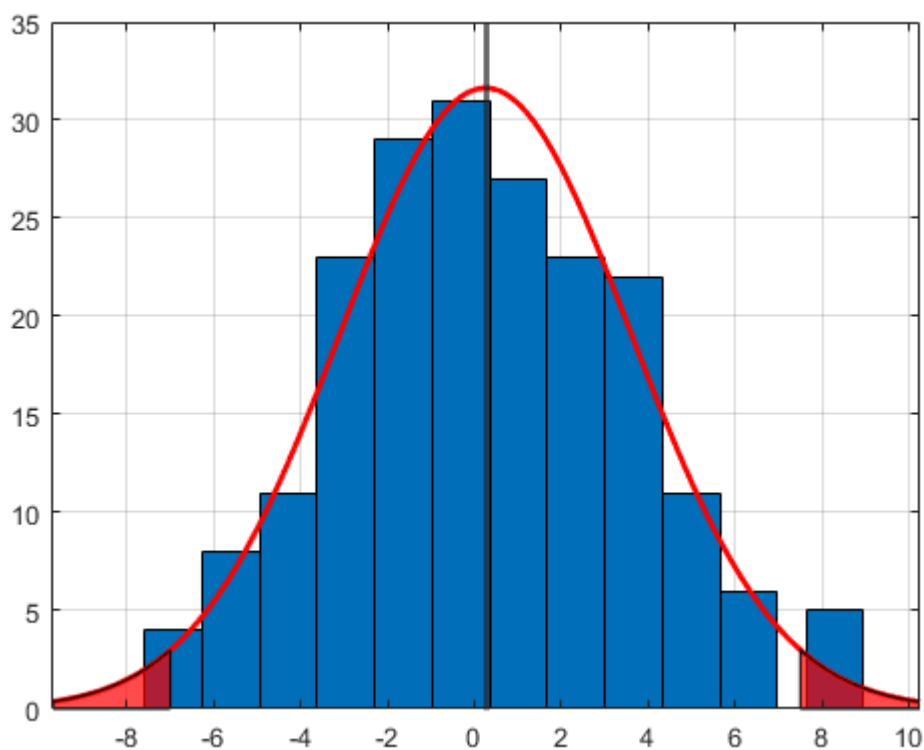
```

```

figure
bar(column_centers,rows,1)
hold on
x1 = round(-10+mu:0.1:10+mu, 1);
z1 = (x1-mu)/sigma;
f1 = exp(-z1.^2/2)/sqrt(2*pi);
n1 = f1*step*length(x)/sigma;
plot(x1,n1, 'r', "LineWidth", 2);
xlim([-10+mu,10+mu])
grid on
mu1 = mu - sigma*z_alph/sqrt(length(x));
mu2 = mu + sigma*z_alph/sqrt(length(x));
xline(mu, 'linewidth', 2)
j1 = find(x1 == round(-z_alph*sigma+mu, 1));
j2 = find(x1 == round(z_alph*sigma+mu, 1));
ar1 = area(x1(1:j1), n1(1:j1)); ar2 = area(x1(j2:end), n1(j2:end));
ar1.FaceColor = 'r'; ar2.FaceColor = 'r';
ar1.FaceAlpha = 0.7; ar2.FaceAlpha = 0.7;
hold off

```





```
hi2 = (rows-n).^2./n; % подсчет экспериментального значения хи-квадрат
sum(hi2)
```

```
ans = 8.7964
```

```
tbl2 = table(column_centers', rows', (column_centers.*rows)', z', f', n', hi2');
tbl2.Properties.VariableNames = {'Середина интервала (xi)', 'Экспериментальная частота (ni)', 'xi*ni', 'zi', 'f(zi)', 'n'}
```

```
tbl2 = 12x7 table
```

	Середина интервала (xi)	Экспериментальная частота (ni)	xi*ni	zi	f(zi)
1	-6.9374	4	-27.7497	-2.1521	0.0394
2	-5.6123	8	-44.8984	-1.7556	0.0854
3	-4.2872	11	-47.1588	-1.3591	0.1584
4	-2.9620	23	-68.1266	-0.9625	0.2510
5	-1.6369	29	-47.4698	-0.5660	0.3399
6	-0.3118	31	-9.6643	-0.1695	0.3933
7	1.0134	27	27.3613	0.2270	0.3888
8	2.3385	23	53.7859	0.6235	0.3285
9	3.6637	22	80.6004	1.0200	0.2371
10	4.9888	11	54.8767	1.4166	0.1463
11	6.3139	6	37.8836	1.8131	0.0771

	Середина интервала (xi)	Экспериментальная частота (ni)	xi*ni	zi	f(zi)
12	8.3016	5	41.5081	2.4078	0.0220

```
writetable(tbl2, 'vbrk.xls')
C = chi2inv(1-alpha,length(rows)-3); % вызов теоретического элемента таблицы хи-квадрат
if sum(hi2) <= C
    disp(['Так как  $\chi^2$  экспериментальный = ' num2str(sum(hi2)) ' < ' num2str(C) ' =  $\chi^2$  теоретический'])
    disp(['Prob[' num2str(mu1) ' <=  $\mu$  <= ' num2str(mu2) '] = ' num2str(1 - alph) ' - доверительность'])
else
    disp('Нулевая гипотеза H0 не подходит')
end
```

Так как  $\chi^2$  экспериментальный = 8.7964 < 18.4796 =  $\chi^2$  теоретический, то нулевая гипотеза H0 подходит  
 Prob[-0.25807 <=  $\mu$  <= 0.76756] = 0.97 - доверительное утверждение для среднего значения

### % Задание номер 3

```
mux = 0.262; bet = alph; d = 0.1*mux; % данные задания
z_bet = icdf('Normal',1-bet,0,1); % вызов табличного значения для z-распределения для ошибки вт
n_neened = round(sigma*(z_alph + z_bet)/d)^2; % размер выборки для проверки новой гипотезы
disp([num2str(n_neened) ' - размер выборки для проверки новой гипотезы'])
```

267289 - размер выборки для проверки новой гипотезы

```
mux1 = mux - sigma*z_alph/sqrt(n_neened); % подсчет левой границы области принятия гипотезы
mux2 = mux + sigma*z_alph/sqrt(n_neened); % подсчет правой границы области принятия гипотезы
disp([num2str(mux1) ' <= X <= ' num2str(mux2) ' - область принятия гипотезы'])
```

0.24797 <= X <= 0.27603 - область принятия гипотезы