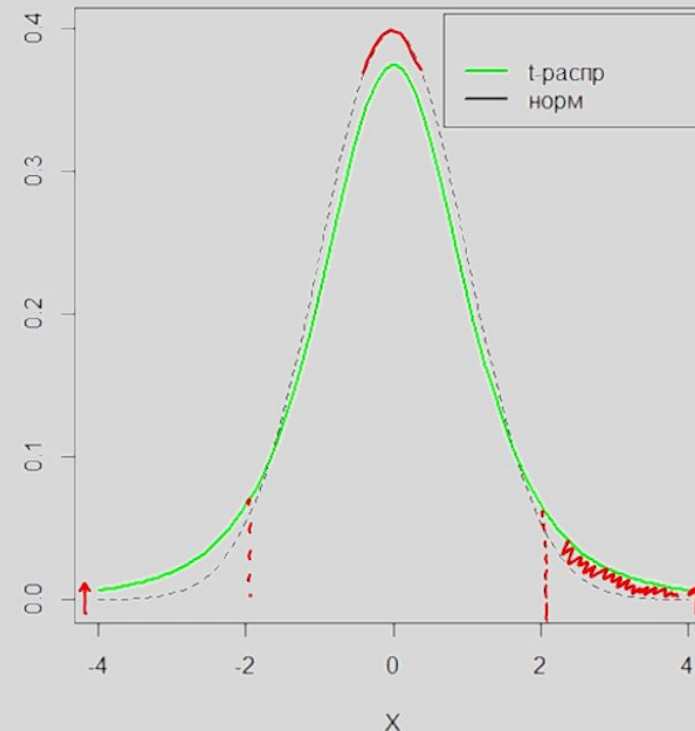


# Распределение Стьюдента

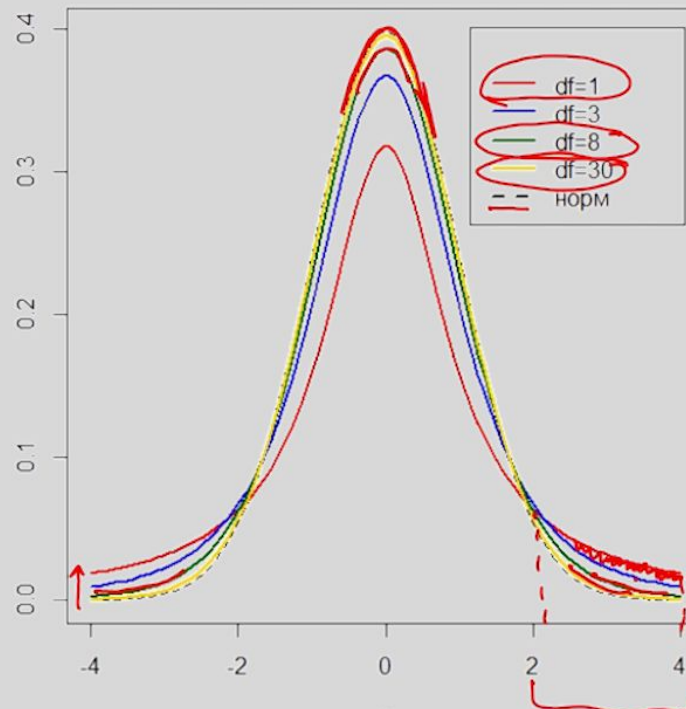
Если число наблюдений невелико и  $\sigma$  неизвестно (почти всегда), используется распределение Стьюдента (t-distribution).

Унимодально и симметрично,  
но: наблюдения с большей  
вероятностью попадают за  
пределы  $\pm 2\sigma$  от  $M$



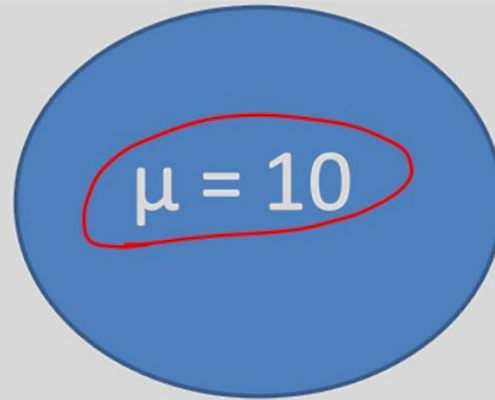
# Распределение Стьюдента

«Форма» распределения определяется числом степеней свободы ( $df = n - 1$ ). С увеличением числа  $df$  распределение стремится к нормальному.



$$n=2 \quad df=2-1=1$$
$$n=9 \quad df=30$$
$$n=31$$

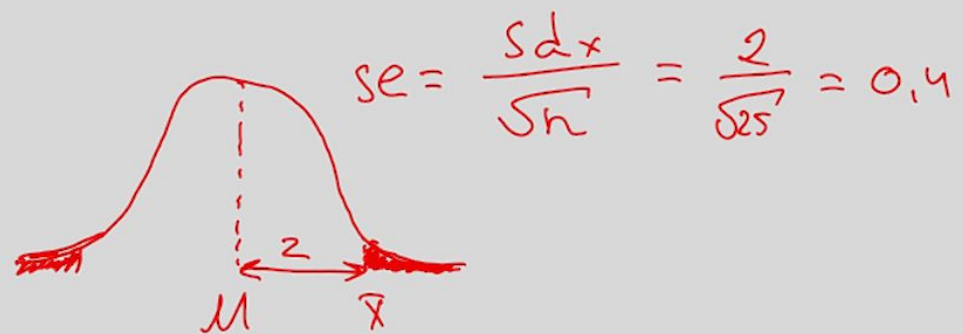




$$\bar{X} = \underline{10,8}$$

$$sd = 2$$

$$\underline{\underline{N = 25}}$$



$$Z = \frac{10,8 - 10}{0,4} = \frac{0,8}{0,4} = 2$$



# Distribution Calculator

Distribution:

t

Degrees of freedom

1 24 50

Model:

 $P(X < a \text{ or } X > b)$ 

Find Area:

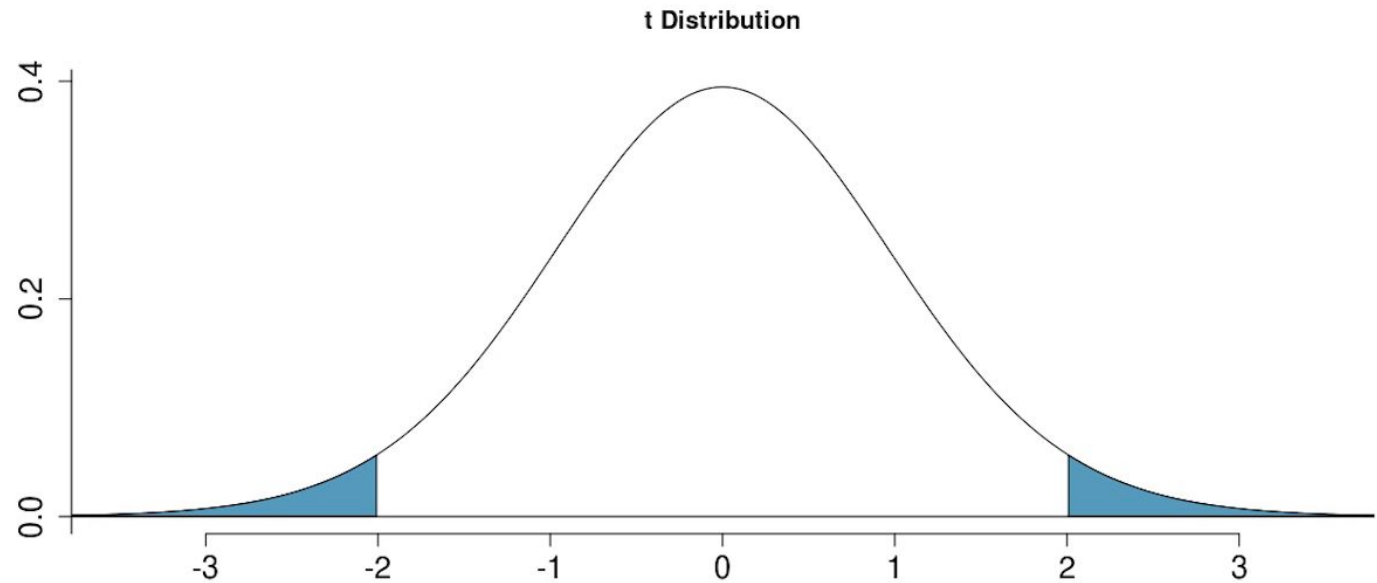
Both Tails

a

-6 -2 6

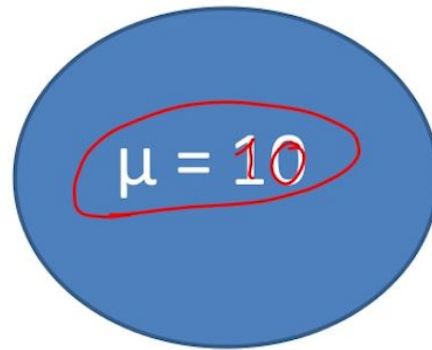
b

-6 2 6

[Rate this app!](#)[View code](#)[Check out other apps](#)[Want to learn more for free?](#)

$$P(X < -2 \text{ or } X > 2) = 0.0569$$

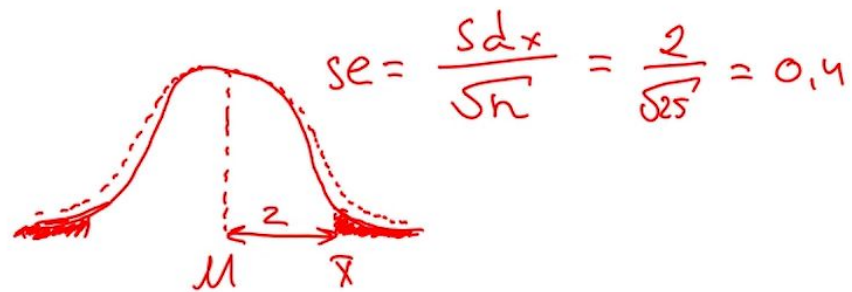




$$\bar{X} = \underline{10,8}$$

$$sd = 2$$

$$\underline{N = 25}$$



$$Z = \frac{10,8 - 10}{0,4} = \frac{0,8}{0,4} = 2$$

$$p < 0,05 \quad \text{No } (?)$$

$$t = \frac{\bar{X} - \mu}{\frac{sd}{\sqrt{n}}} = 2 \quad df = 24 \quad p = 0,056 \quad H_0 ?$$



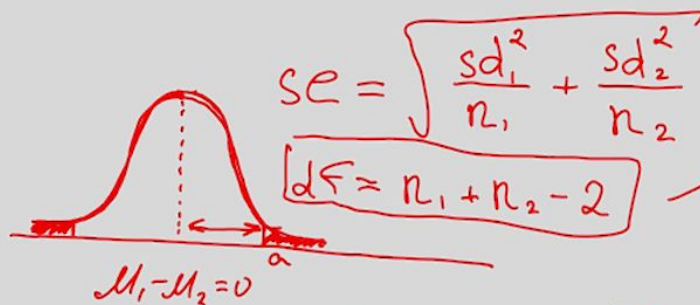
# Критерий t - Стьюдента

$$\begin{array}{cc} \checkmark & \checkmark \\ \overline{X}_1 & \overline{X}_2 \\ sd_1 & sd_2 \\ n_1 & n_2 \end{array}$$

$$\begin{array}{l} \boxed{H_0} \quad \mu_1 = \mu_2 \\ H_1 \quad \mu_1 \neq \mu_2 \end{array}$$

$$\boxed{\overline{X}_1 - \overline{X}_2} = \underline{\underline{a}}$$

$$\boxed{t} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}}$$



$$\boxed{df = n_1 + n_2 - 2}$$

$$df_1 = n_1 - 1 \quad df_2 = n_2 - 1$$

P



# Критерий t - Стьюдента

Процесс денатурации ДНК представляет разрушение водородных связей между двумя цепями этой молекулы и очень сильно зависит от температуры, с которой мы воздействуем на молекулу.

При сравнении двух видов между собой в исследовании были получены следующие различия в средней температуре плавления ДНК:

	M	SD	N
Вид №1	89,9	11,3	20
Вид №2	80,7	11,7	20

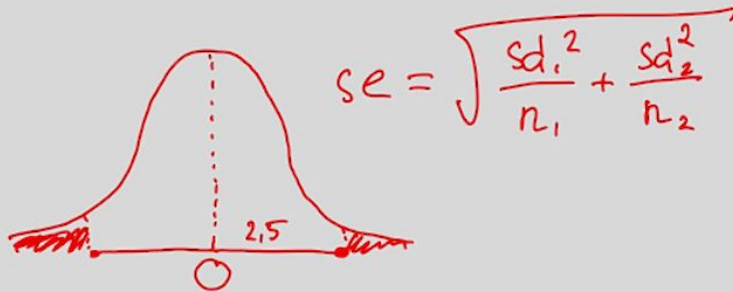
№1	84,7	105,0	98,9	97,9	108,7	81,3	99,4	89,4	93,0	119,3	99,2	99,4	97,1	112,4	99,8	94,7	114,0	95,1	115,5	111,5
№2	57,2	68,6	104,4	95,1	89,9	70,8	83,5	60,1	75,7	102,0	69,0	79,6	68,9	98,6	76,0	74,8	56,0	55,6	69,4	59,5





## Критерий t - Стьюдента

	$M_x$	SD	N
Вид №1	89,9	11,3	20
Вид №2	80,7	11,7	20



$$t = \frac{89,9 - 80,7}{\sqrt{\frac{11,3^2}{20} + \frac{11,7^2}{20}}} \approx 2,5$$

$$df = n_1 + n_2 - 2 = 40 - 2 = 38$$

$$H_0 \quad M_{1\text{ вид}} = M_{2\text{ вид}}$$

$$H_1 \quad M_{1\text{ вид}} \neq M_{2\text{ вид}}$$



# on Calculator

38

50

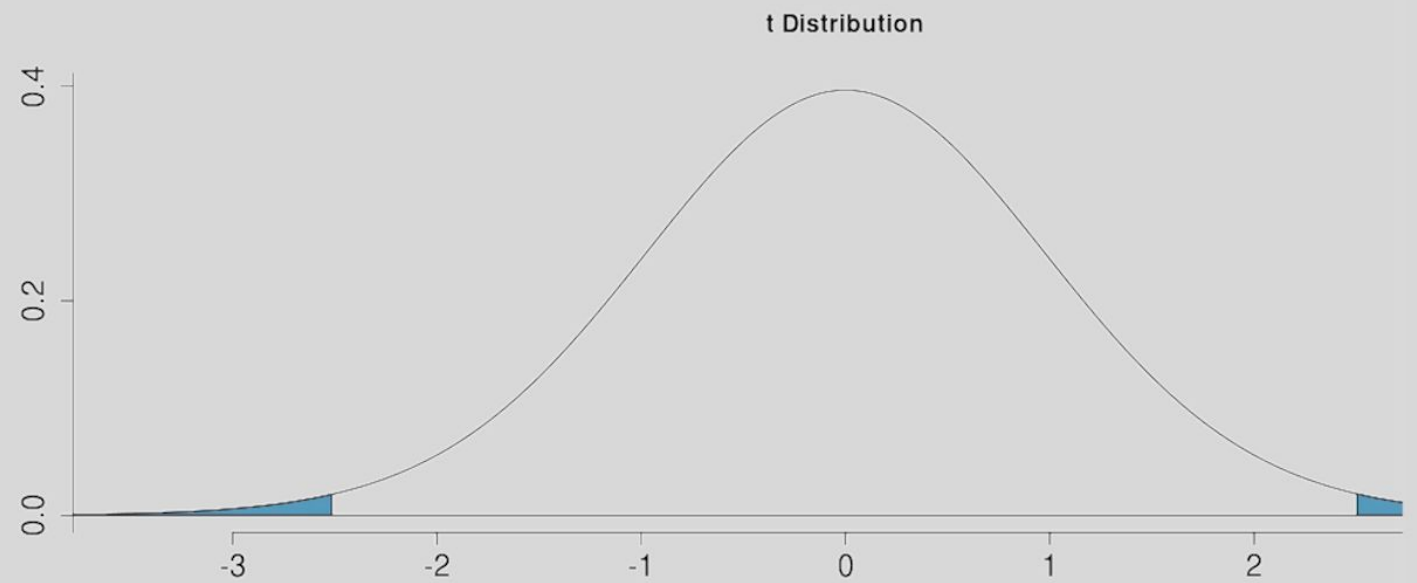
6

6

2.5

6

free?



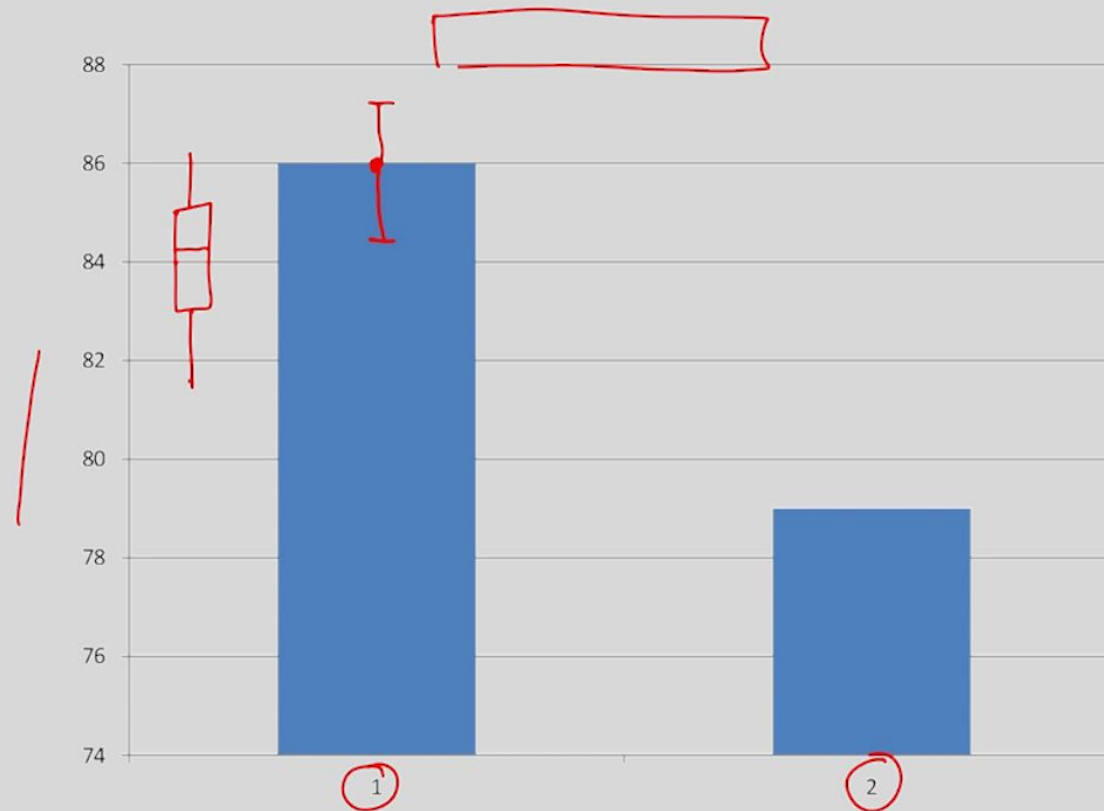
$$P(X < -2.51 \text{ or } X > 2.5) = 0.0167$$

4

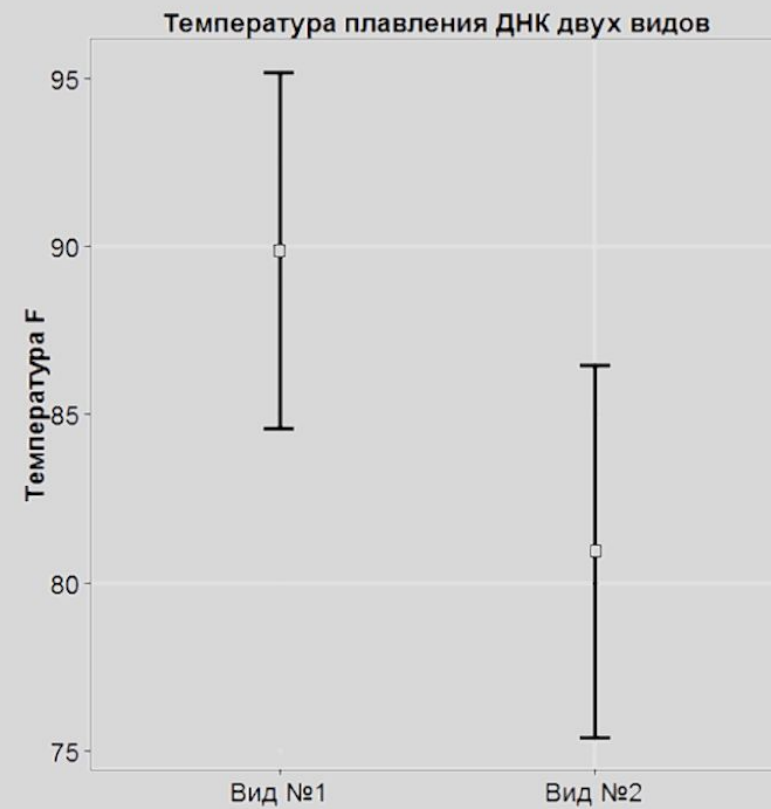
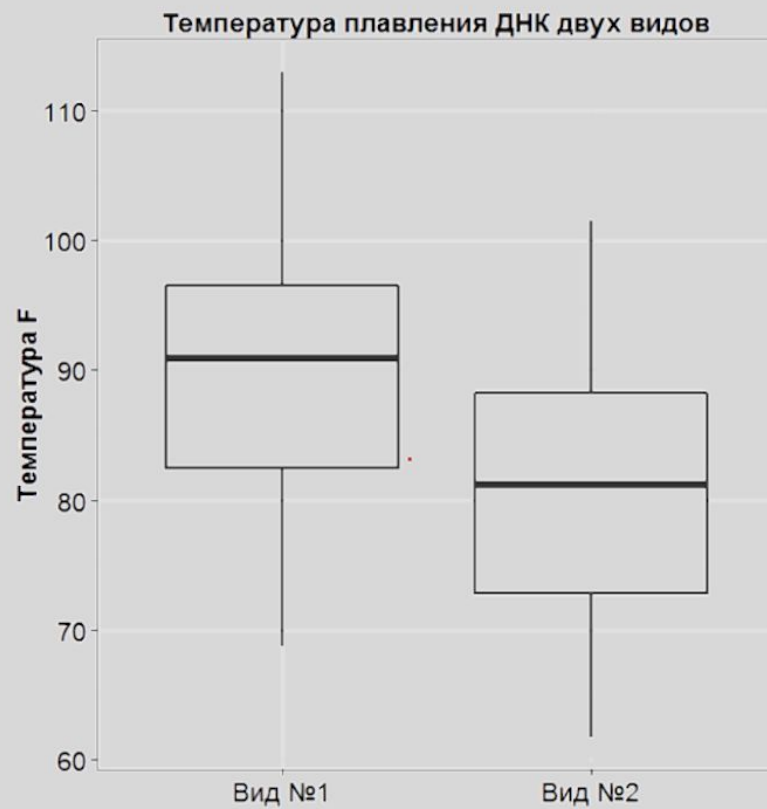


# Строим графики

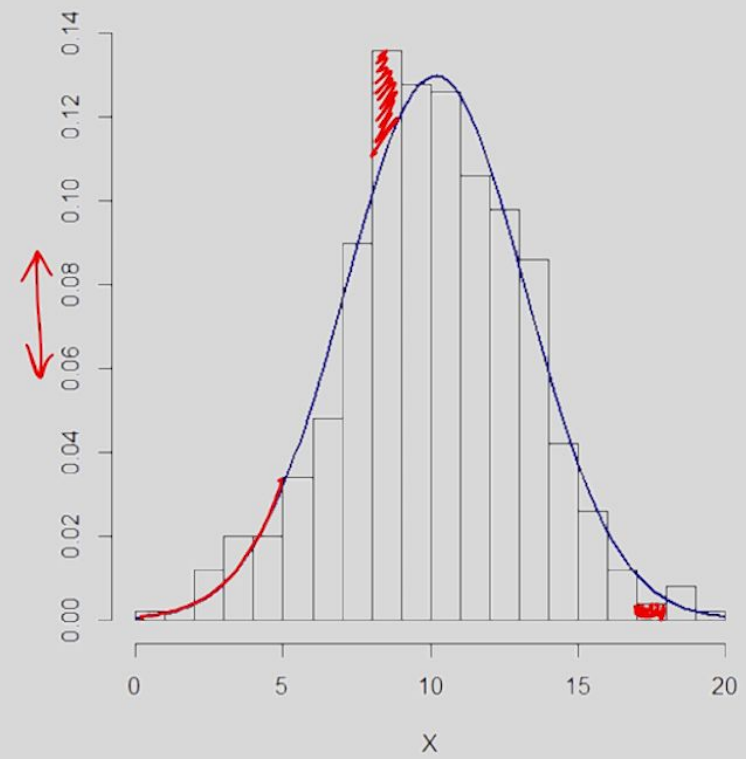
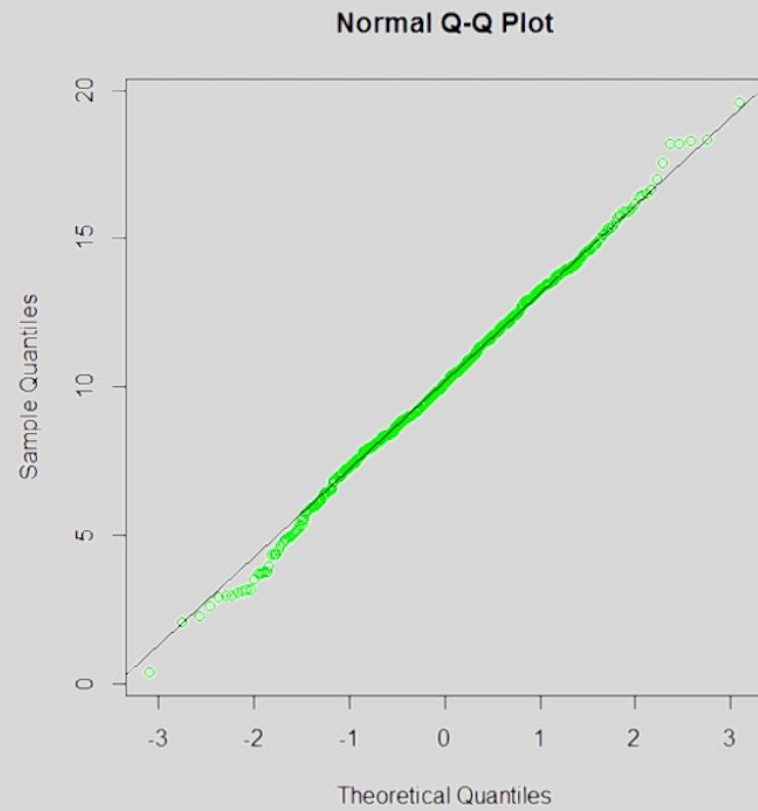
Как не надо делать :)



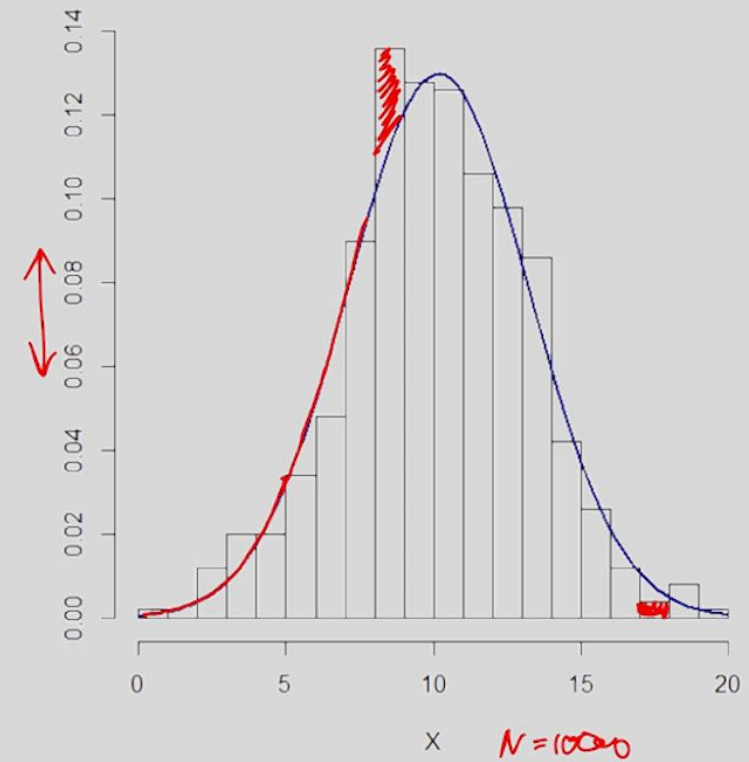
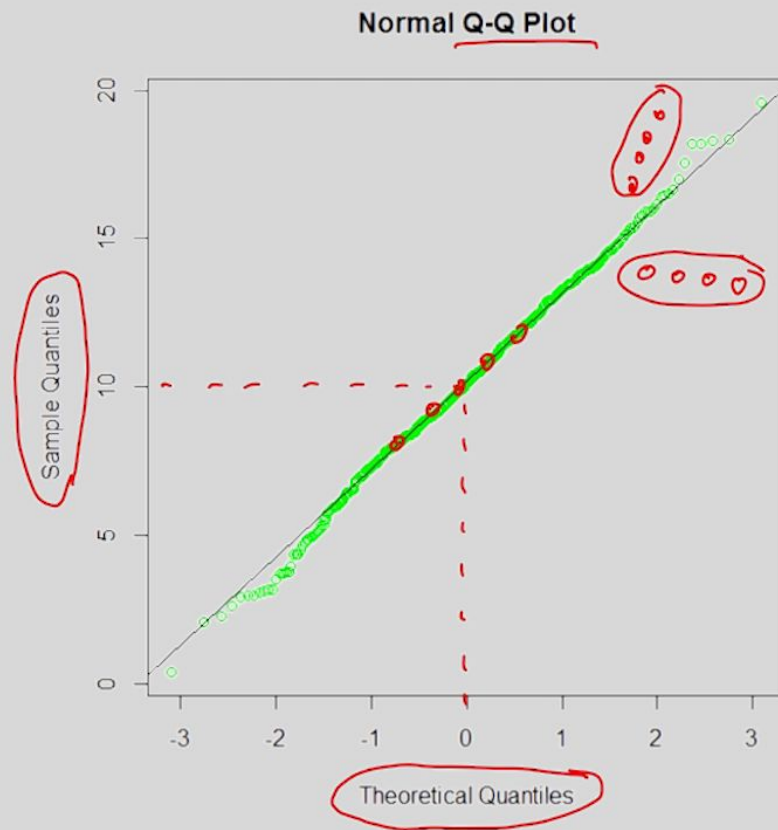
# Строим графики



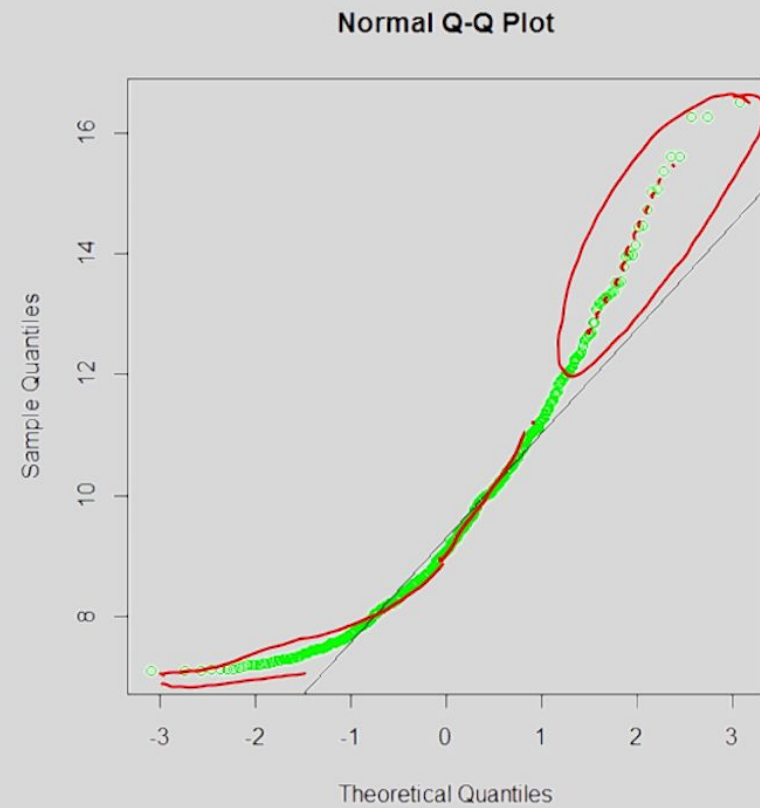
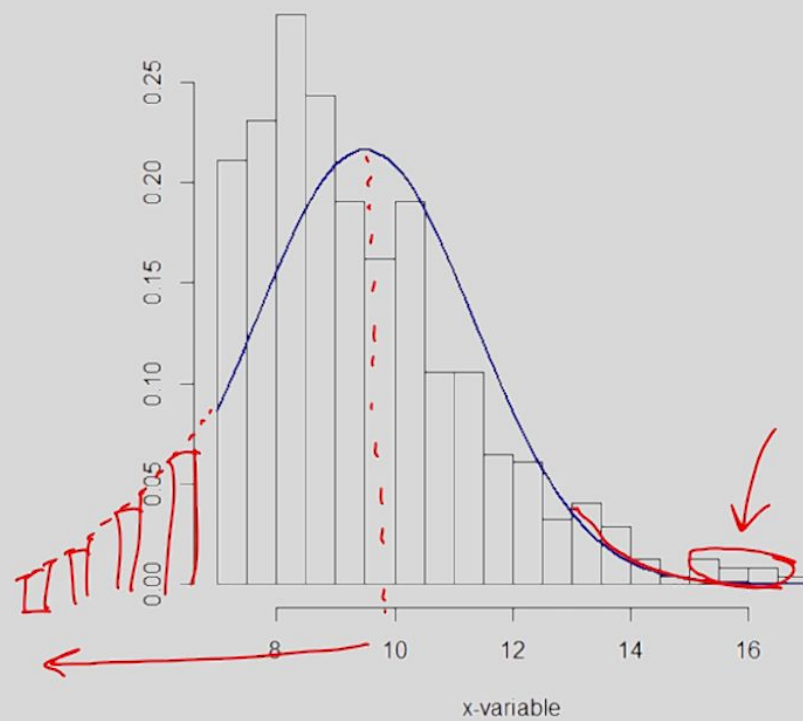
# QQ Plot



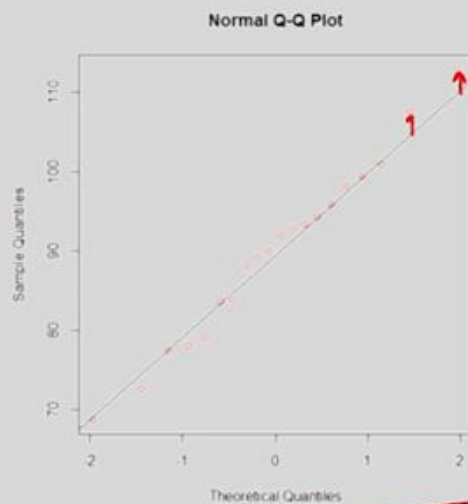
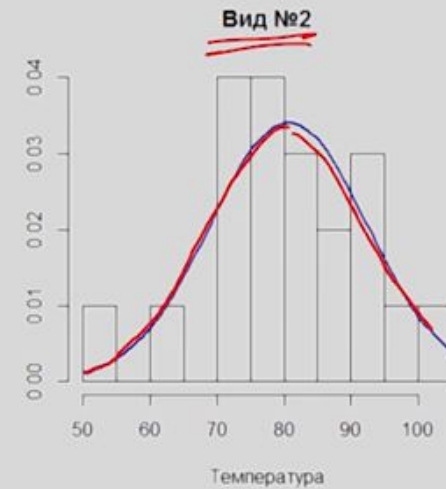
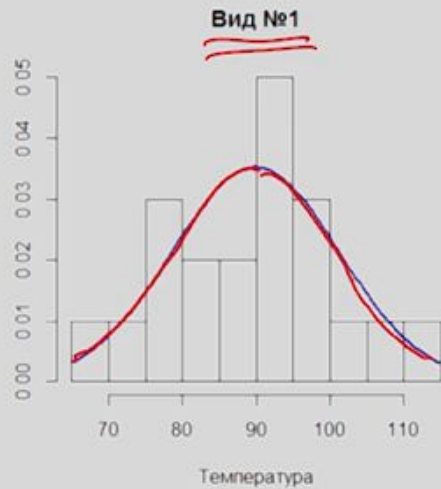
# QQ Plot



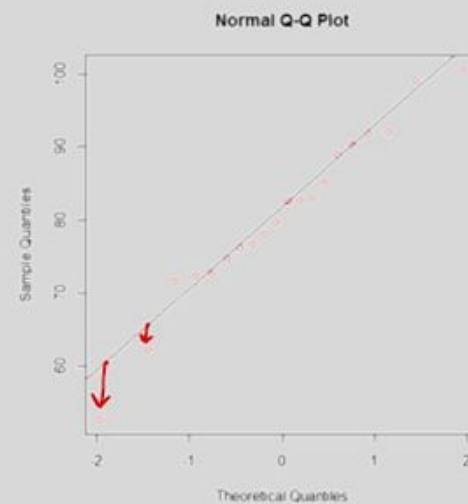
# QQ Plot



# Проверка распределения на нормальность



Shapiro-Wilk normality test,  $p=0.98$

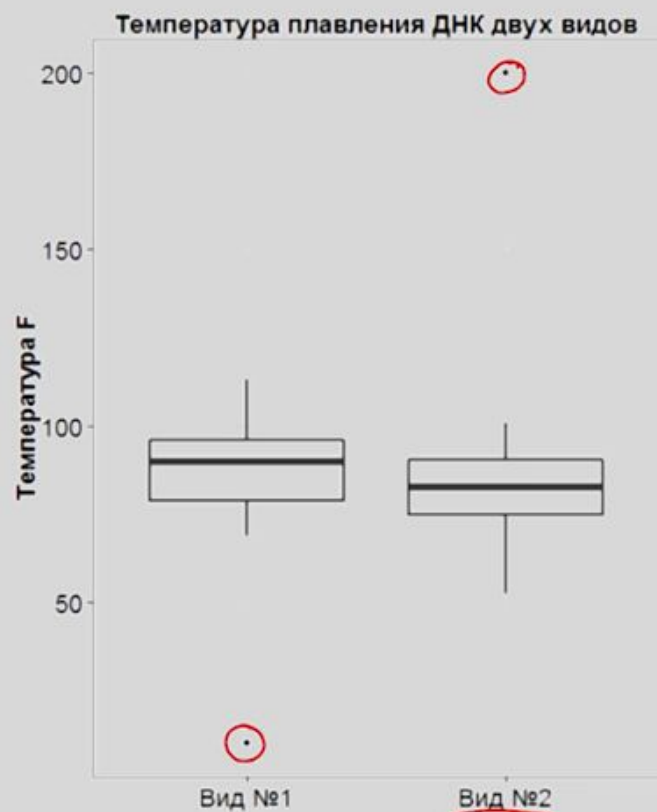


Shapiro-Wilk normality test,  $p=0.78$





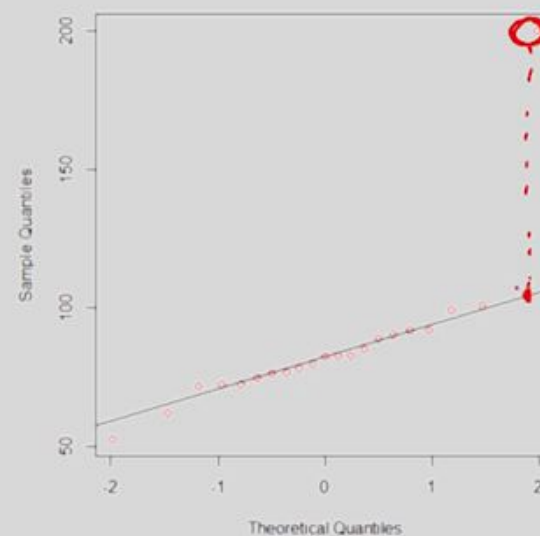
# Берегись выбросов!!!



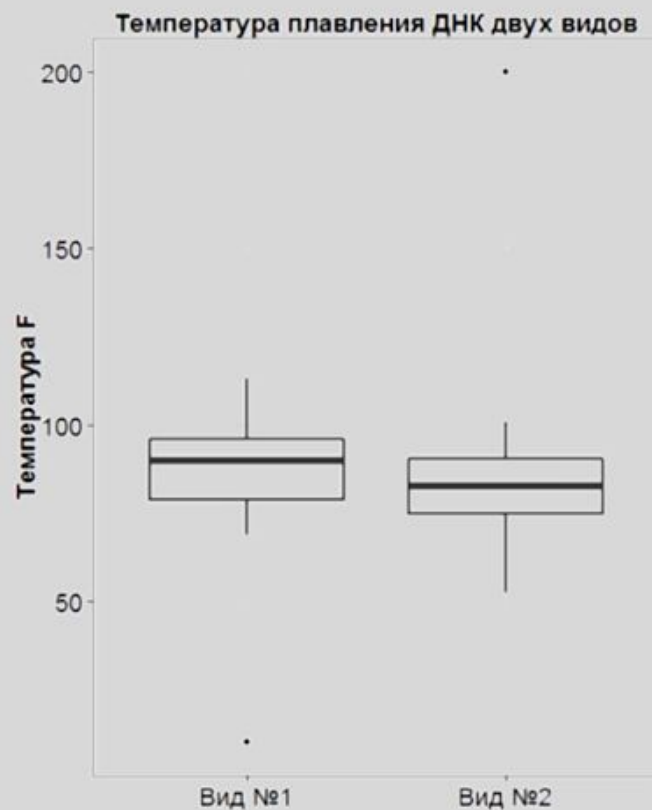
Добавим по одному наблюдению в каждую выборку и в результате:

$t = ?$

$p = ?$



# Берегись выбросов!!!



Добавим по одному наблюдению в каждую выборку и в результате:

$t = 0.03$

$p = 0.97$

Но если использовать непараметрический аналог t-критерия (**Mann — Whitney U-test**):  
 $p = 0.09$



1	2	3
③	5	7
1	3	6
2	4	5

$$H_0 \quad \mu_1 = \mu_2 = \mu_3 \quad \circ \cdots \circ \cdots \circ$$

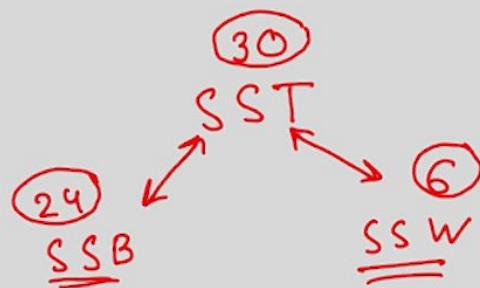
$$H_1 \quad \mu_1 \neq \mu_2 \neq \mu_3$$

$$\bar{X} = \frac{\overbrace{3+1+2}^6 + \overbrace{5+4+3}^{12} + \overbrace{7+6+5}^{18}}{9} = \frac{36}{9} = \underline{\underline{4}}$$

$$F = \frac{\frac{SSB}{m-1}}{\frac{SSW}{N-m}} = \frac{\frac{24}{2}}{6/6} = \underline{\underline{12}}$$

$$SS T = (3-4)^2 + (1-4)^2 + (2-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2 + (7-4)^2 + (6-4)^2 + (5-4)^2 = \underline{\underline{30}}$$

$$df = N - 1 = 8$$



$$SSB = 3(2-4)^2 + 3(4-4)^2 + 3(6-4)^2 = 24$$

$$df = m - 1 = 3 - 1 = 2$$

$$\bar{X} = 2 \quad \bar{X} = 4 \quad \bar{X} = 6$$

$$SSW = (3-2)^2 + (1-2)^2 + (2-2)^2 + (5-4)^2 + (3-4)^2 + (4-4)^2 + (7-6)^2 + (6-6)^2 + (5-6)^2 = 2 + 2 + 2 = 6$$

$$df = N - m = 9 - 3 = 6$$



# Distribution Calculator

Distribution:

F

Degrees of freedom

2 50

Degrees of freedom (2)

1 6 50

Model:

$P(X > a)$

Find Area:

Upper Tail

a

0 12.05 15.27

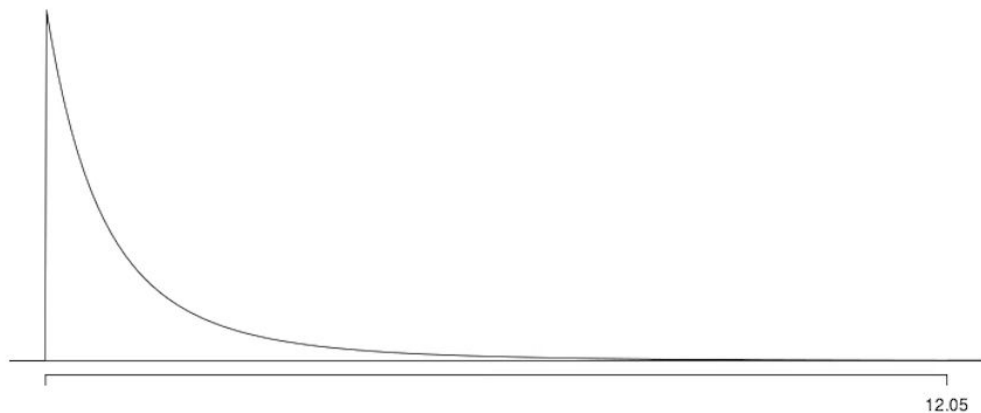
[Rate this app!](#)

[View code](#)

[Check out other apps](#)

[Want to learn more for free?](#)

F Distribution



$$P(X > 12.05) = 0.00792$$



## Однофакторный дисперсионный анализ One-way ANOVA

Генотерапия позволяет корректировать работу дефективного гена, ответственного за развитие заболевания. В эксперименте сравнивалась эффективность четырех различных типов терапии. Результаты исследования представлены в таблице:

Терапия	N	M <sub>x</sub>	SD
A	15	99,7	4,1
B	15	98,8	5,8
C	15	94,4	5,1
D	15	92,3	3,8



Результаты дисперсионного анализа:

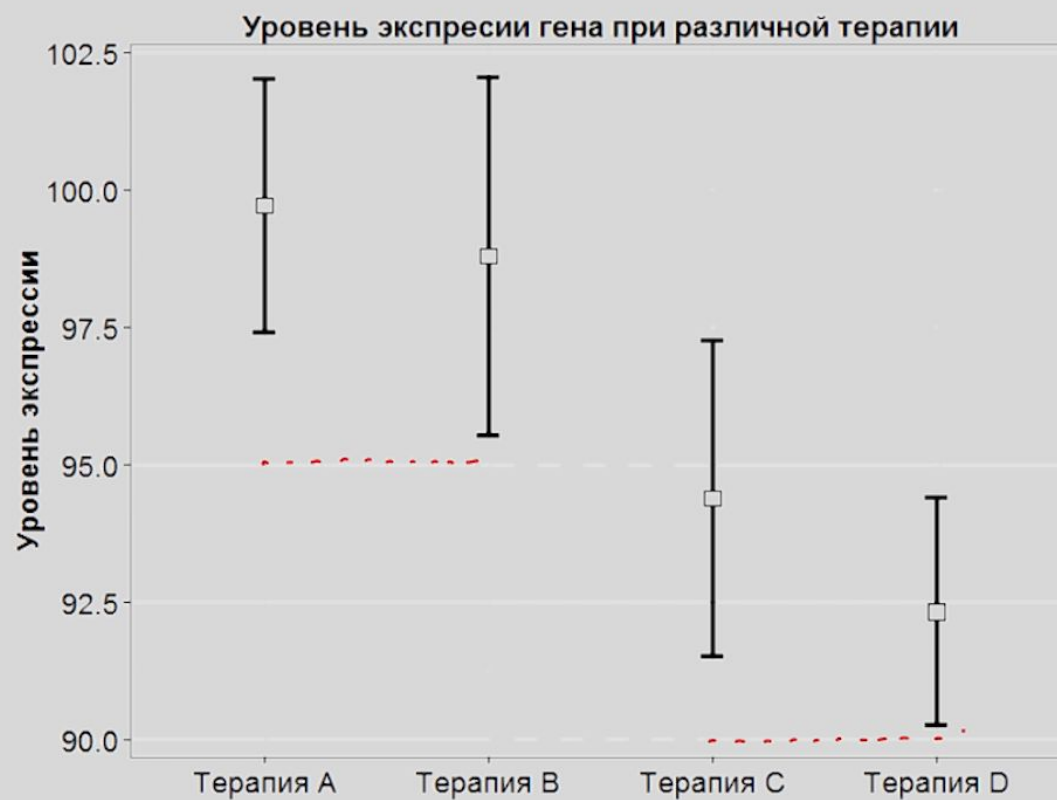
	<u>Df</u>	<u>Sum Sq</u>	<u>Mean Sq</u>	<u>F value</u>	<u>Pr(&gt;F)</u>
* Therapy	3	560.7	186.91	8.037	0.0002
* Residuals	56	1302.3	23.25		

$$df = m - 1 \quad SSB \quad \frac{SSB}{df}$$

$$df = N - m \quad SSW$$



## Строим график и интерпретируем результат



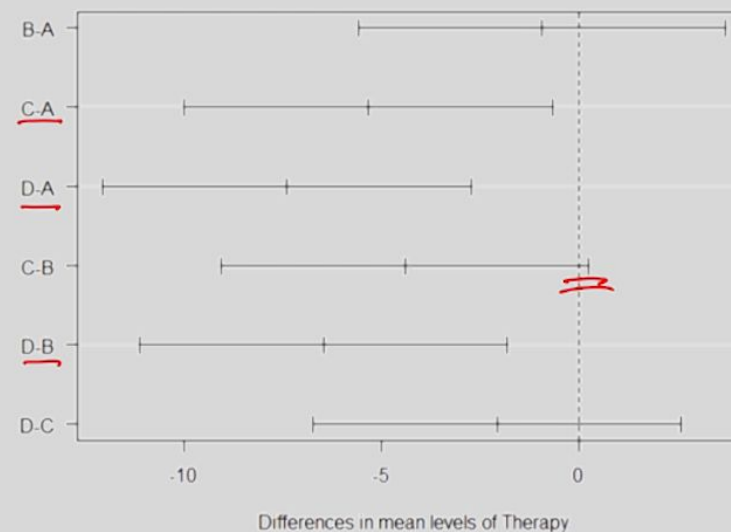
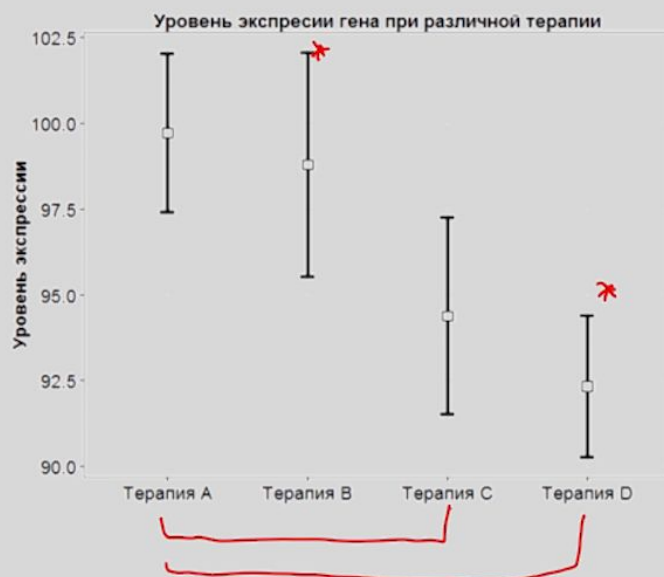
Удалось выявить статистически значимую взаимосвязь типа терапии с показателем уровня экспрессии гена(  $F(3, 56)=8,04$ ,  $p<0,05$  )



# Множественные сравнения в ANOVA критерий Тьюки (Tukey HSD)

$$\bar{X}_A - \bar{X}_B$$

95% family-wise confidence level



	Разница	Верхняя граница	Нижняя граница	p уровень
B-A	-0,9	-5,6	3,7	0,951
C-A	-5,3	-10,0	-0,7	0,019
D-A	-7,4	-12,1	-2,7	0,001
C-B	-4,4	-9,1	0,3	<u>0,071</u> *
D-B	-6,5	-11,1	-1,8	0,003
D-C	-2,1	-6,7	2,6	0,646

$p > 0,05$





## Двухфакторный дисперсионный анализ Two-way analysis of variance

Атеросклероз довольно опасное заболевание – причина ишемической болезни сердца и инсультов. Анализ экспрессии генов лейкоцитов позволяет предсказать вероятность развития данного заболевания. В эксперименте исследовался уровень экспрессии в зависимости от возраста пациентов и дозировки лекарства аторвастатина.

НП  
1  
2

НП  
1  
2

Возраст	Дозировка	N	Mx	SD
молодые	высокая	<u>16</u>	104,8	5,8
молодые	низкая	<u>16</u>	105,5	4,4
пожилые	высокая	16	101	5,1
пожилые	низкая	16	102,3	5,1

Expn Age Dose  
 $x_1$  1 2  
 $x_2$  1 1  
 $x_3$  2 1  
 $x_4$  2 2

Результаты дисперсионного анализа:

$$SS_{total} = SSW + SSB_A + SSB_B + SSB_A * SSB_B$$

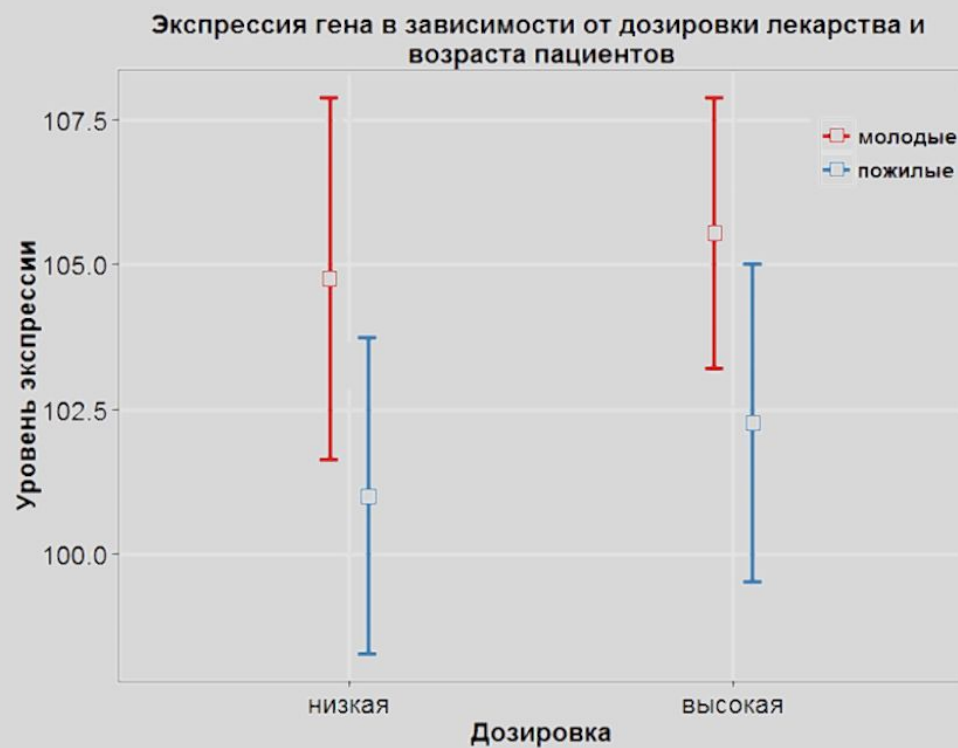
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
* Age	1	197,5	<u>197,45</u>	7,57	<u>0,008</u>
* Dose	1	16,9	<u>16,91</u>	0,64	<u>0,42</u>
Residuals	61	1591,2	26,08		

SSW





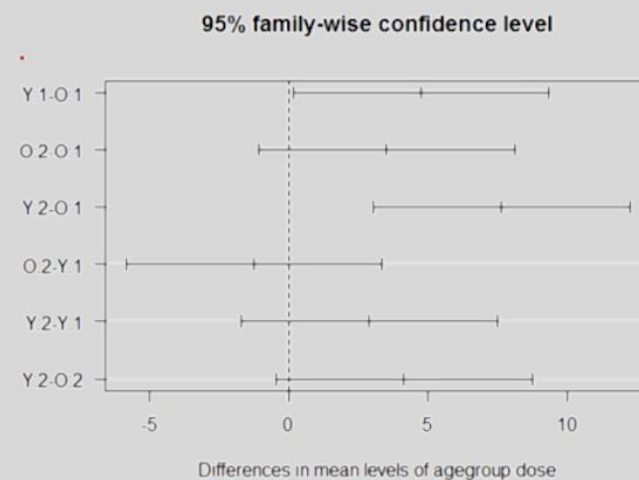
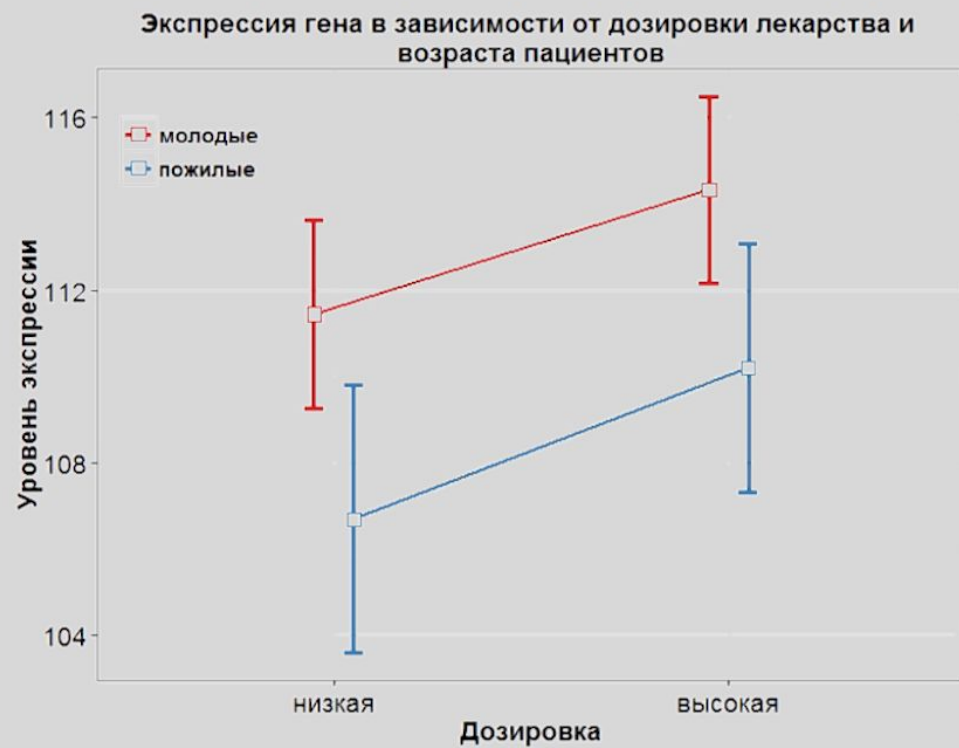
## Строим график и интерпретируем результат



Значимый эффект для фактора возраста пациентов (  $F(1, 61) = 7,57, p < 0,05$  ).  
Незначимый эффект для фактора дозировки (  $F(1, 61) = 0,64, p > 0,05$  )



## Значимы оба фактора



Значимый эффект для фактора возраста пациентов (  $F(1, 61) = 13,25$ ,  $p < 0,05$  ).  
Значимый эффект для фактора дозировки (  $F(1, 61) = 6,87$ ,  $p < 0,05$  )



## Взаимодействие факторов в ANOVA

Исследователей интересовало влияние инъекции некоторого гормона на показатель концентрации кальция в плазме крови у птиц с учетом их пола. В таблице представлены данные экспериментальной и контрольной группы.

НП

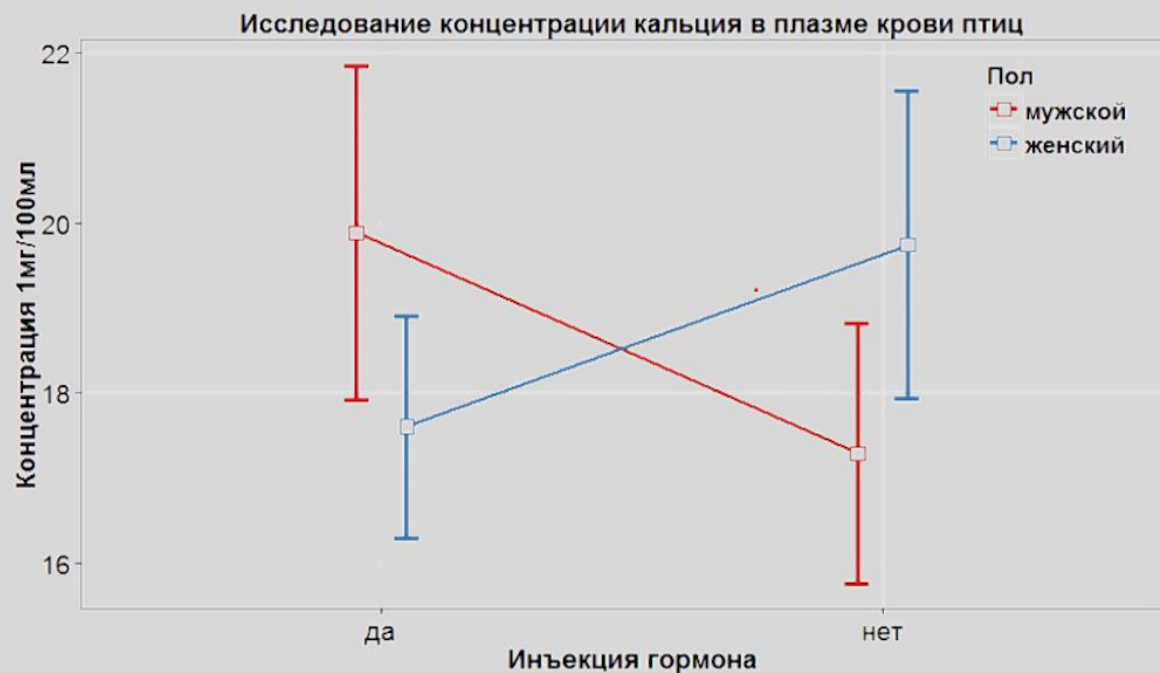
Инъекция	Пол	N	Mx	SD
<u>нет</u>	женский	16	19,9	3,7
нет	мужской	16	17,6	2,4
<u>да</u>	женский	16	17,3	2,9
да	мужской	16	19,7	3,4

Результаты дисперсионного анализа:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hormone	1	0,8	0,85	0,087	0,7697
Sex	1	0,1	0,12	0,012	0,9123
Hormone:sex	1	89,5	89,48	9,136	0,0037
Residuals	60	587,7	9,8		



## Взаимодействие факторов



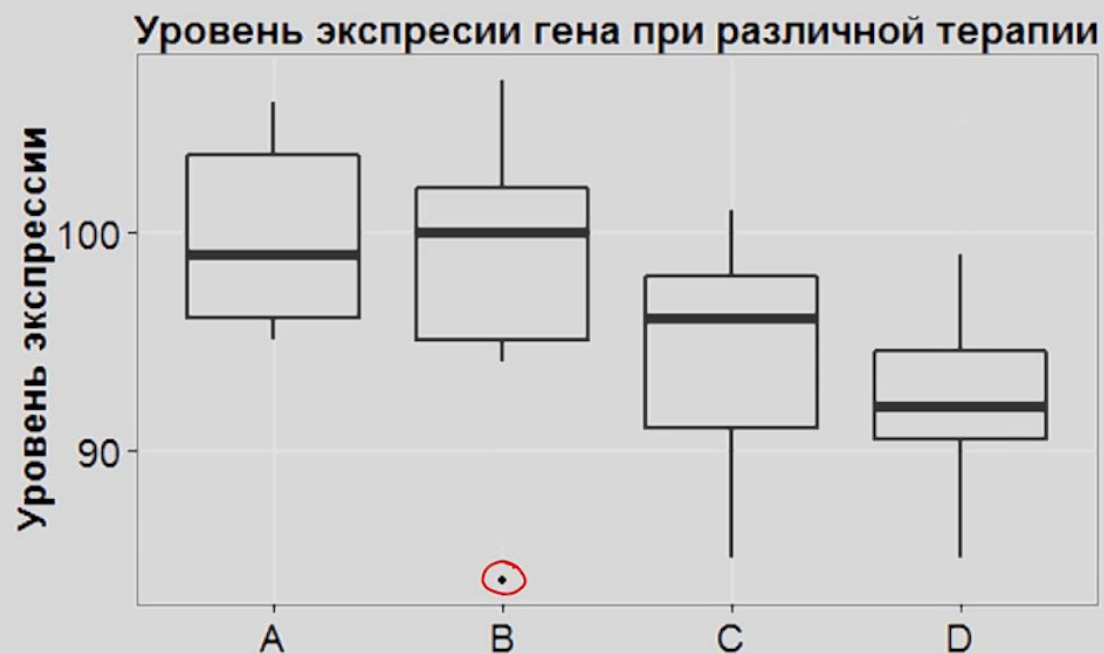
Незначимый эффект для фактора пола (  $F(1, 60) = 0,09$ ,  $p > 0,05$  ).

Незначимый эффект для фактора инъекции гормона (  $F(1, 60) = 0,01$ ,  $p > 0,05$  ).

Значимое взаимодействие факторов (  $F(1, 60) = 9,1$ ,  $p < 0,05$  )



## Гомогенность дисперсий



Levene's Test for Homogeneity of Variance ( $p > 0,05$ )

$$D_1 = D_2 = D_3 = D_4$$



# Нормальное распределение

