Бобин Виталий Леонидович
учитель
МБОУ «СОШ №37»
г. Чебоксары, Чувашская Республика
Бобин Дмитрий Витальевич
старший преподаватель
ФГБОУ ВО «Чувашский государственный
университет им. И.Н. Ульянова»
г. Чебоксары, Чувашская Республика
Кулагина Алевтина Григорьевна
канд. экон. наук, доцент
ФГБОУ ВО «Чувашский государственный
университет им. И.Н. Ульянова»
г. Чебоксары, Чувашская Республика

ИЗУЧЕНИЕ НОРМАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ В ШКОЛЕ

Аннотация*: в статье представлены результаты проектного изучения раздела «теории вероятностей и математической статистики» на уроках математики. В качестве выборочных данных для анализа выступают сведения о собственном росте учеников. Методологической базой исследования являются статистические методы сбора и обобщения данных. Средний объем выборочной совокупности обеспечивает достаточную надежность проверяемой гипотезы и полученных оценок.*

Ключевые слова*: математика, теория вероятностей, статистика, числовые характеристики, выборочная совокупность, нормальный закон распределения.*

Одна из составных задач статистики – это сбор данных в определенной группе с целью выявить закономерности их распределения. При этом данные организуются особым образом, визуализируются для удобства и затем непременно интерпретируются.

Всю группу исследуемых объектов (событий) называют популяцией или генеральной совокупностью. Когда из этого множества отбирают для изучения или проверки некоторой статистической гипотезы часть объектов, то это называют выборочной совокупностью. Например, мы можем отобрать несколько учеников и измерить их рост. Однако при этом необходимо придерживаться определенных правил, чтобы результаты исследования были адекватными. А при достаточно большом объеме выборки мы уже можем говорить о возникновении законов распределения. Закон распределения описывает как данные распределяются между разными значениями на числовой оси. Понимание распределений крайне важно не только в статистике, но и во многих современных видах деятельности, включая аналитику данных, информатику и программирование.

На практике наиболее часто встречающимся распределением является нормальное (рис. 1). Это такое распределение точек на числовой оси, при котором вероятность того, что случайная величина окажется около центра гораздо выше, чем того, что она сильно отклонится от середины. Многие природные явления подчиняются ему, например, рост людей или животных, их вес, цены на рынке и т. д.

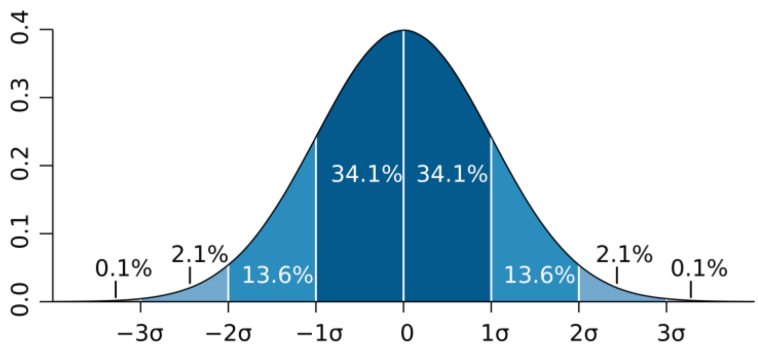

Рис. 1. Кривая нормального закона распределения

Главная особенность, выделяющая нормальный закон среди других законов, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях [2].

Для демонстрации обучающимся 11 класса действия нормального закона распределения в повседневности воспользуемся данными об их росте. Также формально выполним проверку статистической гипотезы.

Для достижения цели поставлены следующие задачи:

1) собрать статистические данные о росте сверстников;

2) найти числовые характеристики показателя роста;

3) построить закон распределения случайной величины;

4) наглядно представит статистическую информацию;

5) обобщить результаты.

Статистические данные некоторой совокупности для начала группируются в таблицах. Например, мы определяем рост в сантиметрах в группе из 61 ученика 11-го класса. Запишем округленные до целых значения роста учеников в виде вариационного ряда (табл. 1).

Таблица 1

Вариационный ряд

Рост, см	152	157	158	159	160	162	163	164	165	166	167	168	169	170
Количество	1	1	3	1	1	3	1	1	5	2	3	5	2	5
Рост, см	171	172	174	175	176	177	179	180	181	182	183	184	185	186
Количество	3	2	1	5	2	2	1	4	2	0	1	1	2	1

Понятие «Середины» или среднего значения закрепилось в статистике. Вероятно, потому что оно даёт нам представление о том, что есть «норма» и насколько мы от нее отклонились. Мы говорим о средней скорости автомобиля, о средних оценках или о среднем балле за всё время обучения в школе и т. д. Среднее арифметическое для наших наблюдений равно:

$$m = (152\cdot 1+157\cdot 1+158\cdot 3+ \ldots +184\cdot 1+185\cdot 2+186\cdot 1)/61 = 170{,}4 \text{ (см)}.$$

А дисперсия, характеризующая разброс наблюдений относительно среднего значения, равна:

$$\sigma 2 = (1522\cdot 1+1572\cdot 1+ \ldots +1852\cdot 2+1862\cdot 1)/61 - 170{,}42 = 61{,}2 \text{ (см2)}.$$

Рост учеников является непрерывной случайной величиной. Поэтому логично будет представить ее в виде интервалов. Вычисляем оптимальное количество интервалов по формуле Стерджесса:

$$k = 1+3{,}322\cdot\lg 61 \approx 7.$$

Оптимальное количество интервалов равно 7. Размах данных, – разница между наибольшим и наименьшим значениями, – составляет 34 (см). Для удобства установим метки на значениях роста, кратных 5: 150, 155, …, 185, 190. В таком случае получится 8 интервалов одинаковой ширины 5 см. Составим интервальное распределение выборки (табл. 2) с указанием получившихся интервалов и соответствующих им частот.

Таблица 2

Интервальное распределение

i	Интервал	Середина интервала	Эмпирическая частота	Относительная частота
1	[150; 155]	152,5	1	0,0164
2	(155; 160]	157,5	6	0,0984
3	(160; 165]	162,5	10	0,1639
4	(165; 170]	167,5	17	0,2787
5	(170; 175]	172,5	11	0,1803
6	(175; 180]	177,5	9	0,1475
7	(180; 185]	182,5	6	0,0984
8	(185; 190]	187,5	1	0,0164
Сумма			61	1

Как известно, при больших выборках относительная частота подмножества приблизительно равна их вероятности [1]. Так, значения в таблице следует интерпретировать следующим образом: у 10 учеников из выборки рост составляет от 160 до 165, т.е. вероятность того, что выбранный наугад ученик 11 класса, имеет рост в этом интервале, равна 0,0984.

Теоретические частоты нормального распределения роста учеников найдем по формуле (1):

$$\frac{h}{\sigma\sqrt{2\pi}}e^{-\frac{(x_i-m)^2}{2\sigma^2}}, (1)$$

где h – длина интервала, m – средняя величина по выборке, σ – стандартное отклонение, xi – середина i-го интервала.

Нарисуем графики кривых эмпирических и теоретических частот распределения роста учеников (рис. 2).

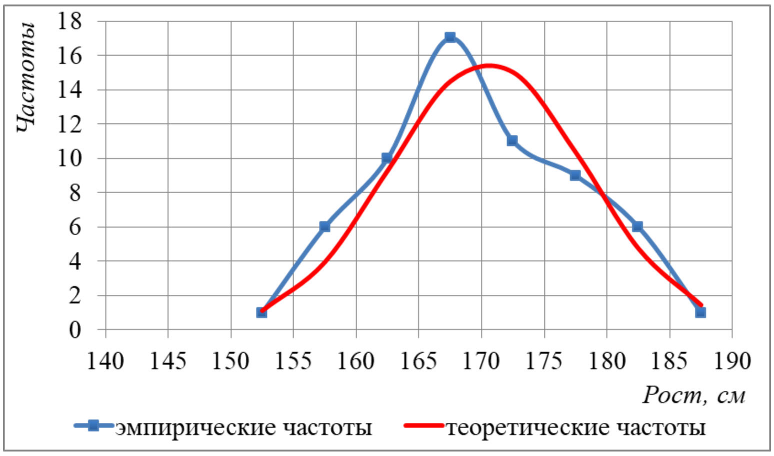

Рис. 2. Графики эмпирических и теоретических частот распределения

Гипотезу о нормальном распределении роста учеников при уровне значимости 0,05 проверим с помощью критерия хи-квадрат Пирсона. Расчетное значение статистики рассчитаем по формуле (2):

$$\chi^2 = \sum_{i=1}^{k=8} \frac{(n_i'-n_i)^2}{n_i} = 3{,}27, \quad (2)$$

где i – номер интервала, k – количество интервалов, n_i' – теоретические частоты, n_i – эмпирические частоты.

Его нужно сравнить со значением из специальной таблицы распределения хи-квадрат при уровне значимости 0,05 и степенях свободы (8–3=5): $\chi_{\text{т}}^2 = 11$. Так как расчетная статистика не превышает табличного значения, то нет оснований отвергнуть гипотезу о нормальном распределении роста учеников.

В результате проведенной работы ученики узнали, что рост и вес учащихся со временем соответствует колоколообразной кривой. Это значит, что гипотеза о нормальном распределении роста учащихся истинна. При этом ученики научились собирать и обрабатывать статистическую информацию, представлять ее в графическом формате, находить статистические характеристики выборки.

Список литературы

1. Бродский И.Л. Вероятность и статистика. 10–11 классы. Планирование и практикум / И.Л. Бродский, О.С. Мешавкина. – М.: Аркти, 2009. – 104 с.

2. Вентцель Е.С. Теория вероятностей: учебник / Е.С. Вентцель. – 12-е изд., стер. – М.: Юстиция, 2018. – 658 с.