СРАВНЕНИЕ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ И ЭМПИРИЧЕСКОЙ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ПРИ СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

Картечина Наталья Викторовна,
заведующая кафедрой математики,
физики и информационных технологий,
доцент, ФГБОУ ВО Мичуринский ГАУ,
г. Мичуринск, РФ
kartechnatali@mail.ru

Бобрович Лариса Викторовна,
профессор кафедры агрохимии,
почвоведения и агроэкологии
ФГБОУ ВО Мичуринский ГАУ,
г. Мичуринск, РФ
bobrovich63@mail.ru

Пчелинцева Наталия Владимировна,
старший преподаватель кафедры
математики, физики
и информационных технологий,
ФГБОУ ВО Мичуринский ГАУ,
г. Мичуринск, РФ
natas79@mail.ru

Картечина Ольга Сергеевна
студентка 1 курса направления подготовки
«Техносферная безопасность» ИУИТ
Российский университет транспорта (МИИТ)
г. Москва, РФ
kartechnatali@mail.ru

Аннотация. В статье рассматривается целесообразность применения нормального распределения и эмпирической функции распределения для обработки результатов эксперимента в плодоводстве.

Ключевые слова. Нормальная функция распределения, эмпирическая функция распределения, выборка, плодоводство, яблоня.

В течение нескольких десятилетий сложилась традиция применения нормального распределения при статистической обработке экспериментальных данных в плодоводстве [1; 2]. Основным критерием применимости кривой Гаусса для оценки вероятности событий стал критерий Пирсона («хи»-квадрат). На примере выборки в 100 единиц попытаемся понять обоснованность применения нормального распределения и его целесообразность. Выборку представим в виде ранжированного ряда длины приростов продолжения скелетных ветвей деревьев яблони сорта Антоновка обыкновенная, измеренных с точностью до 1 см (таблица 1).

Таблица 1

Ранжированный ряд выборки длины приростов деревьев яблони сорта Антоновка обыкновенная (см)

x_i	6	9	10	11	12	13	14	15	16	17	18	19
n_i	1	2	2	1	5	4	5	5	3	9	1	8
x_i	20	21	22	23	24	25	26	28	29	30	31	32
n_i	5	5	4	3	1	8	5	3	1	2	1	1
x_i	33	34	35	36	37	39	40	41	45	46	50	
n_i	3	3	1	1	1	1	1	1	1	1	1	

$$\sum n_i = n = 100\,; X_{Bmin} = 6;\ X_{Bmax} = 50$$

Оценим репрезентативность такой выборки. Обозначим отношение размаха выборки к размаху генеральной совокупности:

$$\frac{X_{B\,max} - X_{Bmin}}{X_{r\,max} - X_{rmin}}$$

Вероятность каждого значения q зависит от объема выборки:

$$P(q) = 1 - nq^{n-1} + (n-1)q^n$$

В нашем случае P(q) = 0,95; n=100=0,96. Объем выборки n=100 позволяет с достаточной надежностью оценить принадлежность распределения генеральной совокупности какому-либо теоретическому.

В результате расчетов получены следующие статистические характеристики: выборочная средняя $\bar{X}_B = 22{,}05$; оценка генерального среднего квадратического отклонения S=8,8; оценка среднеквадратического отклонения средней $S_{\bar{x}} = 0{,}88$. Расчет критерия «хи»-квадрат показал, что его наблюдаемая величина равна 10,53. В то же время табличная с уровнем надежности 0,99 равна 15,09. Расчеты велись исходя из предположения о нормальности распределения генеральной совокупности. Критерий Пирсона с надежностью 0,99 не отверг нашу гипотезу.

Зададим таблично эмпирическую функцию распределения вероятностей, используя приближенное равенство вероятности событий и относительной частоты (таблица 2). Шаг аргумента возьмем 2 см.

Таблица 2

Таблица значений эмпирической функции распределения вероятностей

X_i	8	10	12	14	16	18	20	22	24	26	28	30
W(x_i>x)	0,01	0,03	0,06	0,15	0,25	0,37	0,46	0,56	0,63	0,72	0,77	0,81
X_i	32	34	36	38	40	42	44	46	48	50		
W(x_i>x)	0,84	0,88	0,92	0,94	0,96	0,98	0,98	0,99	0,99	1		

Эмпирическая вероятность неравенства 15<x<30 равна 0,42, а вычисленная по закону нормального распределения равна 0,61. Эмпирический интервал, в который попадает случайная величина с вероятностью 0,99 6<x<50, а нормально распределенная 0,65<x<44,75, то есть теоретически допускаются отрицательные приросты. В многочисленных изданиях, посвященных методике опытного дела пишут, что, начиная с объема выборки n=20, можно считать распределение средних величин нормальным [1; 2].

Сравним интервальные оценки средней с надежностью 0,99, полученные эмпирически и теоретически для некоторых значений объема выборки (таблица 3). Эмпирические границы средней вычисляются усреднением соответствующего числа самых низких и самых высоких значений признака.

Таблица 3

Интервальные оценки средних величин с надежностью 0,99, полученные эмпирически и теоретически для выборок различного объема

n	$\overline{X}_{эмп}$	$\overline{X}_{теор}$	$(\overline{X} + 3\sigma)_{теор}$
20	11,9-36,1	16,4-27,7	16,1-28,0
60	16,1-27,3	19,0-25,1	18,6-25,5
75	18,0-25,4	19,4-24,7	19,0-25,1
80	18,4-24,7	19,5-24,6	19,1-25,0
85	19,3-24,1	19,5-24,6	19,2-24,9
90	20,1-23,4	19,6-24,5	19,2-24,8
95	20,9-22,8	19,5-24,4	19,3-24,8

Анализ приведенной таблицы показывает, что совпадение интервалов с предельной точностью оценки средних 2 см при надежности 0,99 начинается с n=85. Интервальная оценка объема выборки, исходящая из нормальности распределения

$$k = \frac{t^2{}_\gamma\ \sigma^2(1 \mp q)^2}{\Delta^2}$$

дает 86<r<191, что хорошо согласуется с эмпирическими данными. Начиная с n=90 эмпирическое распределение дает оценку средней лучше, чем теоретическое. Это обусловлено тем, что кривая Гаусса далеко не идеально соответствует истинному распределению генеральной совокупности.

Асимметрия теоретического распределения A_s=0,860, эксцесс E_x=0,48. Эти величины существенно отличаются от нуля, а истинное распределение от нормального. Коэффициент Дарбина-Уотсона d=0,8, что указывает на необходимость корректировки или замены вида выбранной функции теоретического распределения. Воздействие асимметрии и эксцесса теоретического распределения на интервальные оценки средней величины практически устраняются, если учитывается дисперсия среднего квадратического отклонения. Причем, при выбранной надежности 0,99 поправочный коэффициент для выборок объемом n<85 берется со знаком «плюс», при n>85 – со знаком «минус». При n = 85 поправка при оценке левой границы интервала берется со знаком «плюс», правой – со знаком «минус».

Большие выборки ($n \geq 100$) позволяют проводить оценки интересующих исследователя величин, не прибегая к определению вида теоретического распределения. Большие выборки позволяют более точно выбирать теоретическое распределение генеральной совокупности.

Список литературы

1. Картечина Н.В., Бутенко А.И., Брижанский Л.В., Пчелинцева Н.В., Бобрович Л.В. Статистическая оценка динамики роста и плодоношения яблони// Вестник Мичуринского государственного аграрного университета. – Мичуринск,2018, № 2. – С. 31-36.

2. Потапов В.А., Завражнов А.И., Бобрович Л.В., Петрушин В.Н. Биометрия плодовых культур. - Мичуринск, Изд-во Мичуринский ГАУ, 2004. - с. 332

COMPARISON OF NORMAL DISTRIBUTION AND EMPIRICAL DISTRIBUTION FUNCTION IN STATISTICAL PROCESSING OF MEASUREMENT RESULTS

Kartechina Natalia Viktorovna,

Associate Professor of the Department mathematics,

physics and information technology

Michurinsk State Agrarian University

Michurinsk, Russia.

kartechnatali@mail.ru

Bobrovich Larisa Viktorovna,

Associate Professor of theDepartment

agrochemistry, soil science and agroecology

Michurinsk State Agrarian University,

Michurinsk, Russia,

bobrovich63@mail.ru

Pchelintseva Natalia Vladimirovna

senior lecturer, Department of mathematics,

physics and information technologies,

Michurinsk State Agrarian University,

Michurinsk, Russia

natas79@mail.ru

Kartechina Ol'ga Sergeevna

2-nd year student of the direction

of training «Technosphere safety»

Russian University of Transport (MIIT)

Moscow, Russia

kartechnatali@mail.ru

Annotation. The article discusses the feasibility of applying the normal distribution and the empirical distribution function for processing the results of an experiment in fruit growing.

Keywords. Normal distribution function, empirical distribution function, sampling, fruit growing, apple tree.