ГОСУДАРСТВЕННЫЙ ЦЕНТРАЛЬНЫЙ
ОРДЕНА ЛЕНИНА ИНСТИТУТ ФИЗИЧЕСКОЙ
КУЛЬТУРЫ

Б.А. Суслаков, З.М. Баранова

СПОРТИВНАЯ МЕТРОЛОГИЯ

Вариационные ряды

Методические разработки для студентов
институтов физической культуры

Москва – 1980

ГОСУДАРСТВЕННЫЙ ЦЕНТРАЛЬНЫЙ ОРДЕНА ЛЕНИНА
ИНСТИТУТ ФИЗИЧЕСКОЙ КУЛЬТУРЫ

Б.А. Суслаков, Э.М. Баранова

Утверждены
Ученым советом ГЦОЛИФКа

СПОРТИВНАЯ МЕТРОЛОГИЯ

Вариационные ряды

Методические разработки для студентов
институтов физической культуры

Москва - 1980

Введение

Результаты спортивных измерений и наблюдений имеют разностороннюю направленность и носят самый разнообразный характер. Как измерения, так и наблюдения направлены на формирование качественного познания тренировочного процесса спортсмена и всего, что окружает и связано с этим.

Процесс получения или сбора информации в спорте может развиваться по направлениям, представленным на рис. I.

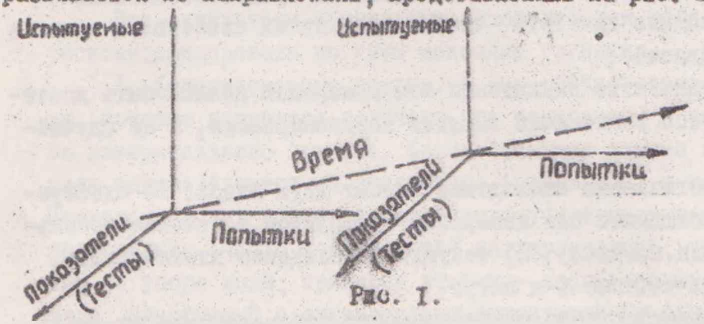


Рис. I.

Большинство спортивных измерений требует специальных методов анализа - методов математической статистики. Все эти методы работоспособны только в том случае, если измерения или наблюдения носят случайный характер. Например, при измерении роста у студентов (специализация - волейбол) мы не отбираем имеющих рост 182 ± 2 см, а измеряем этот показатель у каждого вновь прибывшего на обследование и таким образом получаем случайный набор результатов измерений.

Так как статистический подход к анализу данных объединяет большое число различных методик и приемов, отметим основные его этапы.

I. Статистическое наблюдение представляет собой планомерное научно обоснованное собирание данных, характеризующих изучаемые объекты или явления, к которым предъявляются следующие требования:

а) объекты должны быть однородными или одинаковыми с точки зрения их свойств (например, при обследовании спортсменов-борцов следует учесть возраст, спортивный стаж, вид борьбы, квалификацию, весовую категорию и др.;

б) объекты должны быть массовыми и представительными (это необходимо для того, чтобы обобщить их свойства и выявить характер связей);

в) количество наблюдений или измерений должно быть достаточным, чтобы можно было выявить закономерность, а не случайность.

Статистические наблюдения бывают двух видов: I) одновременные (наблюдение или измерение проводится в течение небольшого периода времени); 2) текущие (наблюдение длится долго, например, в течение 4-х лет).

По степени охвата статистические наблюдения могут быть: а) сплошными (например, социологическое обследование членов сборной команды СССР) и б) частичными (например, обследование при трудоемком и дорогостоящем лабораторном эксперименте).

II. Статистические сводка и группировка являются важным этапом подготовительной части к непосредственно статистическому анализу данных и предусматривают: а) систематизацию или группировку данных (например, распределение информации о спортсменах по подгруппам, различающимся по тем или иным показателям); б) оформление специальных, зависящих от цели исследования, статистических таблиц.

III. Анализ статистического материала - завершающий этап статистического исследования изучаемых объектов или явлений. Статистический анализ проводят, используя соответствующую математическую обработку исходных цифровых данных.

Все вышеизложенное позволяет заключить, что **математическая статистика** - **наука, исследующая количественные соотношения массовых явлений или процессов.** Выделяют два основных направления математической статистики:

I) описательная статистика, позволяющая упорядочить и привести к определенному стандартному виду результаты наблюдений или измерений;

2) проверяющая статистика, предназначенная для исследования и проверки статистических гипотез или предположений, а также построения приближенных статистических моделей с целью практического использования в спорте.

Классификация спортивных данных

Все результаты спортивных измерений или наблюдений можно расклассифицировать по трем основным группам.

I. Количественные данные. К ним обычно относят показатели, которые поддаются подсчету или получаются с помощью любого измерительного прибора. Количественные данные, в свою очередь, подразделяются на дискретные (прерывные) и непрерывные. Примером первых может служить число участников в массовых соревнованиях, число подтягиваний на перекладине, число отжиманий из упора лежа; примером вторых - любой размерный показатель, измеряемый в определенных фиксированных единицах с точными интервалами шкалы измерений: рост, вес, результат в беге на дистанции 100 м и др. Обычно дискретные показатели выражаются только целыми числами, а непрерывные - как целыми, так и дробными в зависимости от разрешающей способности измерительного прибора.

2. Качественные данные. К ним относят показатели, которые не имеют количественной оценки и не могут быть упорядочены или проранжированы при сравнении друг с другом. К таким можно отнести: цвет глаз, цвет волос, разновидности спортивных травм, разнообразие технических приемов в спортивных движениях, разнообразие спортивных снарядов и др. Также часто встречается в спортивных исследованиях случай, когда имеется только два класса наблюдаемого показателя. Например, успешная или неуспешная попытка, обучение или необучение, пол - мужской или женский и др. Эти показатели, являющиеся частным случаем качественных, называются альтернативными.

3. Порядковые данные. Такими показателями в спорте являются всевозможные спортивные достижения, оцениваемые в баллах или очках (фигурное катание, художественная и спортивная гим-

настика и др.), т.е. показатели, для которых точная количественная характеристика либо невозможна, либо нецелесообразна. Однако их можно расположить в определенном порядке или проранжировать.

Составление рядов распределения и их графические представления

В процессе наблюдения или измерения в спорте какого-либо показателя получают ряд цифровых значений. Обозначим эти значения как $x_1, x_2, x_3, \ldots, x_n$. Будем считать, что x_1 - результат измерения изучаемого показателя у I-го спортсмена, x_2 - результат 2-го и т.д. Как видно, всего спортсменов - n.

Такой ряд значений, представленный случайными числами, будем называть выборочной совокупностью или выборкой.

Совокупность всех значений, которые можно было бы получить для изучаемой выборки, называется генеральной совокупностью. Например, рост студентов ГЦОЛИФКа - выборочная совокупность, а рост студентов институтов физической культуры СССР - генеральная, но в то же время рост студентов институтов физической культуры СССР - выборка по отношению к генеральной совокупности всех студентов земного шара, активно занимающихся спортом.

Одной из основных характеристик выборки является ее объем, который всегда равняется n, т.е. числу спортсменов, показателей или попыток в данном исследовании.

Каким же образом проводится анализ выборки? Предположим, что мы измерили у студентов ГЦОЛИФКа, специализирующихся по баскетболу, силу левой кисти. Результаты измерений в килограммах (n = 100 чел.) представлены в табл. I.

В табл. I строка 2 составлена так, как проходили на измерении спортсмены, т.е. случайным образом, и поэтому она представляет собой неупорядоченную выборку; строка 3 - выборка упорядоченная, точнее ранжированная. Ранжированием называют расстановку значений выборки в порядке возрастания или убывания.

Таблица I

I.	I	2	3	4	5	6	99	I00
2. X (кг)	46	50	59	60	55	49	58	60
3. X (ранжир. ряд)	36	36	38	38	40	40 41......	70	74

В больших выборках ($n > 25$-30) всю упорядоченную выборку разбивают на интервалы. В выборках - спортивные результаты, в простейшем случае таких интервалов может быть два. Например: когда по результатам спортивных достижений нужно отобрать лучших и худших, всю упорядоченную выборку делят пополам. Интервалы в этом случае могут быть неравными. В медико-биологических исследованиях при больших выборках рекомендуется выбирать число интервалов в пределах 8-I2. Однако число интервалов можно рассчитать по формуле (правилу) Стургеса (Sturges, I926):

$$k = 1 + 3{,}32 \lg n \ ,$$

тогда величина или шаг интервала определяется как:

$$h = \frac{X_{max} - X_{min}}{k} \ ,$$

где X_{max} - максимальное значение в упорядоченной (ранжированной) выборке; X_{min} - минимальное значение.

Для нашего примера рассчитаем:

$$K = 1 + 3{,}32 \lg 100 = 1 + 3{,}32 \cdot 2 \simeq 8 \ ,$$

$$h = \frac{74 - 36}{8} = 4{,}75 \simeq 5.$$

Для справки число интервалов выборок различного объема следующее:

Объем	I0-I2	30-50	60-90	I00-200	300-400
Число интервалов	4	5-6	7	8	9

Рассчитав показатели K и h строим табл. 2.

Таблица 2

№ интервала	Границы интервала x	Частота m_i	Накопленная частота V_{m_i}	Относительная частота ω_i	Накопленная относительная частота V_{ω_i}	Частота, % $\omega(\%)$
	1	2	3	4	5	6
1.	35,5 - 40,5	7	7	0,07	0,07	7
2.	40,5 - 45,5	10	17	0,1	0,17	10
3.	45,5 - 50,5	19	36	0,19	0,36	19
4.	50,5 - 55,5	29	65	0,29	0,65	29
5.	55,5 - 60,5	18	83	0,18	0,83	18
6.	60,5 - 65,5	11	94	0,11	0,94	11
7.	65,5 - 70,5	5	99	0,05	0,99	5
8.	70,5 - 75,5	1	100	0,01	1	1

В данной таблице столбец 1 получается следующим образом: выбираем значение x_1, нижнюю границу 1-го интервала чуть меньше, чем x_{min} (из табл. 1) - 35,5 + 5, получаем верхнюю границу 1-го интервала - 40,5 (она же есть нижняя граница 2-го интервала), далее 40,5 + 5 = 45,5 и т.д.

Столбец 2 определяет частоту или встречаемость значений выборки в каждом интервале. Другими словами, частота определяется числом значений, попавших в данный интервал. Отметим, что в нашем примере $m_1 + m_2 + m_3 + \cdots + m_8 = n$.

Столбец 3 определяет накопленную частоту. Рассчитывается простым суммированием частот предыдущих интервалов: $V_{m_1} = m_1$;

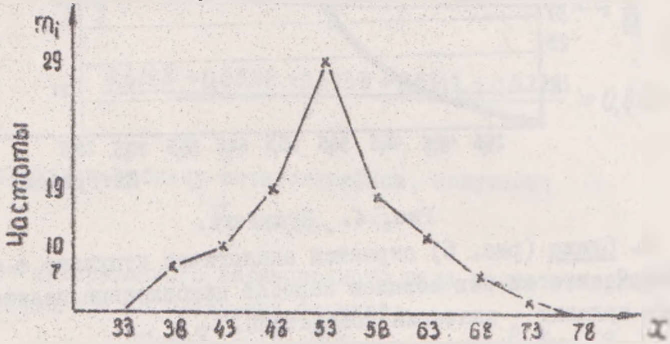
$$V_{m_2} = m_1 + m_2; \ldots, V_{m_8} = m_1 + m_2 + m_3 + m_4 + m_5 + m_6 + m_7 + m_8.$$

Столбец 4 определяет относительную частоту по формуле $\omega_i = \frac{m_i}{n}$. Столбец 5 определяет накопленную частоту и рассчитывается суммированием частот предыдущих интервалов. Столбец 6 есть столбец 4, выраженный в процентах.

Из табл. 2 можно определить, как часто каждое значение изучаемой выборки встречается в данной совокупности (столбцы 1 и 2). В результате получаем распределение результатов измерений или <u>вариационный ряд.</u>

Для вариационного ряда наиболее часто используются следующие графические представления.

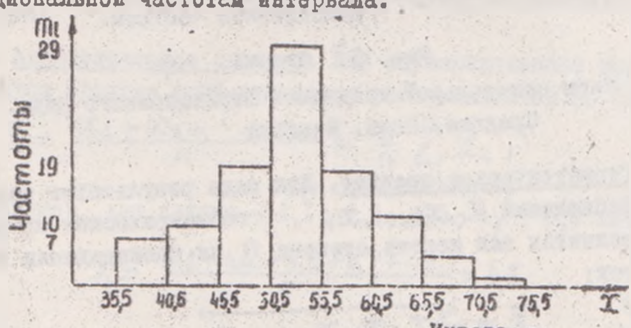
1. Полигон распределения (рис. 2) строится в прямоугольной системе координат. Величины измеряемого показателя откладывают на оси абсцисс, частоты на оси ординат.



Величина показателя - середина интервалов

Рис. 2. Полигон распределения.

2. Гистограмма распределения (рис. 3) строится аналогично полигону распределения, однако на оси абсцисс откладываются не точки - (средины интервалов), а отрезки, изображающие интервал, а вместо ординат, соответствующих частотам или частостям отдельных вариантов, строят прямоугольники с высотой, пропорциональной частотам интервала.



Интервалы

Рис. 3. Гистограмма.

3. Кумулята (кривая сумм) (рис. 4) строится в прямоугольной системе координат. На оси ординат откладывают отрезки, длина которых пропорциональна накопленной частоте того или иного интервала. На абсциссе наносят значения показателя.

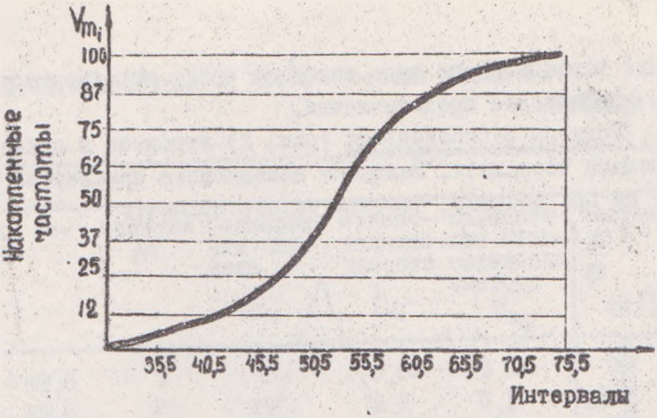


Рис. 4. Кумулята.

4. Огива (рис. 5) строится аналогично кумуляте с той лишь разницей, что на ось абсцисс наносят накопленные частоты, а на ось ординат - значения показателя.

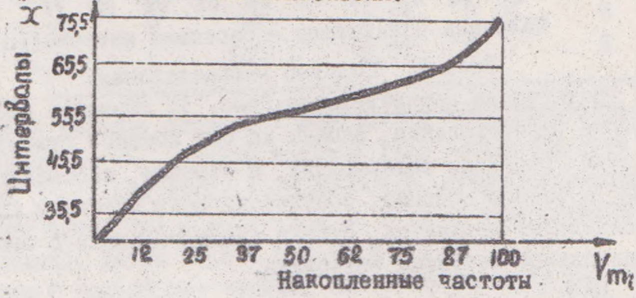


Рис. 5. Огива.

Меры центральной тенденции вариационного ряда. Средние. Мода. Медиана

I. Геометрическое среднее. Для ряда результатов измерений или наблюдений $x_1, x_2, \ldots, x_n$ геометрическое среднее $\bar{x}_g$ определяется как корень степени n из произведения этих результатов:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} .$$

Для удобства вычисления преобразуем (прологарифмируем) это выражение:

$$\lg \bar{x}_g = \frac{1}{n}\left(\lg x_1 + \lg x_2 + \cdots + \lg x_n\right) = \frac{1}{n}\sum_{i=1}^{n} \lg x_i$$

Например, для обычного неупорядоченного ряда – 4,1; 4,5; 4,4; 4,8; 4,7 – геометрическое среднее рассчитывается следующим образом:

$$\lg \bar{x}_g = \frac{\lg 4{,}1 + \lg 4{,}5 + \lg 4{,}4 + \lg 4{,}8 + \lg 4{,}7}{5} =$$

$$= \frac{0{,}6128 + 0{,}6532 + 0{,}6435 + 0{,}6812 + 0{,}6721}{5} = 0{,}6524$$

Используя таблицу антилогарифмов, получаем:

$$\bar{x}_g = 4{,}48 .$$

Для интервального вариационного ряда геометрическое среднее вычисляется по следующей формуле:

$$\lg \bar{x}_g = \frac{m_1 \cdot \lg \bar{x}_{g_1} + m_2 \cdot \lg \bar{x}_{g_2} + \cdots + m_k \cdot \lg \bar{x}_{g_k}}{m_1 + m_2 + \cdots + m_k} ,$$

где $m_1, m_2, \ldots, m_k$ – значения частот соответствующих интервалов, которых всего K, $\bar{x}_{g_1}, \bar{x}_{g_2}, \ldots, \bar{x}_{g_k}$ – средние геометрические тех же интервалов.

Геометрическое среднее используют в случаях, когда результаты измерений или наблюдений пропорционально убывают или возрастают, а также варьируют по некоторому синусоидальному закону.

2. Арифметическое среднее. Для неупорядоченной выборки результатов среднее арифметическое рассчитывается по формуле:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Например, для данных – 4,1; 4,4; 4,5; 4,7; 4,8 вычисляем:

$$\bar{x} = \frac{4{,}1 + 4{,}4 + 4{,}5 + 4{,}7 + 4{,}8}{5} = 4{,}5 .$$

Для интервального вариационного ряда среднее арифметическое вычисляется по следующей формуле:

---

x) Символ $\sum_{i=1}^{n} x_i$ ($\sum$ – греческая буква "сигма") обозначает суммирование всех результатов x_i, когда i – принимает значения 1, 2, 3, 4, ..., n, т.е. можно записать, что
$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$.

$$\bar{X} = \frac{m_1 \cdot \bar{X}_1 + m_2 \cdot \bar{X}_2 + \dots + m_k \cdot \bar{X}_k}{m_1 + m_2 + \dots + m_k} = \frac{\sum_{i=1}^{k} m_i \cdot \bar{X}_i}{\sum_{i=1}^{k} m_i} = \frac{\sum_{i=1}^{k} m_i \cdot \bar{X}_i}{n},$$

где $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ - средние арифметические соответствующих интервалов.

Средняя арифметическая используется для анализа дискретных и непрерывных данных как характеристика, отображающая некоторую среднюю оценку изучаемой выборки.

3. Мода. Модой называют результат выборки или совокупности наиболее часто встречающийся в этой выборке. Для интервального вариационного ряда модальный интервал выбирается по наибольшей частоте. Мода вычисляется по формуле:

$$M_0 = x_{M_0(min)} + h \frac{m_{M_0} - m_{M_0-1}}{2 m_{M_0} - m_{M_0-1} - m_{M_0+1}},$$

где $x_{M_0(min)}$ - нижняя граница модального интервала; h - величина или шаг интервала; m_{M_0} - частота модального интервала; m_{M_0-1} - частота интервала, предшествующего модальному; m_{M_0+1} - частота интервала следующего за модальным:

Для данных табл. 2 вычислим значение M_0:

$$M_0 = 50{,}5 + 4{,}75 \frac{29 - 19}{2 \cdot 29 - 19 - 18} = 52{,}76.$$

4. Медиана. Медиана - результат выборки или совокупности, который находится в середине ранжированного ряда. Для вычислений медианы интервального ряда вначале находят медианный интервал, которому соответствует первая из накопленных частот, превышающая половину всего объема выборки, а потом определяют приближенное значение медианы по следующей формуле:

$$M_e = x_{Me(min)} + h \frac{n/2 - V_{m(Me-1)}}{m_{Me}},$$

где $x_{Me(min)}$ - нижняя граница медианного интервала, $V_{m(Me-1)}$ - накопленная частота, предшествующая медианному интервалу, m_{Me} - частота медианного интервала.

Для данных табл. 2 вычислим значение Me :

$$Me = 50{,}5 + 4{,}75 \; \frac{50 - 36}{29} = 52{,}8 .$$

Моду и медиану иногда называют непараметрическими средними. Из этого следует, что их можно использовать в различных распределениях выборок как для качественных, так и для количественных данных.

Меры вариации или колеблемости вариационного ряда. Размах. Дисперсия. Коэффициент вариации. Квартили. Децили

Часто исследование результатов спортивных измерений ограничивается расчетом средней величины. Однако в ряде случаев одна и та же средняя может характеризовать совершенно различные выборки.

Например, результаты измерения динамометрии левой кисти у баскетболистов имеют следующие обобщающие характеристики:

Лучший (максимальный) результат: $x_{max} = 74$ кГ.

Худший (минимальный) результат: $x_{min} = 36$ кГ.

Среднее арифметическое выборки: $\bar{x} = 53{,}32$ кГ.

Для выборки, характеризующей группу такого же объема, но различной специализации, эти характеристики следующие:

$x_{max} = 82$ кГ; $x_{min} = 38$ кГ; $\bar{x} = 53{,}78$ кГ.

Из сравнения этих выборок видно, что если средние значения различаются мало, то максимальные и минимальные - различаются существенно. Следовательно, знание только средних - недостаточно. Необходимо учитывать вариацию или колеблемость результатов измерений. Для этого существуют следующие статистические характеристики.

I. Размах варьирования определяется как разность между максимальным и минимальным результатами измерения. Например, для группы баскетболистов: $R_1 = x_{max} - x_{min} = 74 - 36 = 38$; для группы различных специализаций: $R_2 = 82 - 38 = 44$.

Однако размах варьирования дает самое общее представление о колеблемости, так как показывает лишь различия между крайними значениями результатов измерений. Для этого вводят характеристику вариации, которая учитывает отклонение каждого результата от средней величины данной совокупности или выборки.

2. Дисперсия определяется как средняя сумма квадратов отклонений отдельных результатов от средней арифметической и обозначается - σ^2 ("сигма малая").

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

Если число испытуемых мало, т.е. $n < 30$, применяется формула:

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

Эту формулу используют, когда результаты измерения представлены неупорядоченной (обычной) выборкой. Например, для первых шести результатов из табл. I расчет проводится следующим образом.

Таблица 3

№ п/п	x	$(x-\bar{x})$	$(x-\bar{x})^2$
I.	46	46 - 53,I6 = -7,I6	$(-7{,}I6)^2$ = 5I,26
2.	50	50 - 53,I6 = -3,I6	$(-3{,}I6)^2$ = 9,98
3.	59	59 - 53,I6 = 5,84	$5{,}84^2$ = 34,I0
4.	60	60 - 53,I6 = 6,84	$6{,}84^2$ = 46,78
5.	55	55 - 53,I6 = I,84	$I{,}84^2$ = 3,38
6.	49	49 - 53,I6 = -4,I6	$(-4{,}I6)^2$ = I7,30

$\sum x = 319$ $\sum(x-\bar{x})^2 = 162{,}83$

$$\bar{x} = \frac{\sum x}{n} = \frac{319}{6} = 53{,}16; \quad \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{162{,}83}{5} = 32{,}56 .$$

Для выборки $n > 30$ также применяют формулу расчета σ^2 следующего вида:

$$\sigma^2 = \overline{x^2} - \bar{x}^2 ,$$

где $\overline{X^2}$ – среднее арифметическое квадратов результатов измерений, $\overline{X}^2$ – квадрат среднего арифметического.

Как видно из определения, дисперсия имеет единицу измерения – единица измерения результата, возведенная в квадрат. На практике часто требуется указать интервалы варьирования среднего значения. Для этого используют среднее квадратическое отклонение, которое определяется как положительное значение корня квадратного из дисперсии, т.е.:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}, \quad (n < 30).$$

В случае интервального вариационного ряда среднее квадратическое отклонение вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2 \cdot m_i}{\sum_{i=1}^{k} m_i}},$$

где K – число интервалов, $\bar{X}_i$ – среднее арифметическое каждого интервала.

Так как среднее квадратическое отклонение имеет те же единицы измерения, что и результат, то для сравнения двух выборок (по их колеблемости) с различными единицами измерения оно (σ) не пригодно. В таких случаях вычисляют следующую меру колеблемости.

3. Коэффициент вариации. Определяется как отношения среднеквадратического отклонения, выраженного в процентах к средней арифметической:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% .$$

Например, коэффициент вариации результатов динамометрии левой кисти баскетболистов вычисляется так:

$$V = \frac{\sqrt{32.56}}{53.15} \cdot 100\% = 10,66\% .$$

В спортивных исследованиях колеблемость результатов измерений в зависимости от величины коэффициента вариации считают небольшой (0%-10%), средней (11%-20%) и большой ($V > 20\%$).

Коэффициент вариации часто используется в спортивных задачах, так как будучи величиной относительной (измеряется в процентах) он позволяет сравнивать между собой колеблемость показателей или признаков, имеющих различные единицы измерения.

Определенное место среди порядковых статистических характеристик, которые имеют важное значение при построении различного рода шкал, занимают квартили, децили и процентили.

4. Квартили. Медиана делит выборку или совокупность пополам, а квартили делят на четыре части. Различают две квартили: нижнюю и верхнюю.

В интервальном вариационном ряду квартили рассчитываются по следующим формулам:

нижний квартиль: $$Q_1 = x_{q_1(min)} + h\frac{n/4 - V_{m(q_1-1)}}{m_{q_1}};$$

верхний квартиль: $$Q_3 = x_{q_3(min)} + h\frac{n/3/4 - V_{m(q_3-1)}}{m_{q_3}},$$

аналогично рассчитываются и два средних квартиля.

$$Q_2 = x_{q_2(min)} + h\frac{n/2 - V_{m(q_2-1)}}{m_{q_2}}.$$

В приведенных формулах используются обозначения:

$x_{q_i(min)}$ - нижняя граница квартильного интервала, соответственно первого, второго и третьего;

h - шаг (величина) интервала исследуемого вариационного ряда;

$V_{m(q-1)}$ - накопленная частота интервала, предшествующего квартильному;

m_q - частота квартильного интервала.

Например, для данных табл. 2 рассчитаны все три значения квартилей:

$$Q_1 = x_{q_1(min)} + h\frac{n/4 - V_{m(q_1-1)}}{m_{q_1}} = 45{,}5 + 4{,}75\,\frac{25 - 17}{19} = 47{,}5;$$

$$Q_2 = x_{q_2(min)} + h\,\frac{n/2 - V_{m(q_2-1)}}{m_{q_2}} = 50{,}5 + 4{,}75\,\frac{50-36}{29} = 52{,}8\,;$$

$$Q_3 = x_{q_3(min)} + h\,\frac{n/3/4 - V_{m(q_3-1)}}{m_{q_3}} = 55{,}5 + 4{,}75\,\frac{75-65}{18} = 58{,}14\,;$$

5. <u>Децили</u> делят всю выборку на десять частей и рассчитываются по следующим формулам:

$$D_1 = x_{A_1(min)} + h\,\frac{1/10 \cdot n - V_{m(A_1-1)}}{m_{A_1}}\,;$$

. . . .

$$D_9 = x_{A_9(min)} + h\,\frac{9/10 \cdot n - V_{m(A_9-1)}}{m_{A_9}}\,.$$

Обозначения в формулах аналогичны обозначениям в формулах квартилей. Для примера на основании данных табл. 2 рассчитаем первые три значения децилей:

$$D_1 = x_{A_1(min)} + h\,\frac{1/10 \cdot n - V_{m(A_1-1)}}{m_{A_1}} = 40{,}5 + 4{,}75\,\frac{10-7}{10} = 41{,}92\,;$$

$$D_2 = x_{A_2(min)} + h\,\frac{2/10 \cdot n - V_{m(A_2-1)}}{m_{A_2}} = 45{,}5 + 4{,}75\,\frac{20-17}{19} = 46{,}25\,;$$

$$D_3 = x_{A_3(min)} + h\,\frac{3/10 \cdot n - V_{m(A_3-1)}}{m_{A_3}} = 45{,}5 + 4{,}75\,\frac{30-17}{19} = 48{,}75$$

и т.д.

Аналогично можно рассчитать значения процентилей (перцентилей), которые делят всю выборку на сто частей.

Меры положения распределения. Асимметрия. Скошенность. Эксцесс

Кривая любого эмпирического распределения, полученная в результате измерения показателя или признака у некоторой группы испытуемых, только в редких случаях представляется идеальной колоколообразной и симметричной (кривой нормального распределения). Для многих распределений характерны смещения кривой вправо или влево. Поэтому различают левостороннюю (рис. 6) и правостороннюю (рис. 7) асимметрию.

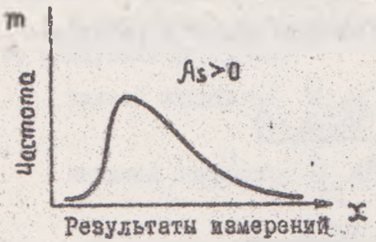


Рис.6. Левосторонняя асимметрия.

Рис.7. Правосторонняя асимметрия.

<u>Асимметричность кривой</u> распределения характеризуют показателями асимметрии и скошенности. Показатель асимметрии рассчитывают по следующей формуле:

$$As = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sigma^3}.$$

Чем больше абсолютная величина этого показателя, тем большую скошенность и асимметричность имеет кривая распределения результатов измерений. Если знак показателя отрицательный, асимметрия - правосторонняя, а если положительный - левосторонняя. Например, сравним показатели двух выборок. Результаты динамометрии левой кисти баскетболистов и общей группы: баскетбол - As = 0,089; общая - As = 0,815.

Видно, что распределение этого показателя имеет большую левостороннюю скошенность для общей группы. Следовательно, в этой выборке преобладают слабые результаты по сравнению с группой баскетболистов.

Для оценки асимметрии и скошенности К.Пирсон предложил более простую характеристику – меру скошенности. В основу этой меры положено отклонение среднего арифметического значения от моды:

$$S_K = \frac{\bar{x} - M_o}{\sigma}.$$

Для выборки (табл. 2) скошенность вычисляется, как:

$$S_K = \frac{53{,}32 - 52{,}80}{7{,}94} = 0{,}065.$$

Как видно, это значение S_K немного отличается от A_S (0,089) для этой же выборки. Свойства меры скошенности аналогичны свойствам асимметрии.

Эксцесс. Кривые распределения разделяют на плосковершинные и островершинные (рис. 8).

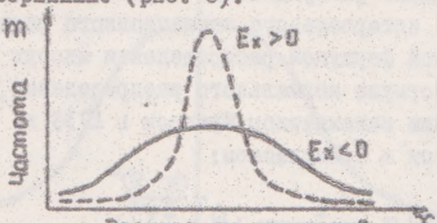


Рис.8. Плосковершинная и островершинная кривые распределений.

Эти свойства кривой распределения характеризуют величиной эксцесса, которую рассчитывают по следующей формуле:

$$E_x = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{\sigma^4} - 3.$$

Если знак эксцесса отрицательный, то имеется тенденция к плосковершинности и наоборот. Сравним эти характеристики для двух рассмотренных выше выборок:

баскетболисты – E_x = –0,214; общая группа – E_x = 0,787.

Выраженная плосковершинность кривой распределения динамометрии баскетболистов говорит о более высокой однородности (одинаковости) этой группы по отношению к другой по изучаемому показателю.

Рассмотренные меры положения распределения имеют важное значение при изучении спортивных показателей или признаков, так как они позволяют оценить эмпирическое распределение и сравнить его с некоторыми видами теоретических распределений.

К наиболее широко используемым распределениям относят нормальный закон распределения, распределение Пуассона, биномиальное распределение и некоторые другие.

Нормальный закон распределения

Как уже было показано выше (с. 9), вариационный ряд графически представляется в виде полигона распределения и гистограмм. При исследовании результатов измерения или наблюдения бывает необходимо оценить степень приближения эмпирического распределения к теоретическому. Обычно под теоретическим распределением понимают распределение вероятностей (а не частот), лежащих в основе интервального вариационного ряда.

Теоретической формулой распределения многих вариационных рядов является формула нормального распределения, которое было найдено английским математиком Муавром в 1733 г. и позднее уточнено К.Гауссом и П.Лапласом:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},$$

где π и e – математические постоянные, соответственно $\pi = 3,141$ и $e = 2,718$; $\bar{x}$ и σ – соответственно среднее арифметическое и среднее квадратическое отклонения; x – результаты измерения или наблюдения; $f(x)$ – обычно называют функцией плотности распределения.

Это уравнение позволяет получить кривую нормального распределения, которая изображена на рис. 9. Также очевидно, что центр группирования частот и форма нормальной кривой определяются $\bar{x}$ и σ.

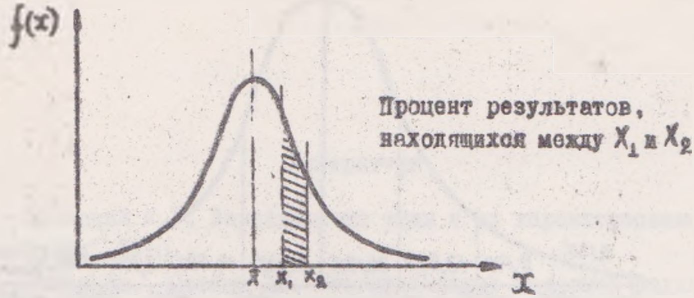


Рис.9. Кривая нормального распределения.

Введя обозначение $u = \frac{x - \bar{X}}{б}$, которое называют нормированным или стандартизованным отклонением, получают выражение для нормированного нормального распределения (рис. 10).

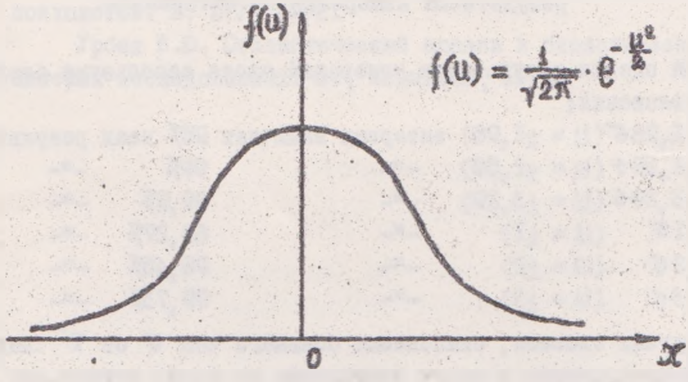


Рис. 10. Кривая нормированного нормального распределения.

Этот график интересен тем, что для него $\bar{X} = 0$ и $б = 1$. Вся площадь, заключенная под кривой, равна I, т.е. все 100% результатов измерений учтены для данной выборки. В теории спортивных тестов и построении шкал оценок представляет интерес процентное содержание результатов измерения в зависимости от размаха варьирования. Эти данные представлены на графике (рис. II).

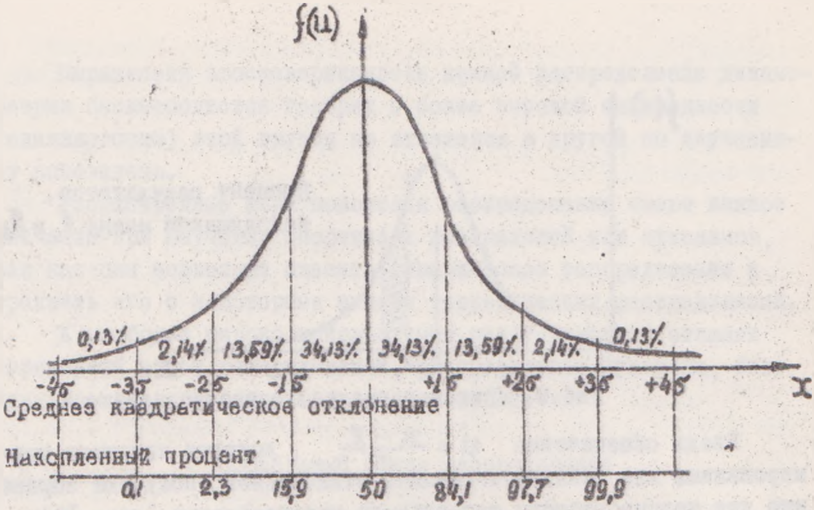


Рис. 11. Процентное выражение распределений относительных и накопленных частот результатов измерений по интервалам.

Для оценки результатов измерений часто используют следующие соотношения:

$\bar{X} \pm 1{,}96\sigma$	$(u = \pm 1{,}96)$	интервал включает	95%	всех результатов,
$\bar{X} \pm 2{,}58\sigma$	$(u = \pm 2{,}58)$	-"-	99%	-"-
$\bar{X} \pm 3{,}29\sigma$	$(u = \pm 3{,}29)$	-"-	99,9%	-"-
$\bar{X} \pm 1\sigma$	$(u = \pm 1)$	-"-	68,27%	-"-
$\bar{X} \pm 2\sigma$	$(u = \pm 2)$	-"-	95,45%	-"-
$\bar{X} \pm 3\sigma$	$(u = \pm 3)$	-"-	99,73%	-"-

Другими словами, отклонения большего чем σ от $\bar{X}$ следует ожидать примерно в одном результате из трех; отклонения, большего чем 2σ - в одном из 22-х и отклонения, большего чем 3σ - в одном из примерно 370. Последнее соотношение для нормального распределения часто называют "правилом трех сигм" и используют его при исключении сильно отклоняющихся "ошибочных" результатов измерений.

В заключение отметим, что нормальная кривая близка по многим одновершинным распределениям, которые часто рассматриваются как нормальные. Особая ценность нормального закона распределения состоит в том, что для этого распределения более подробно разработан целый ряд приемов математико-статистической обработки результатов измерений или наблюдений.

Литература

Венецкий И.Г. Вариационные ряды и их характеристики. М., "Статистика", 1970.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., "Прогресс", 1976.

Липкин М.И. Кривые распределения в экономических исследованиях. М., "Статистика", 1972.

Масальгин Н.А. Математико-статистические методы в спорте. М., ФиС, 1972.

Суходольский Г.В. Основы математической статистики для психологов. Л. ЛГУ, 1972.

Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. М., "Медицина", 1975.

Содержание

СПОРТИВНАЯ МЕТРОЛОГИЯ

Вариационные ряды

Методические разработки для студентов институтов физической культуры

Редактор Н. Масленникова Корректор Т. Арсланова

Объем 0,86 уч.-изд.л. Тираж IOOO экз.
Зак. 979/1339
Издание Редакционно-издательского отдела ГЦОЛИФКа.

Типография В/о "Союзспортобеспечение"
Москва, Мичуринский проспект, 40.