МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РУз
БУХАРСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ ИНСТИТУТ

Кафедра: Общей гигиены и СГОЗ

Л Е К Ц И Я

Тема: СРЕДНИЕ ВЕЛИЧИНЫ

Лектор: Кобилова Г. А.

БУХАРА - 2014

Для характеристики статистической совокупности используют так называемые абсолютные и относительные величины.
Абсолютные числа имеют ограниченное познавательное значение, хотя сами по себе несут важную информацию о размере того или иного явления (количество больных, умерших, родившихся) ими можно пользоваться при характеристике редко встречающихся явлений.
Основной недостаток абсолютных чисел в том, что их нельзя применять для сравнения; сравнивать размеры двух явлений или изучать изменение этих явлений во времени. Для этого необходимо абсолютные числа, выражающие эти размеры, преобразовать в относительные величины.
Например: В районе "К" в 2006 г. было зарегистрировано 50 случаев вирусного гепатита. В 2011 г. в этом же районе было зарегистрировано 70 случаев гепатита. Но в 2011 г. население района "К" составляло 40 ООО человек, а в 2011 г. 60 ООО человек. Для сравнения заболеваемости гепатитом необходимо преобразовать эти абсолютные чис ла в относительные, путем пересчета заболеваемости на 10 ООО человек

Итак: 2006 г.
40 ООО - 50
10 ООО - х
$x = \frac{50 \times 10000}{40000} = 12,5\%$

2011 г.
60 ООО - 70
10 ООО – х
$x = \frac{70 \times 10000}{60000} = 11,1\%$

Отсюда видно, что заболеваемость вирусным гепатитом в 2006 г. составляла на каждые 10000 человек - 12,5, а в 2011 г. на 10 ООО – 11,1
Таким образом для правильного и точного вывода необходимо использовать относительные величины. Они рассчитываются путем отношения (деления) одной абсолютной величины на другую и полученную дробь умножают на 100, 1000, 10000 и т.д. Соответственно этому относительные величины могут быть выражены в процентах ,промилле (‰) или продецимилле (‱) и т.д.
Различают следующие виды относительных величин:
1. Экстенсивный показатель.
2. Интенсивный показатель.
3. Показатели соотношения.
4. Показатели наглядности.

Экстенсивный показатель - это показатель распределения, характеризует отношение части к целому и выражается в процентах.

$$\text{Э. П.} = \frac{\text{абсолютный размер части явления}}{\text{абсолютный размер явления в целом}} \times 100$$

Например: для вирусного гепатита или кори (в %) среди всех инфекционных заболеваний.

Интенсивный показатель - или показатель частоты, распространенности, характеризует частоту явления в той среде, где мы это явление наблюдаем. Явление представляет собой как бы продукт среды. Например: больные (среда) и умершие из их числа (явление), студенты (среда) изаболевшие гриппом из их числа (явление).

$$\text{И. П.} = \frac{\text{абсолютный размер явления}}{\text{абсолютный размре среды(продуцирующий данное явление)}} \times 100, 1000, 10000$$

Показатель соотношения - характеризует соотношение двух, не связанных между собой совокупностей.

$$\text{П. С.} = \frac{\text{абсолютный размер явления}}{\text{абсолютный размер среды(не продуцирующий данное явление)}} \times 100, 1000, 10000$$

Показатели наглядности - учитывают, на сколько процентов или во сколько раз произошло увеличение или уменьшение сравниваемых величин.

$$\text{П. Н.} = \frac{\text{явление}}{\text{такое же явление из ряда сравниваемых, принятое за 1}} \text{ или } 100$$

Теперь рассмотрим- на примере вычисление относительных величин.

В районе "К" в 2012 г. было зарегистрировано 900 случаев заболеваний. Численность населения района 45000 человек. Из числа за регистрированных заболеваний 450 составляет грипп, 300 случаев - гепатит, 100 заболеваний пищеварительного тракта, 50 - травматизм. В 2011 г. в районе число коек - 200, в 2012 г. - 250 коек.

По представленным данным необходимо вычислить возможные относительные показатели.

$$\text{И. П.} = \frac{\text{явление} \times 100}{\text{среда}} = \frac{900 \times 1000}{45000} = 20\%$$

$$\text{Э. П.} = \frac{\text{часть явления}}{\text{явление в целом}} \times 100$$

$$\text{ЭП (грипп)} = \underline{450 \times 100} = 50\%$$

$$\text{ЭП (гепатит)} = \underline{300 \times 100} = 33.3\%$$

$$\text{ЭП (заб.пищев.тракта)} = \underline{100 \times 100} = 11.1\%$$

$$\text{ЭП (травм)} = \underline{50 \times 100} = 5.6\%$$

$$\text{П.С. (2011 г.)} = \underline{200 \times 1000} = 4.4\%$$

$$\text{П.С. (2012г.)} = \underline{250 \times 1000} = 5.5\%$$

$$\text{П.Н.} = \frac{250 \times 100}{200} \text{ или } \frac{250}{200} = 1.25 \text{ раз}$$

С Р Е Д Н И Е В Е Л И Ч И Н Ы

Средние величины применяются в тех случаях когда характеризующий признак присущ всем единицам наблюдения, составляющим статистическую совокупность, если требуется охарактеризовать совокупность одним числом.

Прежде, чем приступить к непосредственному вычислению средних величин, необходимо составить так называемый вариационный ряд.

Вариационный ряд - это ряд числовых значений изучаемого признака, отличающихся друг от друга по своей причине и расположенных в ранговым порядке.

Характеристиками вариационного ряда являются:

вариантa (V) - числовое значение изучаемого признака, частота (P) - частота встречаемости каждой варианты,

общее число наблюдений ($n = \Sigma P$), где Σ - знак суммы.

Вариационный ряд может быть простым и сгруппированным.

Простой вариационный ряд составляется обычно при малом числе наблюдений ($n \leq 30$) и каждая варианта здесь встречается один раз. Например: Дана масса тела пятерых подросков 16-ти лет:

V	P
50	1
51	1
52	1
53	1
54	1
55	1
	n = 5

Сгруппированный вариационный ряд - каждая варианта встречает ся несколько раз, число наблюдений n > 30

V	P
50	4
51	8
52	9
53	10
54	5
55	3
	n=41

Составление сгруппированного вариационного ряда состоит из 5 этапов:

1. Определить количество групп в ряду, которое будет зависить от числа наблюдений, чем больше наблюдений, тем больше групп. Число групп определяется на основании данной таблицы:

Число варивнт	31-45	46-100	101-200	201-500
Число групп	6-7	8-10	11-12	13-17

2. Определить интервал (i) между группами по формуле:

$$i = \frac{V_{max} - V_{min}}{\text{число групп}}$$

3. Определить границы и середину каждой группы.
4.Распределить изучаемую совокупность по группам.
5.Графическое изаображение вариационного ряда.

Например: Дано число дней, проведенных в стационаре 45-ти больными.

Число проведенных дней V	Число больных P
3	1
4	2
5	2
6	2
7	3
8	3
9	4
10	5
11	6
12	4
13	3
14	2
15	2

16	2
17	1
18	1
19	1
20	1
	n=45

Составляем группированный вариационный ряд:

1. Определяем количество групп в ряду на основании таблицы. Оно равно 6, т.к. n=45.

2.Определяем интервал (i):

$$i = \frac{Vmax - Vmin}{\text{число групп}} = 20-3\backslash 6=3$$

3.Определяем границы и середину каждой группы: первая группа -3-5, а ее середина - 4, следующая 6~Ъ, середина группы соответст**евенно** т **7** и т.д.

4. Распределение по группам:

Длительность лечения	Середина группы	Число больных
3-5	4	5
6-8	7	8
9-11	10	15
12-14	13	9
15-17	16	5
18-20	19	3

Составление графического изображения вариационного ряда

После составления простого или сгруппированного вариационного ряда определяется средний уровень признака - средняя величина.

Средняя величина - это обобщающая характеристика признака в статистической совокупности.

Общеупотребительными являются 3 вида вредних величин: Мода (Мо), медиана (Ме), средняя арифметическая СМ).

I. Мода (Мо) - это величина признака чаще других встречающаяся в вариационном ряду.

Например: В нашем ряду Мо- 10, тк. 10 дней лечения провели большее число больных - 15.

V	P
4	5
7	8
10	15
13	9
16	5
19	3

2. Медиана (Ме) - величина признака, занимающего в вариационном ряду серединное положение и делит вариационный ряд пополам.

Определяется медиана следующим образом: при четном числе наблюдений за медиану принимают среднюю величину из двух центральных вариантов. При нечетном

Ме=п+1\2

45 +1\2

Так, в вышеприведенном примере, где п=45 Ме = 23.

Следовательно 23-я. варианта и будет медианой.

Средняя арифметическая (М) « вычисляемая с помощью следующих формул:

а) простая средняя арифметическая по формуле:
М = Ер\п.   при п<30
Этот способ применяется в том случае, когда каждая варианта встречается I раз и п < 30.

V	Р
50	1
51	1
52	1
53	1
54	1

$$\mathbf{M} = \frac{\mathbf{EP}}{\mathbf{П}} = \frac{260}{5} = 52 \text{ кг}$$

Следовательно: Средний вес у пятерых 16-летних подростков 52 кг.
б) взвешенная средняя арифметическая по Формуле:

$$M = \frac{\mathbf{EVP}}{\mathbf{n}}$$

Этот способ применяется в том случае, когда каждая варианта встречается несколько раз и $n > 30$.

V	Р
50	4
51	8
52	12
53	10
54	6
55	3

$$M = \frac{\mathbf{EVP}}{\mathbf{n}} = \frac{(50x4)+(51x6)+(52x12)+(53x10)+(54x6)+(55x3)}{39} =$$

Способ моментов - он применяется в случаях, когда варианты представлены большими числами (например, масса новорожденных в граммах) и имеется число наблюдений, выраженное сотнями или тысячами случаев. При этом средняя арифметическая вычисляется по формуле:

$$M = M1 + \frac{Edp}{n}$$ где

М1 - условная средняя (чаще всего в качестве нее берется Мо);
Р - частота
Е- знак суммы
n - общее число наблюдений,
Выражение $\frac{Edp}{n}$ - среднее отклонение всех вариант от условной средней (или 1-й момент средней),
Этапы расчета М по способу моментов:

1. За условную среднюю М1 рекомендуется принять варианту чаще других повторяющуюся в вариационном ряду М1=Мо.

2. Определяем ***d*** - условное отклонение от условной средней по формуле: ***d =*** V-Мо, т.е. из каждой варианты вычитаем условную среднюю.

3. Умножаем условное отклонение (***d***) на частоту (Р) каждой варианты и получаем произведение ***dp.***

4. Получаем сумму Ed^2p

5. Определяем среднее отклонение от условной средней (момент 1-ой степени) $\frac{Ed}{p}$

n

6. Определяем среднюю арифметическую по способу моментов:

$$M = M1 + \frac{Edp}{n}$$

Средняя арифметическая величина обладает тремя свойствами:

1. Занимает серединное положение в вариационном ряду.
2. Является обобщающей величиной и имеет абстрактный характер.
3. Сумма отклонений всех вариант от средней равна нулю О.

Пример: Вычисление средних величин но способу моментов.

V	P	D	dP	Ed ^{2}P
6	3	-4	-12	48
7	10	-3	-30	90
8	15	-2	-30	60
9	20	-1	-20	20
10	24	0	0	0
11	18	1	18	18
12	16	2	32	64
13	9	3	27	81
14	6	4	24	96
15	2	5	10	50
	n=123		19	527

Вычислить среднюю длительность лечения больных гастроэнтерологического отделения по способу моментов.

1.Находим условную среднюю

Mo = 10.

2.Определяем d=V-M1

3.Умножаем на P.

4.Получаем Ed^2P = 527.

5.Определяем момент первой степени $V = \frac{Ed p}{n}$

6.Расчитываем среднюю арифметическую по способу моментов

$$M = Mo + \frac{Ed p}{n}$$

7. Определение среднего квадратического отклонения δ или Момента 2-й степени

$$\delta \pm \sqrt{\frac{Ed^2 p}{n}}$$

8.Определяем ошибку репрезентативности средних величин. $m_м = \pm \frac{\delta}{\sqrt{n}}$

9.Определяем доверительные границы $M_0 + m_м$

Вывод: Средняя длительность лечения 123-х больных гастроэнтерологического отделения составляет 10,15 дней.

Критерии разнообразия признака.

Величина того или иного признака неодинакова у всех членов совокупности, несмотря на ее относительную однородность.

Вариабильность демонстративно представлена на примере роста 3-х групп мальчиков 12-ти лет. В каждой группе мальчики как бы поставлены в шеренгу по велишне их роста. Сравниваемые три группы не различаются, по числу наблюдений (n = 16), а также по среднему уровню признака, т.к. средняя арифметическая роста для каждой группы одинакова(M=153 см). Однако эти группы различны по разнообразию роста лиц, их составляющих.

Статистика позволяет охарактеризовать это специальными критериями, определяющими уровень разнообразия каждого признака в той или иной группе. К таким критериям относятся:

1) Лимит lim = (Vmax – Vmin)

3. Среднее квадратическое отклонение (δ))

4. Коэффициент вариации (Cv)

Так как каждый из этих критериев имеет свое самостоятельное значение, то следует остановиться на них отдельно.

Это мы сделаем с помощью следующей таблицы:

	1-я группа	2-я группа	3-я группа
V	151,152,153,154,155	153	141,147,159,165
P	1,4,6,4,1	16	4,4,4,4
Число наблюдений	$N_1 = 16$	$n_2 = 16$	$n_3 = 16$
Средняя арифмитическая	$M_1 = 153$ см	$M_2 = 153$ см	$M_3 = 153$ см
Среднее квадратическое отклонение	Δ_1=1,05	δ_2=0	δ_3=8,7

1. Лимит(lim)определяется крайними значениями вариант в вариационном ряду.

lim = Vmax + Vmin

Для 1-ой группы $lim_1 = 155 + 151$

2-ой группы $lim_2=153 + 153$

3-ой группы $lim_3 = 165+141$

Значит $lim_3>lim_2>lim_1$ при малом числе наблюдений $n \leq 30$.

Вывод: наибольший лимит наблюдается в 3-ей группе подростков.

2. Амплитуда (Am) - разность крайних вариант.

1-я группа $Am_1 = 155 – 151 = 4$см

2-я группа $Am_2 = 153 – 153 = 0$ см

3-я группа $Am_3 = 165 – 141 = 24$ см

Значит $Am_3>Am_2>Am_1$. Наибольшая амплитуда наблюдается в 3-й группе.

3. Внутреннюю структуру совокупности характеризует среднее квадратическое отклонение δ (сигма), которая вычисляется

1) при $n \leq 30$

$$\delta=\pm\sqrt{\frac{\Sigma d^2}{n-1}}$$

2) при $n > 30$

$$\delta = \pm\sqrt{\frac{\Sigma d^2 P}{n-1}}$$

3) по способу моментов:

$$\delta = \sqrt{\frac{\Sigma d^2 P}{n} - \left(\frac{\Sigma dP}{n}\right)^2}$$

Выражение $\frac{\Sigma d^2 P}{n}$ называется 2-ой момент средней.

Среднее квадратическое отклонение используется при:

1. Вычислении ошибки репрезентативности.
2. Для расчета коэффициента вариации.
3. Для определения концентрированности вариационного ряда. При нормальном распределении признака в пределах $M \pm \delta$ находится 68% всех случаев, а в пределах $M \pm 2\delta$ -95,5% всех случаев, в пределах $M \pm 3\delta$- 99,7% всех случаев, составляющих совокупность. Таким образом $M \pm 3\delta$ охватывает почти весь вариационный ряд. Можно воспользоваться этим правилам для выяснения вопроса о типичности средней величины. Если 95% всех вариант находится в пределах $M \pm 2\delta$, то средняя является характерной для данного ряда и не требуется увеличивать число наблюдений в совокупности. Для определения типичности средней сравнивается фактическое распределение с теоритическим путем расчета сигмальных отклонений.

Чем меньше δ , тем меньше колебаемоеть ряда, тем точнее и типичнее вычисленная в этом ряду средняя.

4. Коэффициент вариации (C v) - характеризует степень разнообразия признаков и позволяет выявить более устойчивые и менее устойчивые признаки в совокупности. Коэффициент вариации вычисляется по следующей формуле:

$$Cv = \frac{\delta}{M} \times 100 ,$$

где δ – среднее арифмитическое отклонение

M – средняя арифметическая величина

При $Cv < 10\%$ - слабое разнообразие признака

$Cv = 10 – 20\%$ - среднее разнообразие признака

$Cv > 20\%$ - сильное разнообразие признака

Тема:Оценка взаимосвязи между явлениями и признаками.

Все явления и процессы, совершающиеся в природе и обществе взаимосвязаны и взаимообусловлены. В социальной гигиене и организации здравоохранения, в различных разделах медицины и биологии часто приходится проводить статистический анализ всевозможных признаков в совокупности.

Различают все формы проявления связей между явлениями или процессами:

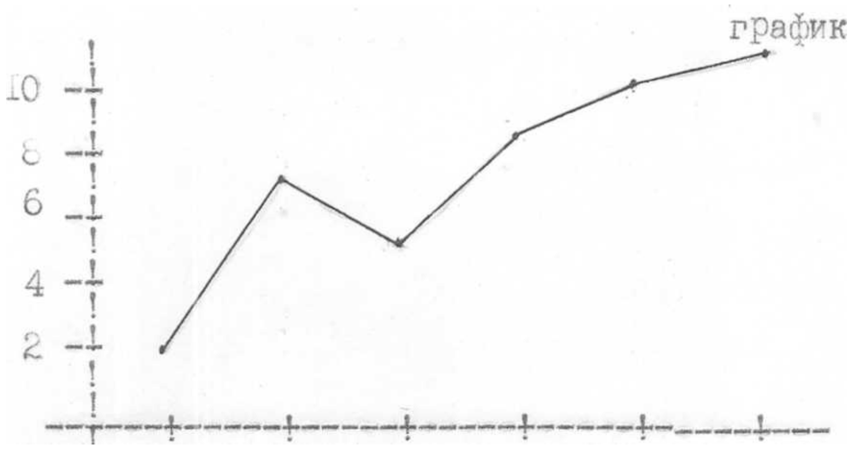
1. функциональная связь.
2. корреляционная связь.

При функциональной связи каждой отдельной величине соответствует определенная взаимосвязанная с ней другая величина, изменение одной величины строго ведет к изменению другой величины.

Так например: с изменением радиуса меняется площадь круга или увеличением температуры нагрева увеличивается объем тела и т.д. тональная связь характерна для физико-химических процессов.

При корреляционной связи одной величине признака соответствует несколько других значений другого взаимосвязанного с ним признака, варьирующих около своей средней.

Примером корреляционной связи может служить зависимость веса тела от его роста. У группы лиц с одинаковым ростом наблюдаются различные колебания массы тела, при этом на величину веса влияет не только рост человека, влияет также питание; состояние здоровья, нерв но-психическое состояние и т.д. Однако эти колебания весьма варьируют в определенных размерах вокруг своей средней величины.



1-5 6-хО 11-15 16-20 20

Другой пример: при одинаковой температуры тела у различных людей наблюдаются индивидуальные колебания частоты сердечных сокращений, варьирующие вокруг своей средней.

Важно отметить, что корреляционная связь проявляется лишь в массе наблюдений, т.е. в совокупности.

Используя методы корреляции, важно помнить о возможности измерять связь между различными признаками только лишь в качественно однородной совокупности нельзя, например, сопоставлять рост и массу тела людей, состоящих из лиц разного пола и возраста.

Корреляция - латинское слово, означающее взаимосвязь, взаимозависимость.

Корреляция может быть представлена в виде таблицы, графика и коэффициента корреляции. Таблицы и графики дают лишь представление о наличии и направлении связи. Так, например, связь между стажем работы ткачих и заболеваемостью.

Стаж работы	Заболеваемость
До 1 года	2
1 – 5 лет	6,8
6 – 10 лет	5,1
11 – 15 лет	8,1
16 – 20 лет	10,2
20 лет и больше	10,5

Как видно из вышеприведенной таблицы между стажем работы ткачих и их заболеваемостью существует корреляционная связь. При этом с увеличением стажа работы заболеваемость повышается. Об этой закономерности можно судить по данным представленным в таблице и на графике. Однако измерить и оценить статистическую достоверность этой связи можно лишь при помощи специального коэффициента корреляции, который дает представление о размерах и характере связи между явлениями и признаками.

По характеру связь может быть: прямой и обратной.

При прямой связи - изменение одного признака идет в том же направлении, что и изменение другого признака.

Например: с увеличением роста ребенка увеличивается его вес. Прямая связь обозначается знаком плюс (+).

При обратной связи с изменением одного признака в одном направлении другой признак изменяется в противоположном направлении.

Например: с увеличением эффективности проводимых профилактических прививок, число инфекционных заболеваний уменьшается или другой пример: с увеличением температуры воздуха, уменьшается число бронхитов. Коэффциент корреляции, характеризующий обратную связь, обозначается знаком (-) .

По силе связи коэффициенты корреляции колеблются от единицы (полная связь) до нуля (отсутствие связи).

Сила связи	Прямая (+)	Обратная (-)
Полная	+1	-1
Сильная	+1; +0,7	-1; -0,7
Средняя	+0,7; +0,3	-0,7; -0,3
Слабая	+0,3; 0	-0,3; 0
Отсутствует связь	0	0

Корреляционная связь может быть прямолинейной и криволинейной.

Прямолинейяная связь характеризуется относительно равномерным изменением средних значений одного признака при равных изменениях другого (например: соответствие между изменениями уровней максимального и минимального артериального давления).

При криволинейной связи - равномерное изменение одного признака влечет за собой то возрастающие, то убывающие средние значения другого признака.

Коэффициент корреляции одним числом измеряет силу связи между изучаемыми явлениями и дает представление о ее направлении.

Существует несколько способов вычисления коэффициента корреляции. Разберем два из них:

1. Способ Спирмана (способ рангов)
2. Способ Пирсона.

1. Способ Спирмана или способ ранговой корреляции - является наиболее простым способом измерения связи, но менее точным. Этот способ применяется:

1)при небольшом числе наблюдений, т.е. когда $n \leq 30$

2)когда нет необходимости в точных расчетах, а нужны лишь ориентировочные сведения.

3)когда признаки имеют качественный характер.

Ро = 1 - $\underline{6\Sigma d^2}$

$n(n^2-1)$,

где Ро – коэффициент ранговой корреляции

1 и 6 – числа постоянные

d – разность ранговой

n – число наблюдений.

Этапы расчета коэффициента ранговой корреляции:

1.Составить вариационный ряд из парных признаков (х и у).

2.Распределить признаки в ранговом порядке.

3.Определить разность рангов $d = x1 - y1$

4.Возвести в квадрат разность рангов d^2

5.Получить сумму квадратов разности рангов - Σd^2

6.Определить Ро по формуле.

7.Определить ошибку репрезентативности (mp) коэффициента корреляции по формулу:

$$mp = \sqrt{\frac{1 - Po}{n - 2}}$$

8.Оценить достоверность коэффициента корреляции по формуле:

$$t = \frac{P}{mPo}$$

9. Сделать выводы.