Севрюк А.П.

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

ГУО «Институт бизнеса Белорусского государственного университета», г. Минск

Аннотация: в статье рассматривается суть нормального закона распределения. На основе проведенного исследования приводится доказательство тому, что нормальное распределение действительно существует в окружающем нас мире и находит применение в различных отраслях науки и техники.

Ключевые слова: нормальное распределение, закон распределения, случайная величина.

Нормальный закон распределения без преувеличения можно назвать философским законом. Наблюдая за различными процессами и объектами в окружающем нас мире, мы часто сталкиваемся с тем, что существует «основная масса», которая соответствует тому или иному признаку и существуют отклонения в обе стороны.

Изучение этого распределения является **актуальным**, потому что такие базовые вещи как рост, вес, физическая сила людей, время вскипания чайника или забега стометровки, все это распределяется в соответствии с данным законом распределения.

В этой связи мною были определены следующие **цели** – изучить суть нормального закона распределения и проверить действительно ли он существует на примерах из окружающего нас мира.

Для достижения целей были использованы следующие **методы**: поиск информации в различных источниках, таких как научные статьи, книги, интернет ресурсы и ее анализ, проведение соцопроса и математических расчетов на его основе.

В школе нам говорили, что завод выпускает столы длиной ±1м. В университете начали говорить, что завод выпускает столы длиной 1м ±1см из-за погрешности изготовления. На самом деле завод выпускает столы длиной 1м ±1см с вероятностью 95%. Т.е. учитывается то, что длина стола является случайной величиной. Случайная величина – это величина, которая существенно зависит от внешних факторов или начального состояния системы, при этом их влияние мы не можем учесть в условиях данного эксперимента.

Длина стола - это непрерывная величина, которая имеет бесконечное количество значений. При этом вероятность появления каждого значения будет равна 0. И это очень просто показать на примере. Мы не можем найти двух столов одинаковой длины. Допустим, мы нашли два стола, которые имеют длину 1м и 1мм, все равно мы можем выбрать такую точность измерения, что их длина будет отличаться на микрометр, нанометр, пикометр и т.д. Несмотря на это, если завод выпускает столы длиной 1 метр, то вероятность, что стол будет иметь длину от 1м до 1м 1мм существенно выше, чем то, что стол будет иметь длину от 1м 10мм до 1м 11мм. Если мы возьмем все столы, которые выпустил завод, найдем среди них самый маленький и самый большой, после чего диапазон длин столов мы разделим на интервалы равной величины, после этого мы сможем посчитать, сколько столов попало в каждый интервал. Количество столов, попавших в каждый интервал, зависит от общего количества столов, которые выпустил завод. Поэтому работать с этой величиной не очень удобно и лучше перейти к частоте. Чтобы рассчитать частоту, нужно взять и разделить частоту столов, попавших в определенный интервал на общее количество столов, тогда сумма частот появления в каждом интервале будет равна единице.

График, на одной оси которого отложены длины сторон, а на второй – частоты их появления в каждом интервале, называется гистограммой. Для хорошей оценки случайной величины очень важен выбор количества интервалов. Если выбрать меньше интервалов, то будет получено мало информации о случайной величине, если выбрать больше интервалов, то будет сложно ее анализировать из-за наличия шумов. Количество интервалов можно рассчитать по следующей формуле: $M = \log_2 N + 1$, где N – размер выборки. Если размер выборки стремится к бесконечности, тогда частота появления в каждом интервале равна вероятности появления в этом интервале. Если же при этом ширина интервала стремиться к нулю, тогда мы получаем непрерывную функцию, которая стремиться к теоретической функции распределения.

Когда вспоминают о функциях распределения, чаще всего на ум приходит график функции нормального распределения. Возникает вопрос, почему же он так часто встречается? Дело в том, что в жизни случайные величины часто зависят от большого количества факторов, при этом факторы имеют неизвестную природу. А по центральной предельной теореме сумма большого количества величин с неизвестным распределением имеет распределение близкое к нормальному. Формула функции распределения для нормального распределения имеет следующий вид $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ $(\sigma > 0)$. Стоит отметить, что она зависит всего лишь от двух параметров: математического ожидания (μ) и дисперсии (D). Чтобы рассчитать эти параметры можно воспользоваться двумя способами. Первый: рассчитать эти значения теоретически, на основе знания о процессе. Второй: рассчитать на основе бесконечной выборки, но получить выборку бесконечной длины мы не можем, поэтому мы пользуемся оценками параметра. Оценка – это число, вычисляемое на основе наблюдений, которое предположительно близко к оцениваемому параметру.

Математическое ожидание равно среднему арифметическому бесконечного количества значений случайной величины, однако, мы будем пользоваться оценкой, которую будем рассчитывать просто как среднее значение всей выборки $\mu = \frac{\sum_{i=1}^{N} Xi}{N}$. Математическое ожидание показывает расположение кривой нормального распределения на графике. Дисперсия характеризует разброс случайной величины относительно ее математического ожидания. Значение дисперсии очень сложно интерпретировать, поэтому часто пользуются значением среднеквадратичного отклонения (σ), которое равно корню из дисперсии и находится по следующей формуле $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(Xi-\mu)^2}{N-1}}$ и показывает степень приплюснотости или заостренности пика кривой нормального распределения, чем больше σ, тем шире пик, чем меньше σ, тем уже пик. Если мы отложим интервалы равные среднеквадратическому отклонению на графике с нормальным распределением, то тогда в интервал математическое ожидание ±среднеквадратическое отклонение попадет 68% всех измерений, в интервал математическое ожидание ±2 среднеквадратических отклонения попадет 95% всех измерений, а в интервал математическое ожидание ±3 среднеквадратических отклонения попадет 99,7% всех измерений.

Что касается практического применения закона нормального распределения, то он широко используется во многих отраслях науки и техники. Например, с показателями надежности тракторов, сельскохозяйственных машин и их элементов закон нормального распределения используется в случае определения характеристик рассеивания (времени и стоимости восстановления работоспособности машины и ее элементов). Примеры нормального распределения встречаются нам повсюду, это и клин перелетных птиц, и износ травяного покрытия спортивных полей, который имеет определенную структуру: отметины находятся в обе стороны от центра поля и по бокам возле лавки тренера, это и плиты дорог, тротуаров и пеших зон, которые чаще всего изнашиваются по краям, а в местах стыков такой износ напоминает кривую нормального распределения, аналогично и с залами ожидания, вокруг наиболее излюбленных или удобных мест общественного пользования, а в особенности возле мест, откуда видна информация о прибытии и отправке транспорта, на полу всегда видны следы износа, также примером могут послужить выборы, процент проголосовавших «за» по всей стране распределен по нормальному закону распределения, погрешность измерительных приборов, рост людей и многое другое.

В рамках написания данной статьи был проведен опрос 20 студентов, обучающихся в одной учебной группе. Студентам необходимо было указать свой рост и средний балл за прошлый учебный год. Далее представлены результаты опроса проранжированные в порядке возрастания. Рост студентов: 159; 163; 165; 167; 168; 170; 172; 172; 172; 173; 173; 173; 173; 176, 178, 182, 184, 186, 193, 200. Средний балл студентов: 7,4; 7,8; 7,9; 8,2; 8,2; 8,3; 8,4; 8,4; 8,5; 8,5; 8,6; 8,7; 8,7; 8,8; 8,9; 9,1; 9,1; 9,1; 9,3; 9,5.

По результатам опроса были проведены математические расчеты с целью получения необходимых данных для построения графиков, наглядно демонстрирующих результаты исследования. Далее приведены результаты математических расчетов, относящихся к росту и среднему баллу студентов соответственно. Рост: размах варьирования равен (w) 41; количество интервалов варьирования (l) 6; длина одного интервала (h) 6,83; частоты (n_i): 3; 6; 6; 3; 1; 1; относительные частоты (w_i): 0,15; 0,3; 0,3; 0,15; 0,05; 0,05; функция распределения (F(x_i)): 0; 0,15; 0,45; 0,75; 0,9; 0,95; 1; среднее выборочное ($\bar{x}_в$) 174,72; выборочная дисперсия ($D_в$) 77,5; выборочное среднее квадратическое ($\sigma_в$) 8,8; исправленная дисперсия ($D_и$) 81,6; исправленное среднее квадратическое ($\sigma_и$) 9 и другие данные. Средний балл: размах варьирования равен (w) 2,1; количество интервалов варьирования (l) 6; длина одного интервала (h) 0,35; частоты (n_i): 1; 2; 3; 8; 4; 2; относительные частоты (w_i): 0,05; 0,1; 0,15; 0,4; 0,2; 0,1; функция распределения

($F(x_i)$): 0; 0,05; 0,15; 0,3; 0,7; 0,9; 1; среднее выборочное ($\bar{x}_в$) 8,59; выборочная дисперсия ($D_в$) 0,19; выборочное среднее квадратическое ($\sigma_в$) 0,44; исправленная дисперсия ($D_и$) 0,2; исправленное среднее квадратическое ($\sigma_и$) 0,45 и другие данные.

На основе полученных данных были построены графики эмпирических функций распределения, полигоны частот и гистограммы относительных частот для роста и среднего балла учащихся, которые наглядно продемонстрировали, что полученные распределения стремятся к нормальному. Благодаря полигону частот мы можем посчитать моду распределения - это самое наиболее часто встречающееся значение (например, положение пика). Нормальное распределение является унимодальным, т.е. у него мода равняется математическому ожиданию. В соответствии с таблицей распределения Пирсона, полученные нами распределения с вероятностью 62% (рост учащихся) и 69% (средний балл учащихся) являются нормальными.

Таким образом, по результатам проведенного исследования, которое касалось роста и среднего балла учащихся одной группы, было доказано, что полученные данные стремятся к нормальному распределению. А также, исходя из всего вышесказанного можно сделать вывод, что разнообразие повсеместно встречающихся примеров нормального распределения вокруг нас подтверждает его актуальность. Об этом также свидетельствует использование законов нормального распределения в различных отраслях науки и техники.

Библиографический список

1. Высшая математика – просто и доступно! [Электронный ресурс] / Нормальный закон распределения вероятностей. – Минск, 2005. – Режим доступа: http://mathprofi.ru/normalnoe_raspredelenie_veroyatnostei.html - Дата доступа: 17.03.2020
2. Чарльз Уилан, Голая статистика. Самая интересная книга о самой скучной науке / Чарльз Уилан: пер. с англ. И. Веригина [науч. ред. А. Минько]. – 2-е изд. – Москва: Издательство «Миф», 2016. – 352с.
3. Кобзарь, А.И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. – 2-е изд., испр. – Москва: ФИЗМАТЛИТ, 2012. - 816 с.
4. Королюк, В. С. Справочник по теории вероятностей и математической статистике / Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф. — Москва: Наука, 1985. — 640 с.
5. Крамер Гаральд, Случайные величины и распределения вероятностей / Крамер Гаральд: пер. с англ. А.М. Ягмола [науч. ред. А.Н. Колмогорова]. – Москва: Государственное издательство иностранной литературы, 1947. – 144с.
6. Моделирование систем [Электронный ресурс] / Лекция 25. Моделирование нормально распределенных случайных величин. – Минск, 2005. – Режим доступа: http://stratum.ac.ru/education/textbooks/modelir/lection25.html - Дата доступа: 17.03.2020.