

## Элементы математической статистики

### Построение гистограммы

Исходным понятием статистики является понятие **генеральная совокупность**, объединяющее обычно какое-либо множество испытуемых (учащихся) по одному или нескольким интересующим признакам. Главное требование к выделению изучаемой совокупности — это ее качественная однородность, например, по уровню знаний, росту, весу и другим признакам.

Применение большинства статистических методов основано на идее использования небольшой случайной совокупности испытуемых из общего числа тех, на которых можно было бы распространить (генерализовать) выводы, полученные в результате изучения совокупности. Эта небольшая совокупность в статистике называется **выборочной совокупностью** (или короче — **выборкой**).

Любая статистическая работа предполагает выбор генеральной совокупности. Генеральная совокупность – совокупность однородных статистических данных, отличающихся друг от друга, но при этом имеющих сходные свойства. Но, с генеральной совокупностью на практике работать невозможно из-за огромного числа объектов, входящих в нее. Поэтому на практике формируют выборку. Выборка – часть генеральной совокупности. На выборке измеряют переменные. Переменная – это то, что можно измерять, контролировать или чем можно манипулировать в исследованиях. Иными словами, переменная – это то, что варьируется, изменяется, а не остается постоянным. Например, измеряя давление или содержание лейкоцитов в крови можно получить различные значения у различных пациентов или значения для одного пациента в разное время суток. Можно привести примеры разных переменных. И, понятно, что все они будут отличаться своими характеристиками. Так как значения переменных не постоянны, нужно научиться описывать их изменчивость.

Для этого придуманы описательные статистики: минимум, максимум, среднее, дисперсия, стандартное отклонение, медиана, квартили, мода и т.д. Они дают общие представления о значениях, которые принимает переменная.

**Минимум и максимум** – это минимальное и максимальное значения переменной. **Среднее** – это сумма всех значений переменной, деленное на количество переменных.

$$\bar{x} = \frac{\sum x_i}{n}$$

Дисперсия и стандартное отклонение - наиболее часто используемые меры изменчивости переменной.

**Дисперсия** рассчитывается по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

**Стандартное отклонение** равно квадратному корню из дисперсии. Формально имеем:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Чем выше стандартное отклонение и дисперсия, тем сильнее разбросаны значения переменной относительно среднего

**Медиана**- это величина, относительно которой ряд распределения делится на две равные части: в обе стороны от медианы располагается одинаковое число вариантов. При наличии небольшого числа вариантов медиана определяется довольно просто. Для этого все данные нумеруют, и при нечетном числе вариантов центральная варианта и будет его медианой. При четном числе определяется как среднее значение двух величин, находящихся в середине ряда.

Для величин, по которым построена гистограмма, медиану можно определить следующим способом. Необходимо найти класс, в котором содержится медиана. Для этого необходимо складывать частоты встречаемости по классам до тех пор, пока сумма частот не превзойдет половину всех членов ряда. Данный класс называется медианным. Тогда медиану можно найти по формуле:

$$Me = x_n + \lambda \left( \frac{\frac{n}{2} - \sum f_i}{f_{Me}} \right)$$

где  $x_n$ - нижняя граница интервала, содержащего медиану,  $\sum f_i$ - сумма накопленных частот, стоящая перед медианным классом,  $\lambda$ - величина классового интервала,  $f_{Me}$  – частота медианного класса,  $n$ - общее число наблюдений.

**Мода**- это величина, наиболее часто встречающаяся в данной совокупности. Класс с наибольшей частотой называется модальным. Моду можно найти по формуле:

$$Mo = x_n + \lambda \left( \frac{f_2 - f_1}{2f_2 - f_1 + f_3} \right)$$

где  $x_n$ - нижняя граница модального класса,  $f_2, f_1$ - частота класса, предшествующего модальному,  $f_3$ - частота класса, следующего за модальным,  $\lambda$ - ширина классового интервала.

**Коэффициент асимметрии** характеризует симметричность графика распределения относительно среднего значения. Он принимает значения от -1 до 1.

Коэффициент асимметрии определяется по формуле:

$$As = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{3}{2}}}$$

**Экссес**, в свою очередь, характеризует узковершинность или плосковершинность распределения. Экссес определяется по формуле:

$$\Theta = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} - 3$$

Далее необходимо характеристики, полученные на выборке, перенести на всю генеральную совокупность. Для этого вводят понятие доверительного интервала.

### **Интервальная оценка параметров генеральной совокупности**

По известным выборочным характеристикам можно построить интервал, в котором с той или иной вероятностью находится генеральный параметр. Вероятности, признанные достаточными для уверенного суждения о генеральных параметрах на основании известных выборочных показателей, называют доверительными. Обычно в качестве доверительных используют вероятности  $P_1=0.95$ ,  $P_2=0.99$ ,  $P_3=0.999$ .

Это означает, что при оценке генеральных параметров по известным выборочным показателям существует риск ошибиться в первом случае один раз на 20 испытаний, во втором- один раз на 100 испытаний и в третьем- один раз на 1000 испытаний.

Доверительным вероятностям соответствуют следующие величины нормированных отклонений:

вероятности  $P_1=0.95$  соответствует  $t_1=1.96$

вероятности  $P_2=0.99$  соответствует  $t_2=2.58$

вероятности  $P_3=0.999$  соответствует  $t_3=3.29$

**Доверительный интервал**- это интервал, в который попадает среднее значение генеральной совокупности с той или иной вероятностью.

$$\bar{x} - t_p \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + t_p \frac{\sigma}{\sqrt{n}} \text{ -формула доверительного интервала}$$

где  $\bar{X}$  - среднее значение выборки;

$t_p$  – нормированное отклонение;

$S_x$ - стандартная ошибка на генеральной совокупности;

$\sigma_x$  - стандартная ошибка на выборке;

$n$  – объём выборки;

$\mu$ -среднее значение генеральной совокупности.

В качестве наглядного представления выборочной совокупности можно рассмотреть гистограмму. Рассмотрим алгоритм построения гистограммы на конкретной задаче.

**Гистограмма**-это ступенчатая фигура, состоящая из прямоугольников, основании которых равны ширине класса, а высоты-функции плотности вероятности.

## Задача № 1

Дан рост группы детей. Согласно полученным данным построить гистограмму.

70 69 72 73 71 66 73 67 68 73  
71 69 67 74 71 70 70 67 71 69  
70 70 70 71 74 74 71 69 72 71

### Алгоритм построения гистограммы

Предположим, что в результате эксперимента получен ряд значений случайной величины  $-X_i$


$X_1 \quad X_2 \quad X_3 \quad \dots \quad X_n$

1. Строят вариационный ряд-все данные располагают в порядке возрастания.
2. Находят размах варьирования-  $R=X_{\max}-X_{\min}$ .
3. При большом ряде прибегают к группировке. Число групп или классов находят по формуле:  $K=2Lnn$ .
4. Находят величину класса:  $d = \frac{R}{K}$
5. Разбивают выборку на классы:
  1.  $X_{\min}- X_{\min}+d$
  2.  $X_{\min}+d- X_{\min}+2d$
  3.  $X_{\min}+2d- X_{\min}+3d$  и т.д.
6. Находят число измерений, попавших в каждый класс (частота попадания- $h_i$ ).
7. Определяют эмпирическую плотность вероятности случайной величины- $f(x) = \frac{h_i}{nd}$
8. Строят гистограмму: по оси абсцисс откладывают границы классовых интервалов, по оси ординат-значения функции плотности вероятности- $f(x)$ .

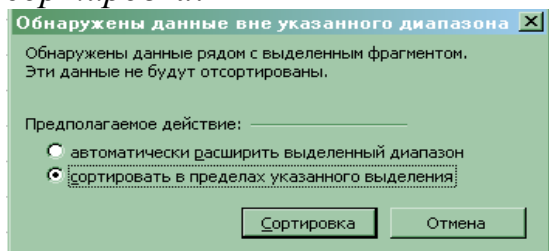
### Ввод данных

1. В ячейку A1 ввести заголовок *РОСТ*
2. В ячейки A2-A31 ввести данные задачи.

### Построение вариационного ряда

1. В ячейку B1 ввести заголовок *вариаци. ряд*
2. Скопировать данные ячеек A2-A31 в ячейки B2-B31. Для этого выделить ячейки A2-A31, щелкнуть правой кнопкой мыши по выделенным ячейкам, выбрать пункт меню *копировать*, встать на ячейку B2, щелкнуть правой кнопкой мыши, выбрать пункт меню *вставить*.
3. Выделить ячейки B2-B31. На панели инструментов выбрать кнопку  в появившемся окне выбрать

предполагаемое действие *сортировать в пределах указанного диапазона*, нажать *сортировка*.



### Расчет необходимых величин

1. В ячейке C1 вводим заголовок *Размах*.
2. В ячейке C2 вводим формулу «=B31-B2». Для этого необходимо в ячейке C2 с клавиатуры ввести знак =, щелкнуть мышкой по ячейке B31, с клавиатуры ввести знак -, щелкнуть мышкой по ячейке B2. После нажатия Enter, в ячейке C2 появится число 8.
3. В ячейке D1 вводим заголовок *число Кл.*
4. В ячейке D2 вводим число классов 4.
5. В ячейку E1 вводим заголовок *Ширина*.
6. В ячейку E2 вводим формулу «=C2/D2». Для этого необходимо в ячейке E2 ввести с клавиатуры знак =, щелкнуть мышкой по ячейке C2, с клавиатуры ввести знак /, щелкнуть мышкой по ячейке D2. После нажатия Enter, в ячейке E2 появится число 2.
7. В ячейку F1 вводим заголовок *Объем*.
8. В ячейку F2 вводим объем выборки 30.

### Определение границ интервалов

1. В ячейках E5, E6, E7, E8 вводим заголовки 1кл, 2кл, 3кл, 4кл.
  2. В ячейках F4,G4 ввести заголовки нижняя, верхняя.
  3. В ячейке F5 ввести минимальное число, т.е. содержимое ячейки B2. Для этого в ячейку F5 ввести формулу «=B2». После нажатия Enter, в ячейке F5 появится число 66.
  4. В ячейку G5 вводим формулу «=F5+\$E\$2». Для этого необходимо в ячейке G5 ввести с клавиатуры знак =, щелкнуть мышкой по ячейке G5, с клавиатуры ввести знак +, щелкнуть мышкой по ячейке E2. Для того, чтобы зафиксировать ячейку E2 для процесса копирования, необходимо ее адрес окружить знаками \$. После нажатия Enter, в ячейке G5 появится число 68.
  5. В ячейку F6 необходимо перенести данные ячейки G5. Для этого в ячейку F6 необходимо ввести формулу «=G5». После нажатия Enter, в ячейке F6 появится число 68.
  6. В ячейки F7,F8 копируем содержимое ячейки F6. Для этого необходимо выделить ячейку F6, подвести указатель мыши к правому нижнему углу ячейки F6, нажать правую кнопку мыши и, не отпуская ее, протянуть до ячейки F8.
  7. В ячейки G6-G8 аналогичным способом скопировать ячейку G5.
- Выполнив эти действия получим:

	E	F	G	H
		нижняя	верхняя	
1кл		66	68	
2кл		68	70	
3кл		70	72	
4кл		72	74	

## Расчет логических функций

Для подсчета частоты попадания в каждый класс необходимо ввести логическую функцию, которая проверяла бы, попадает ли данное число в данный класс или нет. В качестве такой функции можно использовать логическую функцию ЕСЛИ.

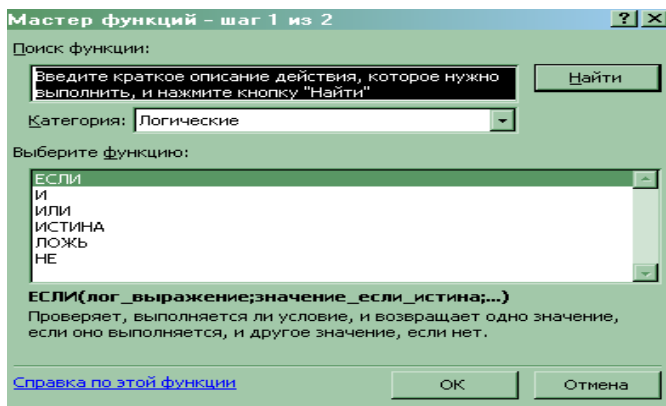
Структура данной функции

**ЕСЛИ(лог\_выражение; Значение\_если\_истина;...)**

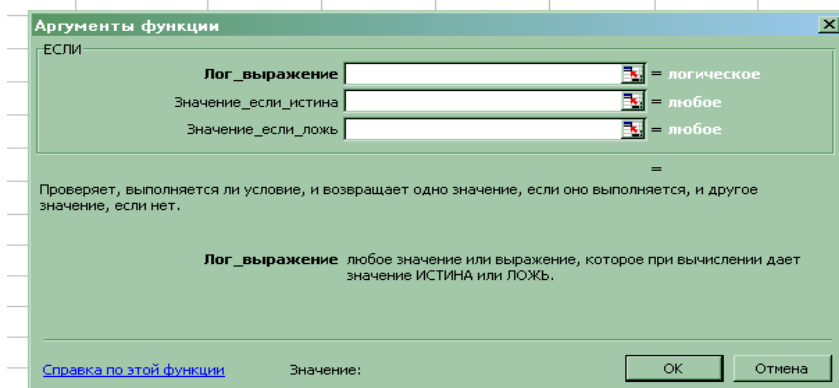
Данная функция проверяет, выполняется ли условие, и возвращает одно значение, если оно выполняется, и другое значение, если нет.

1. В ячейку H1 вводим заголовок 1 класс, I1- 2 класс, J1- 3 класс, K1- 4 класс.

2. В ячейку H2 вводим логическую функцию ЕСЛИ. Для этого в ячейке H2 вводим с клавиатуры знак =, в пункте меню *Вставка* выбрать *Функция*, в категории выбрать *Логические* из предлагаемых функций выбрать ЕСЛИ. нажать ОК.



3. В окне *Аргументы функции* в пункт *Лог\_выражение* ввести «B2<\$G\$5», в пункт *Значение\_если\_истина* 1, в пункт *Значение\_если\_ложь* 0.



После чего нажать ОК. При этом происходит проверка: если число вариационного ряда меньше, чем последнее число первого класса, то в ячейке Н2 появится число 1, иначе число 0.

4. Скопировать ячейку Н2 в ячейки Н3-Н31. Для этого выделить ячейку Н2, подвести указатель мыши к правому нижнему углу ячейки Н2, протянуть указатель мыши до ячейки Н31.

5. В столбце I рассчитываем частоту попадания во второй класс. Для этого в ячейку I2 вводим формулу: «=ЕСЛИ(Н2=1;0;ЕСЛИ(В2<\$G\$6;1;0))». Первое условие для чисел, которые уже попали в первый класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<\$G\$6;1;0) проверяет условие: если число вариационного ряда меньше, чем последнее число второго класса, то в ячейке I2 появится число 1, иначе число 0.

6. Скопировать ячейку I2 в ячейки I3-I31. Для этого выделить ячейку I2, подвести указатель мыши к правому нижнему углу ячейки I2, протянуть указатель мыши до ячейки I31.

7. В столбце J рассчитываем частоту попадания в третий класс. Для этого в ячейку J2 вводим формулу: «=ЕСЛИ(Н2+I2=1;0;ЕСЛИ(В2<\$G\$7;1;0))». Первое условие для чисел, которые уже попали в первый и второй класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<\$G\$7;1;0) проверяет условие: если число вариационного ряда меньше, чем последнее число третьего класса, то в ячейке J2 появится число 1, иначе число 0.

8. Скопировать ячейку J2 в ячейки J3-J31. Для этого выделить ячейку J2, подвести указатель мыши к правому нижнему углу ячейки J2, протянуть указатель мыши до ячейки J31.


9. В столбце K рассчитываем частоту попадания в четвертый класс. Для этого в ячейку K2 вводим формулу: «=ЕСЛИ(Н2+I2+J2=1;0;ЕСЛИ(В2<=\$G\$8;1;0))». Первое условие для чисел, которые уже попали в первый, второй или третий класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<=\$G\$8;1;0) проверяет условие: если число вариационного ряда меньше или равно, чем последнее число четвертого класса, то в ячейке K2 появится число 1, иначе число 0.

10. Скопировать ячейку K2 в ячейки K3-K31. Для этого выделить ячейку K2, подвести указатель мыши к правому нижнему углу ячейки K2, протянуть указатель мыши до ячейки K31.

### Подсчет частоты попадания в каждый класс

Для нахождения частоты попадания в каждый класс, необходимо просуммировать значения столбцов Н, I, J, K.

1. Выделить ячейки Н2-Н31.

2. На панели инструментов выбрать кнопку *Автосумма* . После чего в ячейке Н32 появится число 4.

3. Выделить ячейки I2-I31.

4. На панели инструментов выбрать кнопку *Автосумма*. После чего в ячейке I32 появится число 5.

5. Аналогичные действия для ячеек J и K приведут к тому, что в ячейках I32 и J32 появятся числа 13 и 8.

### Расчет функции плотности вероятности

1. В ячейку A34 вводим заголовок *Частота*.

2. Для удобства расчетов перенесем данные из ячеек H32-K32 в ячейки A35-A38. Для этого в ячейку A35 вводим формулу «=H32», в ячейку A36 – «=I32», в ячейку A37 – «=J32», в ячейку A38 – «=K32».

3. В ячейку B34 вводим заголовок: *Функция*.

4. В ячейке B35 вводим формулу: «=A35/(\$F\$2\*\$E\$2)». Для этого необходимо в ячейке B35 с клавиатуры ввести знак =, щелкнуть мышкой по ячейке A35, с клавиатуры ввести знак /, щелкнуть мышкой по ячейке F2, с клавиатуры ввести знак \*, щелкнуть мышкой по ячейке E2, с клавиатуры ввести знак), добавить знаки \$. После нажатия Enter, в ячейке B35 появится число 0.666667.

5. Скопировать ячейку B35 в ячейки B36-B38. Для этого необходимо выделить ячейку B35, подвести указатель мыши к правому нижнему углу ячейки B35, протянуть указатель мыши до ячейки B38. Выделяем ячейки B35-B38. С помощью пункта меню *Формат/ячейки* выбрать числовой формат *Числовой*, число десятичных знаков 2. При этом появится таблица

33			
34	частота	функция	
35	4	0,07	
36	5	0,08	
37	13	0,22	
38	8	0,13	
39			

### Построение гистограммы

1. Для более наглядного построения гистограммы, перенесем границы интервалов в такой форме:

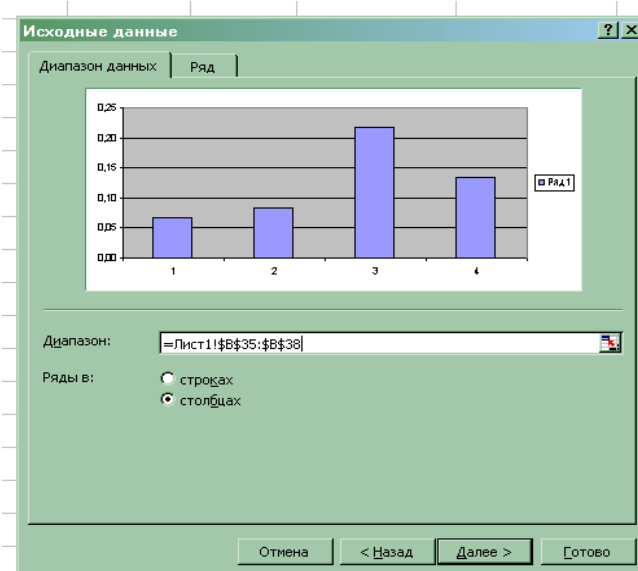
39				
40				
41	частота	функция	границы	
42	4	0,07	66	68
43	5	0,08		70
44	13	0,22		72
45	8	0,13		74
46				
47				

1. На панели инструментов выбрать кнопку *Мастер диаграмм* .

В появившемся окне выбрать тип *Гистограмма*, затем нажать *Далее*.

2. Появится окно *Мастер диаграмм*.

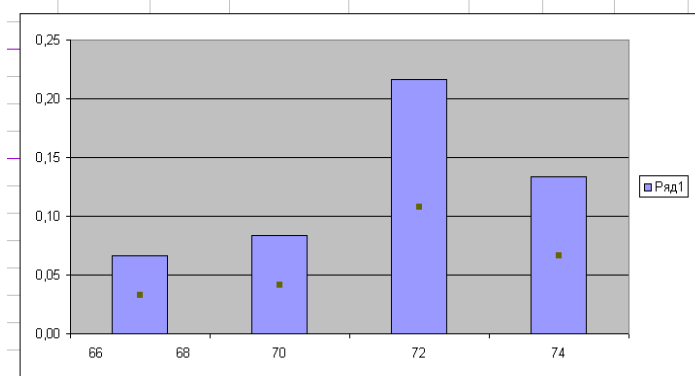




В поле *Диапазон* удалить появившуюся запись, щелкнуть по ячейке В42, протянуть мышку до ячейки В45.

3. Зайти во вкладку *Ряд*. Щелкнуть мышкой в поле *Подписи оси X*, затем щелкнуть мышкой по ячейке С42, протянуть указатель мыши до ячейки С45. Затем нажать *Далее* и *Готово*.

4. Щелкнуть мышкой по одному из прямоугольников гистограммы так, чтобы в центре каждого из прямоугольников появилась точка. Щелкнуть правой кнопкой мыши и в появившемся окне выбрать пункт *Формат рядов данных*.



5. В появившемся окне выбрать вкладку *Параметры*. Затем уменьшить ширину зазора до 0, нажать *Ок*. При этом получим гистограмму

### Задачи для самостоятельного решения

1. Замеры систолического давления у больных гипертонической болезнью 3 степени по выборке (мм. рт. ст.):

227 219 215 230 218 223 220 222 218 219  
 222 221 227 226 226 209 211 215 218 220  
 216 220 220 221 225 224 212 217 219 220

Построить гистограмму.

2. Измерена частота пульса (уд в мин) у здоровых людей. Построить гистограмму согласно полученным данным.

70 69 72 73 71 66 73 67 68 73 71 67 69 74 71 70  
70 67 71 69 70 70 70 71 69 71 74 74 71 69 72 71

3. Значения временного интервала между зубцами R (сек) ЭКГ:

0,74 0,76 0,76 0,76 0,77 0,76 0,76 0,72 0,72 0,69 0,7 0,76 0,77  
0,77 0,79 0,78 0,8 0,69 0,71 0,76 0,76 0,78 0,76 0,77 0,72 0,79 0,75  
0,82 0,86 0,91 0,9 0,84 0,82 0,83 0,82 0,76 0,74 0,7 0,8 0,78

Построить гистограмму.

4. Рост новорожденных (см). Построить гистограмму.

47 51 49 54 48 53 54 52 50 50 50 52 50 55 50  
51 50 46 50 51 49 51 51 53 51 49 51 51 49 49

5. Систолическое давление (мм. рт. ст.) у практически здоровых людей:

127 119 115 130 132 123 120 122 118 119 122 121 127 126 126 109  
111 115 118 120 116 120 120 121 125 124 112 117 119 120

Построить гистограмму.

6. Диастолическое давление (мм. рт. ст) у практически здоровых людей:

67 71 69 74 68 73 74 72 70 70 70 72 70 75 71 70 69 71 71 69  
69 71 70 66 70 71 69 71 71 73

Построить гистограмму.

7. Вес животных при рождении (в кг):

27 32 32 31 32 28 37 35 26 28 32 39 34 30 37 26 27 40 35 37  
28 43 26 35 45 26 35 32 32 35 35 28 32 36 32 36 37 33 28 31

Построить гистограмму.

8. Содержание кальция (мг %) в сыворотке крови обезьян. Построить гистограмму.

13,60 12,90 12,30 9,90 12,73 11,72 10,83 10,42 10,91 10,21 13,10 10,91  
11,96 11,13 13,52 13,53 11,25 10,10 13,96 10,00 11,94 10,82 11,05 12,57  
12,98 10,27 12,67 11,81 12,07 10,65 12,67 10,49 11,18 11,86 9,66 10,05  
9,55 12,50 8,99 12,30

9. Даны значения роста студентов (см) 1 курса. Построить гистограмму.

164 170 164 165 174 180 182 176 169 175 170 169 170 174 156 168  
170 174 167 168 171 182 180 173 178 172 180 168 169 158 169 169  
170 168 172 169 162 167

**10.** Содержание кальция (мг %) в сыворотке крови обезьян:

12,30 14,20 12,60 11,70 12,20 12,30 11,60 12,00 12,50 13,50 11,60 11,90 11,40  
12,00 14,70 11,25 14,20 13,20 12,50 13,80 13,60 12,90 12,30 9,90 12,73 11,72  
10,83 10,42 10,91 10,21 13,10 10,91 11,96 11,13 13,52 13,53 11,25 10,10 13,96  
10,00

Постройте гистограмму.

**11.** При исследовании процесса газообмена лягушек в естественных условиях был получен следующий вариационный ряд:

3,2 4,2 5,3 5,6 5,6 5,9 6,4 6,5 6,8 7,1 7,1 7,3 7,3 7,3 7,3  
7,3 7,4 7,4 7,4 7,4 7,7 9,8 7,3 7,6 9,8 9,8 9,8 10,2 10,6 11,3  
12,3 14,2 7,7 7,7 7,7 7,8 7,9 7,9 8,0 8,3 8,3 8,3 8,3 16,3 8,8  
8,9 9,2 9,4 8,7 8,8 8,5

Построить гистограмму.

**12.** У 60 человек исследовалось количество воды, выпиваемой в течении суток при физической работе в условиях жаркого климата. Получены следующие числовые данные (в литрах). Построить гистограмму.

4.2 4.3 3.4 2.6 4.4 4.8 3.7 4.0 3.2 3.0 5.4 4.4 3.5 4.1  
4.2 5.0 4.7 3.9 3.7 4.5 3.9 3.6 4.6 3.6 4.3 4.5 3.2 3.6  
4.5 4.3 3.7 5.0 5.1 4.5 4.1 4.1 4.7 3.5 4.4 4.1 4.2 4.2  
4.5 4.5 4.1 3.8 4.9 4.0 3.5 3.8 3.7 4.0 3.2 3.9 3.7 3.7  
4.0 3.6 4.4 4.3

**13.** Наблюдения за сахаром крови у 50 человек дали такие результаты:

3.94 3.84 3.86 4.06 3.67 3.97 3.76 3.61 3.96 4.04  
3.82 3.94 3.98 3.57 3.87 4.07 3.99 3.69 3.76 3.71  
3.81 3.71 4.16 3.76 4.00 3.46 4.08 3.88 4.01 3.93  
3.92 3.89 4.02 4.17 3.72 4.09 3.78 4.02 3.73 3.52  
3.91 3.62 4.18 4.26 4.03 4.14 3.72 4.33 3.82 4.03

Построить гистограмму.