

УДК 519.23

**Рыбак Г.В., студент аспирантуры
2 курс, высшая инженерная школа «Технология
лесозаготовительных и деревоперерабатывающих производств»
Северный (Арктический) федеральный университет
Россия, г. Архангельск**

АНАЛИЗ ПОЛУЧЕННЫХ ИЗМЕРЕНИЙ ДЕРЖАЩЕЙ СИЛЫ МОДЕЛИ ЯКОРЯ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ

Аннотация: *Статье рассматривается нормальность распределения значений держащей силы якоря, визуальная оценка нормальности при помощи графика, оценка нормальности по критериям: Колмогорова-Смирного и Шапиро – Уилка.*

Annotation: *The article considers the normality of the distribution of values of the holding force of the anchor, the visual assessment of normality using the graph, the assessment of normality according to the criteria: Kolmogorov-Smirny and Shapiro - Wilk.*

Ключевые слова: *(нормальность распределения, критерий Колмогорова-Смирного и Шапиро – Уилка, описательная статистика)*

Key words: *(normal distribution, Kolmogorov-Smirny and Shapiro-Wilk criteria, descriptive statistics)*

В результате исследований, как правило, получают большое количество однотипных значений наблюдаемой величины. Все эти значения называют статистической совокупностью. Генеральная статистическая совокупность содержит в себе все возможные значения случайных чисел. В большинстве случаев при исследованиях получают так называемую выборочную совокупность, в которой содержится лишь некоторая часть случайных чисел

генеральной совокупности [1]. Выборочную совокупность называют выборкой, а число экспериментов в выборке – объемом выборки. При планировании и проведении экспериментов в большинстве случаев создаются подобные условия. При этом появляется закономерность в частоте появления результатов опытов.

Наиболее мощным программным обеспечением (ПО) для обработки статистических данных является Statistica. Большое количество литературных источников [2] наглядно показывают, как можно пользоваться данным программным комплексом.

В работе рассмотрен подбор закона распределения значений державей силы с помощью ПО Statistica. Чтобы установить, по какому закону распределены значения случайных величин, необходимо воспользоваться программным модулем Distribution fitting, позволяющим определить соответствие анализируемых данных ряду математических распределений.

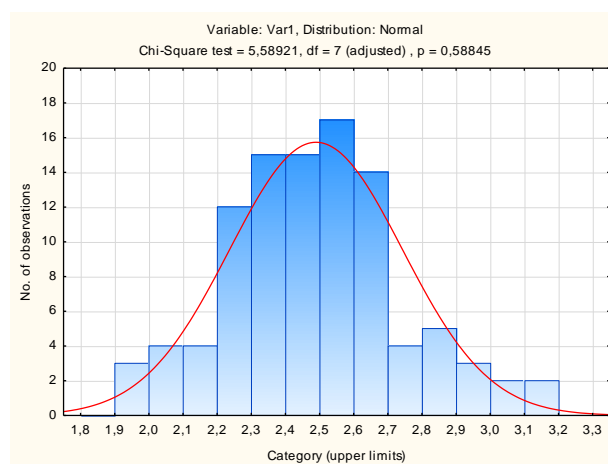


Рисунок 1. График наблюдаемого и ожидаемого распределения

Визуально по построенному графику рисунок 1, можно сделать вывод о том, что случайные значения наблюдаемой величины подвержены нормальному закону распределения. Выдвинем гипотезу о том, что имеющиеся эмпирические данные подчиняются нормальному закону распределения.

Для проверки гипотезы о том, что эмпирическое распределение случайной величины соответствует теоретическому распределению вероятностей, используют критерии согласия.

Существует большое количество критериев согласия, наиболее распространенными из которых являются: критерий Пирсона (χ^2), критерий Колмогорова (Колмогорова–Смирнова).

1. Критерий Пирсона (χ^2). Наиболее часто употребляемый критерий при достаточном объеме выборки ($n > 100$) для проверки гипотезы о принадлежности наблюдаемой выборки случайной величины теоретическому закону распределения. Его можно представить, как сумму отношений квадратов расхождений между теоретическими и эмпирическими частотами к теоретическим частотам.

2. Критерий Колмогорова (Колмогорова–Смирнова). Предназначен для проверки гипотезы о принадлежности выборки случайных чисел определенному закону распределения.

Основан на определении максимального расхождения между накопленными эмпирическими и теоретическими частотами.

Одним из основных условий использования критерия Колмогорова, является достаточно большое число наблюдений.

В некоторых случаях мощность теста χ^2 (критерий Пирсона) при проверке нормальности распределения случайной величины относительно невысока. Поэтому можно воспользоваться другими критериями согласия, например, Колмогорова – Смирнова и Шапиро – Уилка.

Для определения распределения случайной величины в ПО Statistica построена гистограмма, на которой указаны два критерия согласия, рисунок 2.

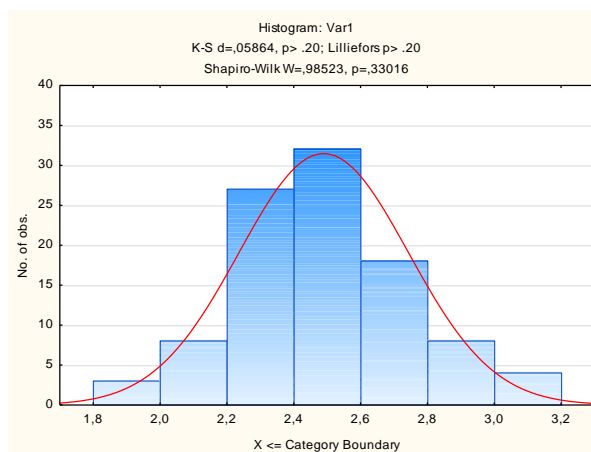


Рисунок 2. Проверка нормальности распределения с помощью критериев согласия Колмогорова – Смирного и Шапиро – Уилка

Проанализируем данные, полученные при построении гистограммы. По критерию согласия Колмогорова–Смирнова K–S на рисунок 2, можно сделать вывод, что уровень значимости (p) больше 0,2, следовательно, гипотеза о нормальности имеющегося распределения случайной величины не отклоняется. По Шапиро–Уилка (Shapiro-Wilk на рисунок 2, уровень значимости больше 0,05, следовательно, гипотеза о нормальности имеющегося распределения случайной величины не отклоняется.

Кроме критериев согласия на нормальность распределения могут указать и описательные статистики. Полученные описательные статистики для выборки представлены на рисунке 3.

Variable	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
Var1	2,489820	2,491500	1,951000	3,151000	0,253530	0,243726	0,241380	0,114795	0,478331

Рисунок 3. Описательные статистики для выборки

Проведем анализ полученных статистик.

1. Среднее значение (mean) и медиана (median) близки по своим значениям. Это является одним из косвенных признаков нормальности распределения.

2. Оценим коэффициент асимметрии (skewness). Его величина может быть положительной (для правосторонней асимметрии) и отрицательной (для левосторонней асимметрии). Если асимметрия по модулю выше 0,5, то она считается значительной. Если значение коэффициента асимметрии более чем в 2 раза превышает стандартную ошибку асимметрии (std. err. skewness), то это указывает на наличие асимметрии распределения. По рисунку 3, можно сделать вывод, что асимметрия незначительна и гипотеза о нормальности распределения не отклоняется.

3. Оценим значение эксцесса (kurtosis). Если его значение больше нуля, то график распределения случайной величины будет островершинным, в обратном случае – плосковершинным. Для нормального распределения значение эксцесса должно быть близко к 0.

Если значения эксцесса и его стандартной ошибки (std. err. kurtosis) различаются по модулю менее, чем в 2 раза, то это является одним из косвенных признаков нормальности распределения. По рисунку 3, можно сделать вывод, что значения эксцесса и его стандартной ошибки различаются по модулю более, чем в 4 раза, следовательно, гипотеза о нормальности распределения отклоняется.

Полученный нормально-вероятностный график представлен на рисунке

4.

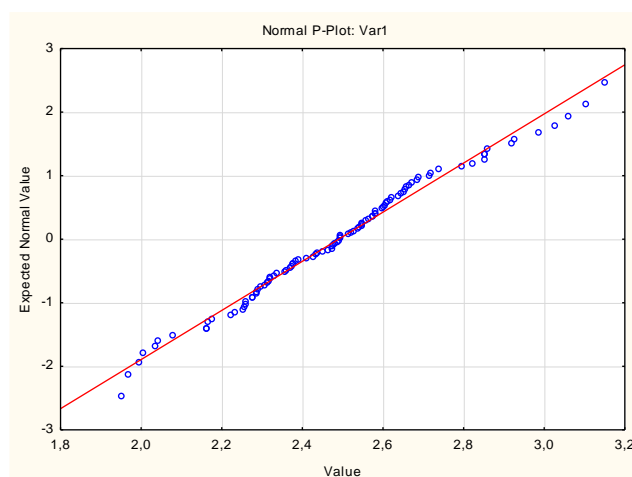


Рисунок 4. Нормально-вероятностный график

В том случае, когда экспериментальные данные достаточно близко расположены вдоль теоретической нормальной прямой рисунок 4, гипотеза о нормальности не отклоняется.

Нормальность распределения случайной величины также может быть оценена с помощью «ящичной» диаграммы (BoxPlot). В ПО «Statistica» данный тип диаграммы получают в модуле Graphs. Выбираем 2D и диаграмму BoxPlot рисунок 5.

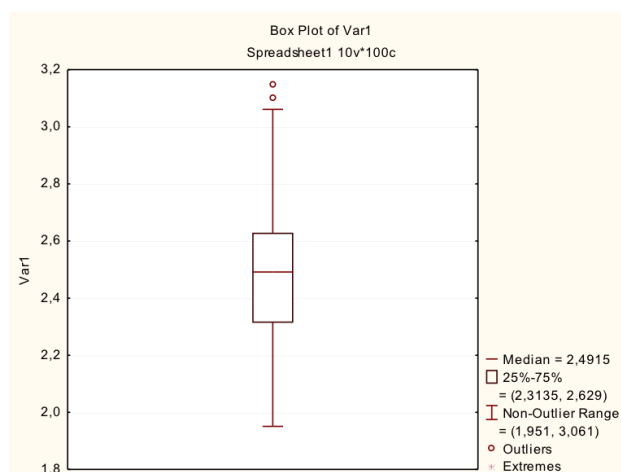


Рисунок 5. Диаграмма Box Plot

Диаграмму Box plot можно оценить визуально: усы достаточно симметричны, медиана находится практически по центру, в целом распределение можно считать нормальным. В верхней части диаграммы видны две точки, которые говорят о наличии грубых ошибок. При наведении курсора на данные точки можно увидеть значение грубой ошибки. Исключим их из выборки и проверим ее на «нормальность».

Определим новые значения критериев согласия Колмогорова–Смирнова и Шапиро–Уилка

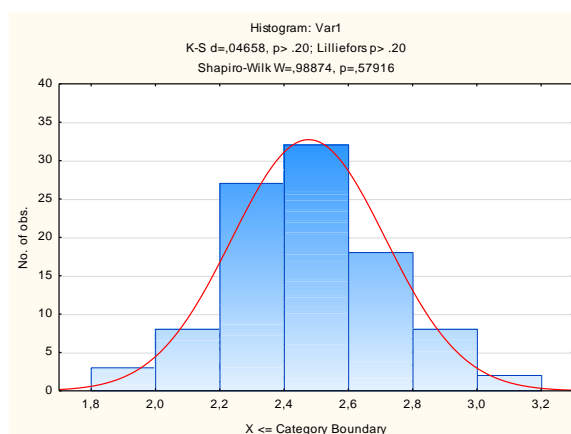


Рисунок 6. Проверка нормальности распределения

По критерию согласия Колмогорова–Смирнова уровень значимости больше 0,2, по Шапиро–Уилка – больше 0,05, следовательно, гипотеза о нормальности имеющегося распределения случайной величины не отклоняется. По сравнению с рисунок 2, вероятность того, что распределение случайной величины близко к нормальному, стала выше.

Построим и проанализируем нормально-вероятностный график, рисунок 7.

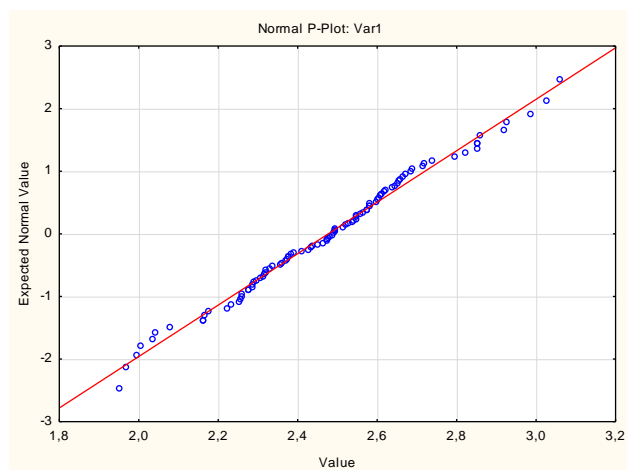


Рисунок 7. Нормально-вероятностный график

Экспериментальные данные достаточно близко расположены вдоль теоретической нормальной прямой, гипотеза о нормальности не отклоняется. По сравнению с рисунком 4, экспериментальные данные на графике рисунок 7, находятся ближе к теоретической нормальной прямой.

Построим и проанализируем диаграмму BoxPlot рисунок 8.

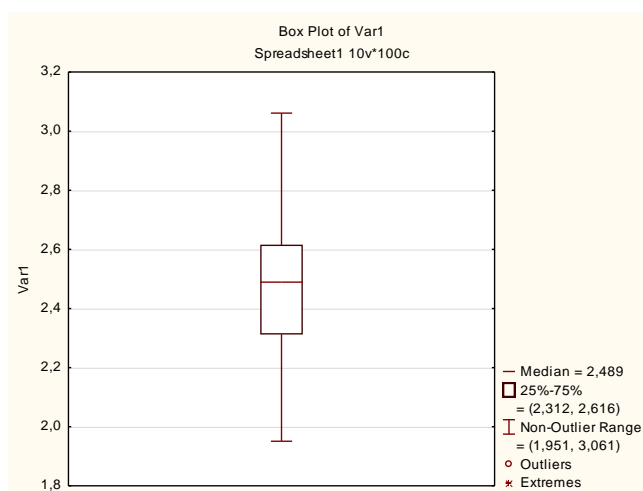


Рисунок 8. Диаграмма Box Plot

Ящик и усы достаточно симметричны, медиана находится практически по центру. Распределение случайной величины в выборке можно считать нормальным.

По сравнению с рисунком 5, диаграмма рисунок 8, визуально выглядит более приближённой к описанию нормального распределения. Исключенные грубые ошибки из выборки отсутствуют.

В результате исследований теория о нормальности распределения значений держащей силы подтверждена.

Использованные источники:

1. Северный Арктический Федеральный Университет им. М.В. Ломоносова. – Электронные текстовые данные. – Архангельск: САФУ, 2018. – 94 с.
2. Перфильев П.Н, Мурашова О.В., Задраускайте Н.О., Моделирование и оптимизация технологических процессов предприятий лесопромышленного комплекса: Северный Арктический Федеральный Университет им. М.В. Ломоносова. – Электронные текстовые данные. – Архангельск: САФУ, 2018. – 94 с.