**ОБРАБОТКА ЛЕСОВОДСТВЕННОЙ ИНФОРМАЦИИ**

**С ПОМОЩЬЮ СТАТИСТИЧЕСКОГО ПАКЕТА «STATGRAPHICS»**

Тихонова В.С., магистрант СибГУ ……………………….

**Введение.** В настоящее время на смену ручной обработке данных приходит автоматическая работа с информацией, которая в значительной мере позволяет ускорить процесс вычислений и больше внимания уделять непосредственно анализу данных.

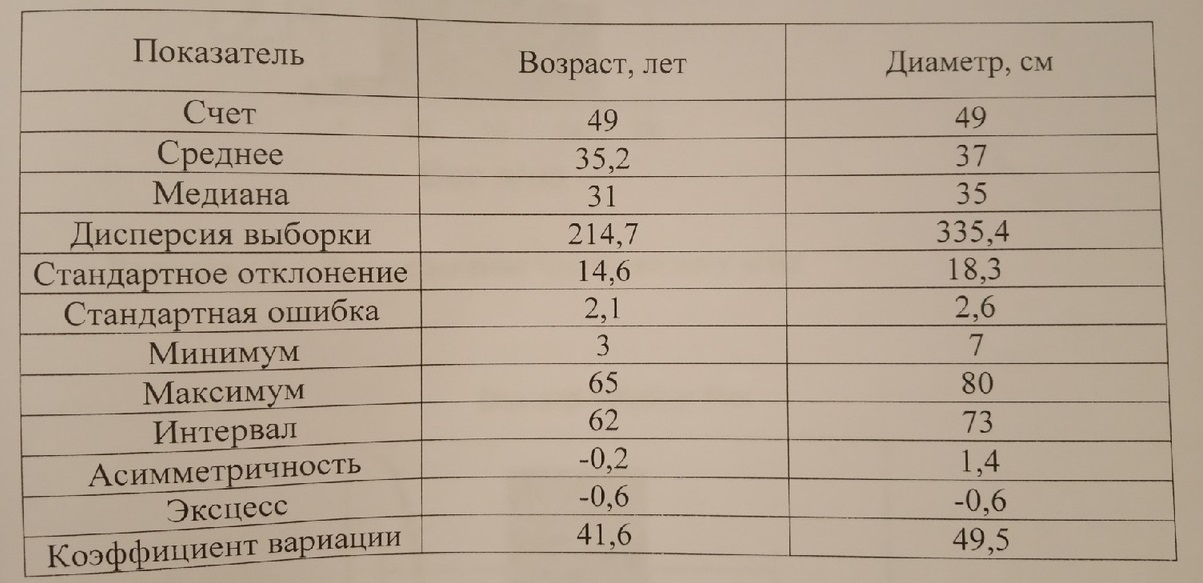
Большое распространение получили пакеты прикладных программ, предназначенных для статистической и графической обработки данных [1,2,3]. К ним относятся ППП «STATGRAPHICS», «SPSS», «STATISTICA», «CurveExpert», пакет анализа в «Excel». Все эти программы обладают различными инструментами и интерфейсами. Главным критерием для отбора программ являются простота интерфейса и возможности по первичному преобразованию переменных.

**Результаты и их обсуждение.** В данной работе изучены методы обработки лесоводственной информации с помощью программы «STATGRAPHICS». В качестве примера для работы с исходными данными были взяты данные популяции брусники Саралинского лесничества по двум признакам: вес ягоды и диаметр ягоды.

Для обработки данных эмпирического исследования обязательным является применение статистических методов. С помощью программы были получены статистические показатели, такие как минимум, максимум, асимметричность, эксцесс, медиана и др., что помогает представить количественные данные в сжатой форме, чтобы упростить последующий анализ (таблица 1).

Так же программа позволила получить рисунки в виде графиков и диаграмм для большей наглядности распределения частот по весу и диаметру. Облегчена возможность изучения медианы, нижнего и верхнего квартилей, минимального и максимального значения выборки, выбросов. Полученные рисунки являются быстрым способом изучения признаков в графическом виде.

Таблица 1 – Статистические показатели



Помимо стандартных графиков распределения частот программа позволяет получить так называемую диаграмму «ящик с усами» (рисунок 1).

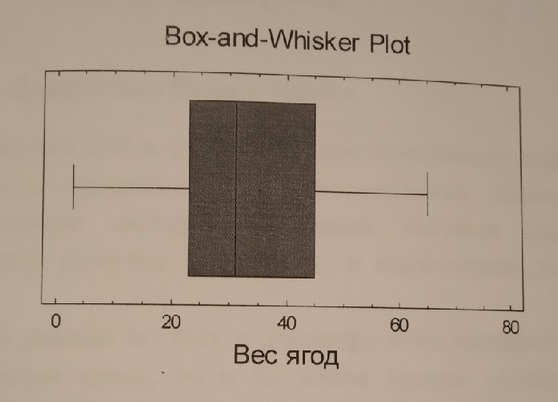


Рисунок 1 – Диаграмма «ящик с усами»

Расстояние между различными частями ящика позволяют определить степень разброса (дисперсии) и асимметрии данных и выявить выбросы. Асимметрия данных выявляется не только по медиане, смещенной влево, но и по длине «усов», выходящих из ящика.

Далее с помощью полученных таблиц и рисунков было наглядно продемонстрированно как программа позволяет провести сравнение выборочных данных с нормальным распределением и выполнить подгонку, что является одним из наиболее распространенных методов статистического анализа.

Для обнаружения связи между переменными, исследования ее силы и направленности используется корреляционный анализ, в рамках которого оцениваются различные показатели связи и их значимость. Для данных, не показывающих «нормального» распределения (качественных признаков и некоторых количественных) используется коэффициент корреляции рангов. Программа позволяет получить простой численный ответ на вопрос о степени корреляционной зависимости переменных, а так же продемонстрировать графические изображения корреляционного анализа (рисунок 2).

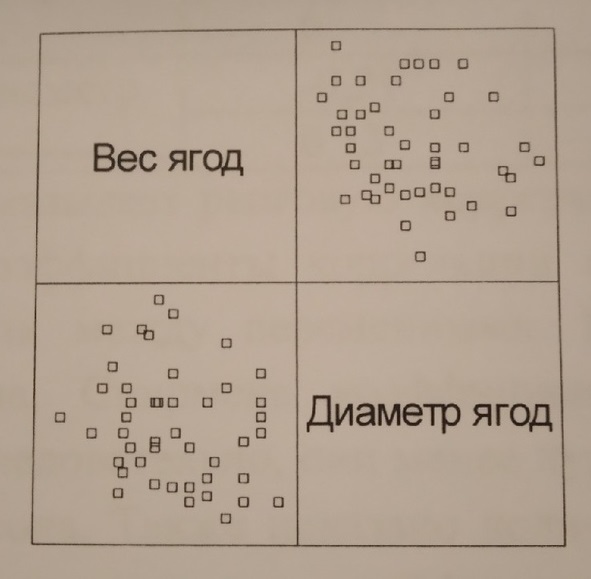
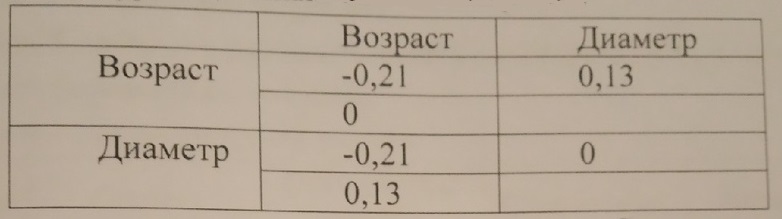


Рисунок 2 – Корреляционный анализ веса и диаметра ягод

Ранговая корреляция Спирмена между двумя признаками (таблица 2). Эти коэффициенты корреляции в диапазоне от -1 до +1 и измеряют силу связи между переменными. В отличие от более обшей корреляции Пирсона, коэффициент Спирмена вычисляется из рядов значений данных. Следовательно, они менее чувствительны к выбросам, чем коэффициенты Пирсона.

Таблица 2 – Корреляционный анализ Спирмена



Так же с помощью программы «STATGRAPHICS» был проведен дисперсионный и регрессионный анализ данных с их результатами в виде таблиц и графические изображения линейных и полиномиальных регрессий различных связей. Данные показывают результаты подгонки линейной модели для описания отношения между весом ягод и диаметром ягод (рисунок 3).

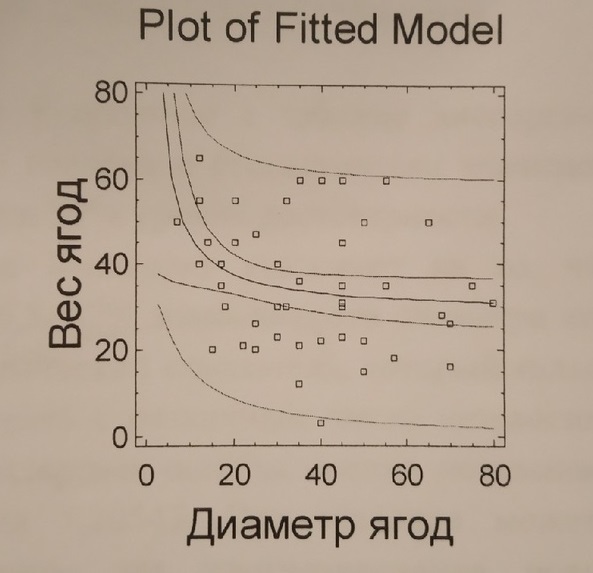


Рисунок 3 - Линейная регрессия

Уравнение подогнанной модели:

Диаметр = 9,72985 + 0,140554\*Вес ягод

Поскольку P-значение в таблице дисперсионного анализа составляет менее 0,01, это говорит о статистически значимой связи между Диаметром ягод и Весом при 99% уровне достоверности. Статистика R-квадрат указывает на то, что модель объясняет 66,8 % изменчивости диаметра ягод. Скорректированный R- квадрат статистический показатель, который является более подходящим для сравнения моделей с различным числом независимых переменных, он равен 66,7 %. Стандартная ошибка оценки показывает стандартное отклонение остатков, равна 5,2 см. Это значение может быть использовано для создания границы для прогнозирования новых наблюдений. Средняя абсолютная ошибка (4,1) – это среднее значение невязки. Поскольку р-значение меньше, чем 0,05, имеются признаки возможной автокорреляции.

Было произведено сравнение альтернативных регрессионных моделей и получено уравнение зависимости между диаметром и весом ягод:

Диаметр ягод = 1/(0,0522586-0,000120535\*Вес ягод).

Из ряда моделей наиболее адекватной оказалась обратная функция (рисунок 4).

Поскольку Р-значение в таблице дисперсионного анализа составляет менее 0,01, есть статистически значимая связь между Диаметром и Весом ягод с 99% уровнем достоверности. Коэффициент R-квадрат статистики показывает, что модель объясняет, как они установлены 75,4 % вариабельности в Диаметре, после преобразования во взаимный масштаб, чтобы линеаризовать модель.

Коэффициент корреляции равен - 0,87, что указывает на умеренно прочные отношения между переменными. Стандартная ошибка оценки показывает стандартное отклонение невязок - 0,004. Это значение может быть использовано для построения границ предсказания для новых

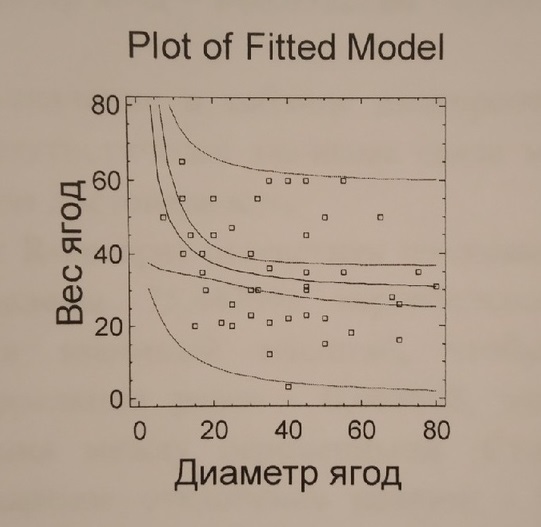


Рисунок 4 – Зависимость диаметра от веса ягод (модель – обратная по Y)

наблюдений. Поскольку p-значение меньше чем 0,05, имеются признаки возможной автокорреляции.

**Выводы.** По итогу изучения можно заключить, что программа «STATGRAPHICS» позволяет относительно просто произвести все необходимые для статистической обработки данных манипуляции, а так же предоставить их окончательные и промежуточные результаты в удобном формате. К недостаткам программы относится наличие только англоязычной версии.

**Список литературы**

1. Дюк В. Обработка данных на ПК в примерах. – СПб: Питер, 1997. – 240 с.
2. Боровиков В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – СПб: Питер, 2001. – 656 с.
3. Алексеев Е.Р., Чеснокова О.В. MATLAB 7. – М: НТ Пресс, 2006. – 464 с.