## Элементы математической статистики

**Выборка и её представление. Статистическое оценивание**.

**Математическая статистика-**это раздел математики, посвященный математическим методам систематизации, обработки и использования данных для научных и практических выводов.
**Генеральная совокупность-**это совокупность объектов, которые отличаются друг от друга, но схожие определенным признаком.
**Выборка**-это часть генеральной совокупности.
О свойствах генеральной совокупности можно судить по свойствам выборки, поэтому она должна быть репрезентативной.
Вариационный ряд-это данные расположенные в порядке возрастания.
Для наглядности данные представляют в виде полигона или гистограммы распределения.
**Гистограмма-**это ступенчатая фигура, состоящая из прямоугольников, основании которых равны ширине класса, а высоты-функции плотности вероятности.

**Построение гистограммы.**
Предположим, что в результате эксперимента получен ряд значений случайной величины –$X_i$
$X_1$ $X_2$ $X_3$ …….. $X_n$
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: К=2Ln**n**.
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).

**Задача:** Измерена концентрация сывороточного альбумина (г/.л) в крови 50 женщин, включённых в одно обследование. По полученным данным построить гистограмму.

42 41 42 44 44 36 38 41 42 44 42 39 49 40 45 32
34 43 37 39 41 39 48 42 43 33 43 35 32 39 35 43

44 47 40 39 42 41 46 37 49 41 39 43 42 47 48 51
52 34

**Решение:**

1. Строят вариационный ряд-все данные располагают в порядке возрастания:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 32 | 32 | 33 | 34 | 35 | 35 | 35 | 35 | 36 | 37 |
| 37 | 38 | 39 | 39 | 39 | 39 | 39 | 39 | 40 | 40 |
| 41 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 |
| 42 | 42 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 |
| 46 | 46 | 47 | 47 | 48 | 48 | 49 | 49 | 51 | 52 |

2. Находят размах выборки:**$R=X_{max} - X_{min.}$**
   **R=52-32=20**
3. Выбирают количество классов: **к=4**;
4. Находят ширину одного класса по формуле: **d=R/k; d=20/4=5;**
5. Разбивают вариационный ряд на классы и находят частоту попадания в каждый класс:

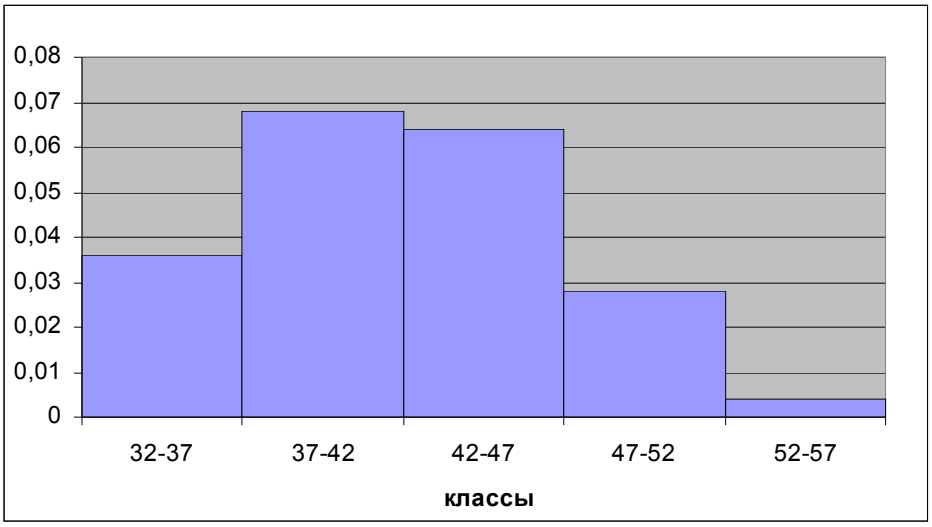
**a) 32-37 $h_1=9$**
**b) 37-42 $h_2=17$**
**c) 42-47 $h_3=16$**
**d) 47-52 $h_4=7$**
**e) 52-57 $h_5=1$**

6.Расчитывают функцию плотности вероятности по каждому классу по формуле: $f(x)=\frac{h_i}{nd}$

1. $f_1=9/250 =0.036$
2. $f_2=17/250=0.068$
3. $f_3=16/250=0.064$
4. $f_4=7/250 =0.032$
5. $f_5=1/250 =0.004$

7 Строят гистограмму, откладывая по оси X значения случайной величины, а по Y-(F)-значения функции плотности вероятности:

| **№ класса** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| классы | 32-37 | 37-42 | 42-47 | 47-52 | 52-57 |
| F | 0,036 | 0,068 | 0,064 | 0,028 | 0,004 |




**Расчёт моды и медианы.**

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:

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

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.

Подставим числовые данные в формулу и рассчитаем медиану:

$$Me = 37 + 5(\frac{\frac{25}{2} - 9}{17}) = 41.7$$

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

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

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.

Подставим числовые данные в формулу и рассчитаем моду:

$$Ìî = 37 + 5(\frac{17 - 9}{2 \cdot 17 - 9 + 16}) = 38$$

**Расчёт коэффициентов ассиметрии и эксцесса.**

Коэффициент ассиметрии определяется по формуле:

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

Эксцесс определяется по формуле:

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

**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;**

$\bar{x} - t_p \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + t_p \frac{\sigma}{\sqrt{n}}$ -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.

**Задача**:

Распределение кальция в сыворотке крови обезьян, как было установлено выше, характеризуется следующими выборочными показателями: $\bar{X}$=11.94 мг, σ =1.27 мг, n =100. Построить 95% доверительный интервал для генеральной средней **μ** этого распределения.

**Дано:**

$\bar{X}$ =11.94 мг

σ=1.27 мг

n =100

P=0.95

_____

**μ=?**

**Решение:**

Применяют формулу доверительного интервала.

$\bar{x} - t_p \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + t_p \frac{\sigma}{\sqrt{n}}$

Подставляют численные данные:

$11.94-1.96 \frac{1.27}{\sqrt{100}} \le \mu \le 11.94+1.96 \frac{1.27}{\sqrt{100}}$

или $11.70 \le \mu \le 12.18$

Следовательно, с вероятностью P=0.95 можно утверждать, что генеральная средняя данного нормального распределения находится между 11.70 и 12.18 мг.

**Ответ:** $11.70 \le \mu \le 12.18$

**Задача:**

Исследователь хочет установить средний уровень гемоглобина в определенной группе населения. Сколько человек он должен обследовать, если в 99 случаев из 100 Δ=±5г/л, а σ=32г/л.?

**Решение:**

**Дано:**

$\sigma_x$ =32

n =10

P=0.99

Δ=5_____

**μ=?**

Решение:

Применяем формулу необходимого объёма выборочной совокупности:

$n = \frac{t^2 \cdot \sigma_x^2}{\Delta^2}$

Где: $\Delta = X - \mu$ -ошибка эксперимента

$\bar{X}$ - среднее значение выборки;

$t_p$ – нормированное отклонение;

$\sigma_x$ - стандартная ошибка на выборке;

**n** – необходимый объём выборки

μ-среднее значение генеральной совокупности.

n= $(\frac{2.58 \cdot 32}{5})^2 \approx 273$

Ответ: n=273

**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 3.91 3.62 4.18 3.82 3.94 3.98 3.57 3.87 4.07 3.99 3.69 3.76 3.71 4.26 4.03 4.14 3.81 3.71 4.16 3.76 4.00 3.46 4.08 3.88 4.01 3.93 3.72 4.33 3.82 3.92 3.89 4.02 4.17 3.72 4.09 3.78 4.02 3.73 3.52 4.03
Построить гистограмму.

**14.** При изучении роста лабораторных крыс коэффициент вариации веса крыс был примерно 13% , а $\overline{X}$=200г. Чему равны среднеквадратическое отклонение и дисперсия веса крыс?

**15.** У группы лиц исследовались функции:
А) потоотделения,
Б) величина кровяного давления,
В)частота пульса при мышечной работе.

Получены следующие характеристики этих процессов.

А: $\overline{X}_1$=200мл $\sigma_1$=22 мл.

Б: $\overline{X}_2$ =160 мм. рт. ст. $\sigma_2$=8 мм рт ст.

В: $\overline{X}_3$=120 уд в мин. $\sigma_3$=16 уд в мин.

1) Сравнить данные процессы по степени их изменчивости.
2) Какой процесс является более изменчивым при мышечной работе человека?

**16.** При исследовании газообмена лягушек в естественных условиях были получены следующие числовые значения для количества кислорода, потребленного за один час (в $см^2$ на 100 г веса): 6,7,7,7,8,8,8,9,9,10,11
Определить среднее количество потребленного кислорода в течение часа, найти дисперсию и среднеквадратическое отклонение.

**17.** При изучении длины листьев садовой земляники сделана выборка. Среднее квадратическое отклонение равно 1.32 см. С вероятностью 0.95 определить такое минимальное число измерений, чтобы отклонение выборочной средней от математического ожидания не превышало 0.06 см.

**18.** Измерено 9 листьев земляники. Получены значения Xср=5см, стандартное отклонение 1.5 см. Каковы доверительные интервалы для $\mu$ при уровнях значимости 0.05; 0.01?

**19.** Для определения средней урожайности овса взято 20 проб ( на 1 $м^2$ ) и для них определено Xср=0.125кг. Среднее квадратическое отклонение равно 0.052. Определите, в каких границах заключена средняя урожайность с 1 $м^2$ по всему полю, если вывод следует сделать с надежностью 0.9.

**20.** С помощью случайной выборки, состоящей из 16 витаминных драже, исследовалось содержание витамина Е. Среднее значение оказалось равным 18,1 весовой единицы, а стандартное отклонение 1,2. Найдите границы 95 процентного интервала содержания витамина Е во всей совокупности витаминных драже.

**21.** С помощью случайной выборки состоящей из 625 человек, исследовался диаметр мышцы бедра, среднее значение которого оказался равным 17,1 см, а стандартное отклонение 1,4 см. Найдите границы 95 и 99 процентного доверительного интервала.

**22.** В результате десяти измерений диаметра капилляра (мкм)в стенке лёгочных альвеол были получены следующие данные: 2,83; 2,82; 2,81; 2,85; 2,87; 2,86; 2,83; 2,85; 2,83;2,84. Вычислить точечную и интервальную оценки для диаметра капилляра с доверительной вероятностью P=0,95

**23.** При определении микроаналитическим способом содержания азота в данной пробе были получены следующие результаты: 9,29; 9,38; 9,35; 9,43; 9,53; 9,48; 9,61; 9,68 (%). Оценить среднее содержание азота в пробе, среднеквадратическое отклонение при доверительной вероятности P = 0,95. Найдите доверительный интервал.

**24.** При фотоэлектроколориметрическом определении концентрации ацетилсалициловой кислоты на основании реакции с сульфатом меди и пиридином были получены следующие результаты: 99,2%; 99,0%; 98,9%; 99,3%; 98,8%; 99,1 %. Вычислить среднее значение концентрации ацетилсалициловой кислоты, среднеквадратическое отклонение при доверительной вероятности P = 0,95. Найдите доверительный интервал.

**25.** При анализе лекарственного препарата (с целью контроля его качества) метазона – 1%-ного раствора для инъекций – найдены следующие значения pH этого раствора: 4,50; 4,52; 4,55; 4,60; 4,70; 4,75. Вычислить среднюю величину pH раствора, среднеквадратическое отклонение при доверительной вероятности P = 0,99. Найдите доверительный интервал.

**26.** В десяти одинаковых пробах были получены следующие значения содержания марганца: 0,69; 0,70; 0,67; 0,66; 0,67; 0,68; 0,67; 0,69; 0,68; 0,68 (%). Вычислить среднюю величину содержания марганца, среднеквадратическое отклонение при доверительной вероятности P = 0,95. Найдите доверительный интервал.

**27.** При определении посторонних примесей в образце лекарственного препарата найдено суммарное содержание примесей : 1,3; 1,4; 1,5; 1,6; 1,6 (%) Вычислить среднюю величину содержания примесей, среднеквадратическое отклонение при доверительной вероятности P = 0,99. Рассчитайте доверительный интервал.

**28.** Высота стебля кукурузы X-случайная величина, имеющая нормальное распределение. Сколько необходимо отобрать растений, чтобы Xср отличалось от $\mu$ меньше, чем на 2 см, если известно, что по результатам проведенных предыдущих измерений стандартное отклонение -6см. Результат найти с надежностью 0.95.

**29.** Сколько следует изучить историй болезни больных дизентерией, чтобы определить средние сроки их лечения, имея в виду, что при одинаковых условиях в 95 случаях из 100 $\Delta=\pm 0{,}5$ дня, а $\sigma=\pm 1{,}5$ дня.

**30.** Наблюдения за дневным удоем восьми коров, случайно отобранных из стада, дали следующие результаты:

| | | | | | |
|---|---|---|---|---|---|
| УДОЙ | 12 | 13 | 15 | 16 | 18 |
| ЧИСЛО ГОЛОВ | 1 | 1 | 3 | 2 | 1 |

a) Определить вероятность того, что средний удой по всему стаду будет отличаться от среднего удоя восьми голов не более, чем на 2.5кг.

b) С P=0.95 найти доверительный интервал для среднего удоя по стаду.

**31.** Исследователь хочет установить средний уровень гемоглобина в определенной группе населения. Сколько человек он должен обследовать, если в 95 случаев из 100 $\Delta=\pm 2$г/л, а $\sigma=24$г/л.?

**32.** Определить минимальное число семей, которое нужно обследовать с целью установления среднего размера семьи с точностью среднего результата, не превышающего 0,2 ($\Delta\leq 0{,}2$) при доверительной вероятности P=0,99. При проведении пробного исследования 10 семей установлено, что среднеквадратическое отклонение составляет 1,3.

**33.** Определить необходимое для исследования число женщин 20-летнего возраста для получения среднего роста с точностью до 0,5см ($\Delta\leq 0{,}5$см) при доверительной вероятности P=0,95. При пробном исследовании 10 женщин получено значение среднеквадратического отклонения 5,5.

**34.** Рассчитать минимальное число наблюдений при исследовании окружности грудной клетки у женщин, если $\bar{X}=88$см, $\sigma=3{,}2$см при заданной точности исследования ($\Delta\leq 0{,}9$) и доверительной вероятности P=0,95.