

ГУАП

КАФЕДРА № 41

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

доцент

должность, уч. степень, звание

подпись, дата

Г. В. Титова

инициалы, фамилия

ОТЧЕТ О ПРАКТИЧЕСКОЙ РАБОТЕ №2

СТАТИСТИЧЕСКАЯ ПРОВЕРКА КРИТЕРИЕВ

по курсу: СТАТИСТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ гр. №

4316

подпись, дата

А. А. Жукова

инициалы, фамилия

Санкт-Петербург 2025

Цель работы

Целью практической работы является освоение методов статистической проверки гипотез на основе реальных экономических данных, а также применение параметрических и непараметрических критериев для анализа различий между выборками.

1 Описание данных

Для выполнения лабораторной работы были взяты данные из датасетов «Объемы производства овощной продукции.xls» (далее датасет «Овощи») и «Объемы производства фруктов, напитков и пряностей.xls» (далее датасет «Фрукты»), полученного с официального сайта Турецкого статистического института (TURKSTAT, <https://data.tuik.gov.tr>). Это данные из бюллетеня «Вторая оценка производства растений, 2025 г.». Ссылка на информационный бюллетень <https://data.tuik.gov.tr/Bulten/Index?p=Bitkisel-Uretim-2.Tahmini-2025-53938>.

Использован исключительно 2025 год в соответствии с требованием задания о недопустимости анализа временных рядов.

Датасет «Овощи» содержит в себе переменные объёма производства в тоннах (т), единица наблюдения одна овощная культура, общее количество наблюдений 62 культуры, диапазон значений от 59 т (горький огурец) до 13 500 000 т (томаты). Для проверки критериев были образованы выборки: плодовые овощи, корнеплоды и листовые овощи. Плодовые овощи – овощи, выращиваемые ради плодов, количество культур 12, пример: томаты, огурцы, перец (все виды), баклажаны, кабачки, тыква, бамя. Корнеплоды – овощи, с подземными съедобными частями (корень, клубень, луковица). Количество культур 6, пример: лук (сухой), чеснок (сухой), морковь, свёкла, красная редька, хрен. Листовые овощи – овощи, употребляемые в пищу за счёт листьев, стеблей или соцветий, количество культур 17, пример: салат (все виды), капуста (все виды), шпинат, петрушка, укроп, рукола, сельдерей (стебель). Принцип формирования групп: ботанико-агрономическая классификация, общепринятая в статистике растениеводства (ТуркСтат, FAO). Группы однородны по типу потребляемой части растения.

Датасет «Фрукты» содержит в себе переменные объёма производства в тоннах (т), единица наблюдения одна фруктовая культура, общее количество наблюдений 57 культур, диапазон значений от 85 т (пепино) до 2 620 000 т (виноград). Для проверки критериев были образованы выборки: косточковые фрукты, цитрусовые. Косточковые фрукты – фрукты, с твёрдой косточкой внутри плода (семейство Rosaceae, подсемейство Prunoideae), количество культур 6, пример: персик, нектарин, слива, абрикос, вишня, черешня. Цитрусовые – фрукты, семейства Rutaceae, с кожистой кожурой и высоким содержанием

кислот, количество культур 4, пример: апельсин, мандарин, лимон, грейпфрут. Принцип формирования групп: систематическая ботаническая классификация (семейство, тип плода). Группы выбраны для обеспечения внутренней однородности и контрастности между группами (различия в агротехнике, рынке, объемах).

На рисунке 1 представлено начало таблицы «Объемы производства овощной продукции.xls», для работы были использованы столбцы «2025» и «Share (%)».

(Ton - Tonnes)					
Sebzeler - Vegetables	Üretim - Production				
	Pay (%)		Pay (%)		Değişim (%)
	2024	Share (%)	2025	Share (%)	
Toplam - Total	33 572 553	100,0	33 315 678	100,0	-0,8
Yumru ve kök sebzeler					
Root and tuberous vegetables	4 197 401	12,5	4 246 770	12,7	1,2
Soğan (taze) - Onion (green)	108 184	0,3	132 430	0,4	22,4
Soğan (kuru) - Onion (dry)	2 632 000	7,8	2 700 000	8,1	2,6
Sarımsak (taze) - Garlic (green)	50 870	0,2	59 000	0,2	16,0
Sarımsak (kuru) - Garlic (dry)	125 000	0,4	135 700	0,4	8,6
Pırasa - Leek	151 277	0,5	145 900	0,4	-3,6
Havuç - Carrots	892 155	2,7	858 300	2,6	-3,8
Şalgam - Turnip	3 054	0,0	2 800	0,0	-8,3
Pancar (kırmızı) - Beets (red)	24 405	0,1	16 000	0,0	-34,4
Kereviz (kök) - Celeriac	29 003	0,1	29 600	0,1	2,1
Turp (bayır) - Horse radish	12 879	0,0	12 100	0,0	-6,0
Turp (kırmızı) - Red radish	165 129	0,5	151 100	0,5	-8,5
Turp (beyaz) - White radish	3 445	0,0	3 840	0,0	11,5
Meyvesi için yetiştirilen sebzeler					
Vegetables cultivated for their fruits	26 866 588	80,0	26 143 496	78,5	-2,7
Domates - Tomatoes	14 617 000	43,5	13 500 000	40,5	-7,6
Hıyar - Cucumber	1 742 900	5,2	1 708 700	5,1	-2,0
Acur - Hairy cucumber	46 700	0,1	40 600	0,1	-13,1
Biber (salçalık, kapya) - Pepper (for processed, capia)	1 941 156	5,8	1 818 000	5,5	-6,3
Biber (dolmalık) - Pepper (bell)	421 267	1,3	426 000	1,3	1,1

Рисунок 1 – Таблица «Объемы производства овощной продукции.xls»

На рисунке 2 представлено начало таблицы «Объемы производства фруктовых продуктов, напитков и пряностей.xls», для работы были использованы столбцы «2025» и «Share (%)».

(Ton - Tonnes)					
Meyveler, içecek ve baharat bitkileri	Üretim - Production				
	Pay (%)		Pay (%)		Değişim (%)
	2024	Share (%)	2025	Share (%)	
Fruits, beverages and spices crops					
Toplam - Total	28 389 775	100,0	19 767 012	100,0	-30,4
Üzüm - Grapes	3 468 000	12,2	2 620 000	13,3	-24,5
Diğer meyveler ve sert kabuklular					
Other fruits and nuts	23 077 485	81,3	15 436 106	78,1	-33,1
Tropikal ve subtropikal meyveler					
Tropical and subtropical fruits	1 406 749	5,0	1 386 470	7,0	-1,4
Muz - Bananas	875 000	3,1	861 000	4,4	-1,6
İncir - Figs	375 000	1,3	370 000	1,9	-1,3
Kivi - Kiwi	92 249	0,3	77 470	0,4	-16,0
Avokado - Avocado	64 500	0,2	68 300	0,3	5,9
Frenk inciri ⁽¹⁾ - Prickly pear ⁽¹⁾	-	-	387	0,0	-
Diğer tropikal ve subtropikal meyveler^{(1),(2)}					
Other tropical and subtropical fruits ^{(1),(2)}	-	-	9 313	0,0	-
Turuncgiller - Citrus	5 481 576	19,3	4 801 794	24,3	-12,4
Portakal - Oranges	1 610 000	5,7	1 368 000	6,9	-15,0
Mandalina - Mandarin	1 988 000	7,0	2 130 000	10,8	7,1
Limon - Lemons	1 730 000	6,1	1 128 000	5,7	-34,8
Greyfurt (altıntop) - Grape fruits	150 000	0,5	172 000	0,9	14,7
Turunc - Sour oranges	3 576	0,0	3 540	0,0	-1,0
Diğer turuncgiller^{(1),(3)} - Other citrus fruits^{(1),(3)}					
	-	-	254	0,0	-
Diğer meyveler - Other fruits	10 638 860	37,5	5 741 284	29,0	-46,0

Рисунок 2 – Таблица «Объемы производства фруктовых продуктов, напитков и пряностей.xls»

2 Наглядное представление данных

Для анализа распределения и сравнения объёмов производства в 2025 году были применены следующие типы графиков: гистограмма с кривой плотности, столбчатая диаграмма, круговая диаграмма. Цель использования гистограммы с кривой плотности заключается в том, чтобы оценить форму распределения. Это позволяет визуальнo оценить асимметрию, мультимодальность и отклонение от нормальности — ключевые предпосылки для выбора статистических критериев. Цель использования столбчатой диаграммы заключается в том, чтобы сравнить средние значения по группам. Это наиболее наглядно демонстрирует различия в среднем объёме производства между категориями. Цель использования круговой диаграммы заключается в том, чтобы оценить доли групп в общем объёме. Круговая диаграмма эффективна при небольшом числе групп и необходимости показать вклад каждой категории в итоговый результат.

Была построена гистограмма распределения объёмов производства овощей с наложенной кривой плотности (KDE) (рис. 3).

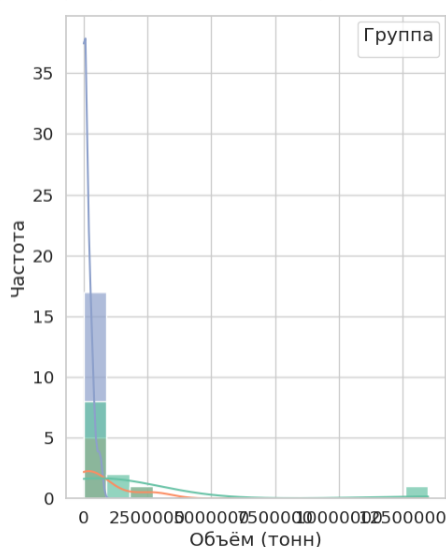


Рисунок 3 - Гистограмма распределения объёмов производства овощей с наложенной кривой плотности (KDE), 2025 год

Гистограмма демонстрирует сильную правостороннюю асимметрию распределения объёмов производства овощей. Основная масса наблюдений сосредоточена в диапазоне до 1 млн тонн, однако присутствуют значительные выбросы (в частности, томаты — 13.5 млн т). Плодовые овощи (n=12) формируют длинный правый хвост. Корнеплоды (n=6) и листовые овощи (n=17) сконцентрированы в нижней части диапазона. Кривая KDE подтверждает ненормальность распределения (множественные пики, асимметрия).

Была построена столбчатая диаграмма для среднего объёма производства по группам овощей (рис. 4).

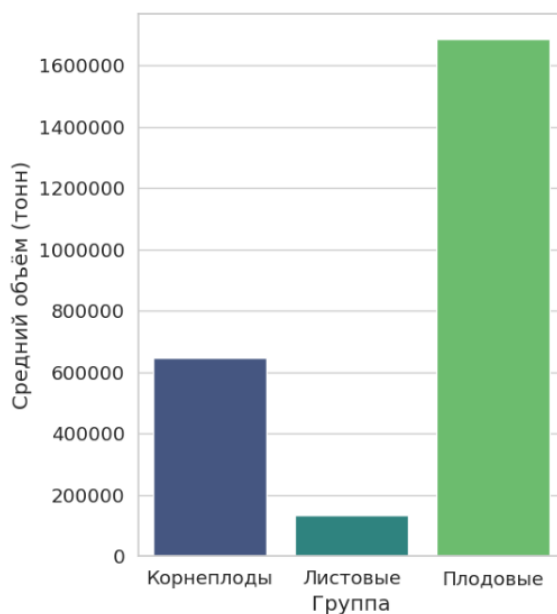


Рисунок 4 - Средний объём производства по группам овощей, 2025 год

Столбчатая диаграмма наглядно иллюстрирует значительные различия в среднем объёме производства. Плодовые овощи: 2 751 208 т, корнеплоды: 978 867 т, листовые овощи: 190 987 т. Можно сделать вывод, что плодовые овощи доминируют по объёму производства, что объясняется высокой урожайностью и площадью посевов (томаты, огурцы, перец).

Была построена круговая диаграмма для долей групп овощей в общем объёме производства (рис. 5).

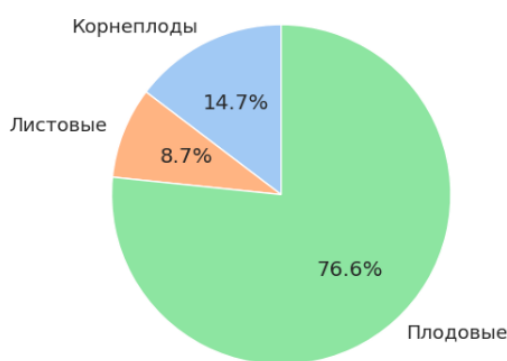


Рисунок 5 - Доля групп овощей в общем объёме производства, 2025 год

Круговая диаграмма отражает структуру производства. Плодовые овощи: 78.5%, корнеплоды: 12.7%, листовые овощи: 8.8%. Можно сделать вывод, что плодовые овощи

составляют абсолютное большинство, что подчёркивает их экономическую значимость в овощеводстве Турции.

Была построена гистограмма распределения объёмов производства фруктов с KDE (рис. 6).

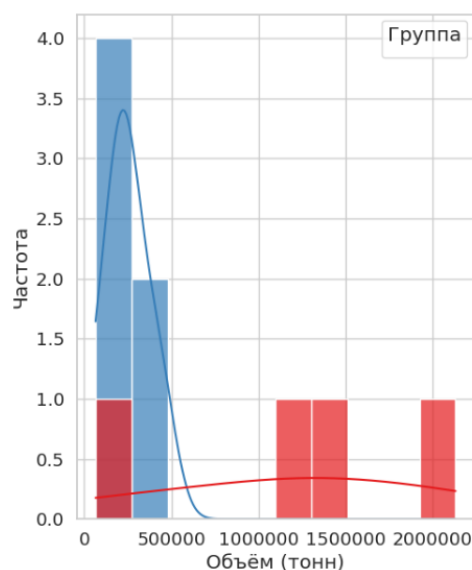


Рисунок 6 - Гистограмма распределения объёмов производства фруктов с KDE

Гистограмма показывает два кластера. Цитрусовые ($n=4$): объёмы от 172 000 до 2 130 000 т. Косточковые ($n=6$): от 65 200 до 443 000 т. Распределение также асимметрично, с правым хвостом у цитрусовых. Наличие разрыва между группами подтверждает их контрастность.

Была построена столбчатая диаграмма для среднего объёма производства по группам фруктов (рис. 7).

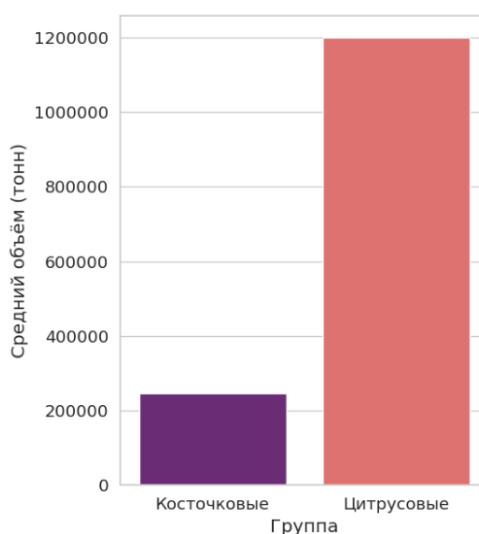


Рисунок 7 - Средний объём производства по группам фруктов, 2025 год

Столбчатая диаграмма наглядно иллюстрирует значительные различия в среднем объеме производства. Цитрусовые: 1 247 000 т, косточковые: 263 217 т. Можно сделать вывод, что цитрусовые значительно превосходят косточковые по среднему объему, что связано с экспортной ориентацией (мандарины, апельсины).

Была построена круговая диаграмма для долей групп фруктов в общем объеме производства (рис. 8).

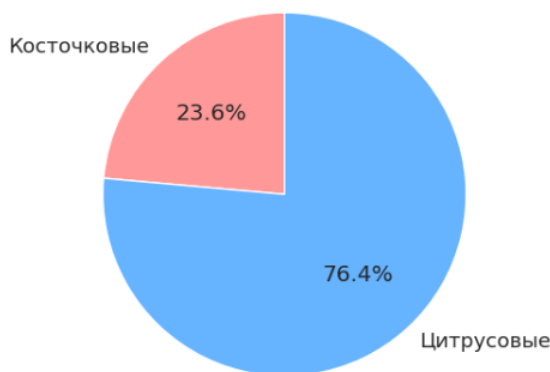


Рисунок 8 - Доля групп фруктов в общем объеме производства, 2025 год

Круговая диаграмма отражает структуру производства. Цитрусовые: 24.3%, косточковые: 6.5%. Можно сделать вывод, что цитрусовые — одна из ведущих товарных групп, несмотря на меньшее число культур.

3 Описание каждого применяемого критерия

Для анализа данных были выбраны следующие критерии: критерий Шапиро-Уилка, критерий Манна-Уитни, медианный критерий, критерий Краскела–Уоллиса, критерий Спирмена, F-критерий, t-критерий.

Критерий Шапиро–Уилка. Используется для проверки гипотезы о нормальности генерального распределения (для выборок небольшого объема $n < 50$). Критерий будет применен для всех выборок датасета «Овощи» и для всех выборок датасета «Фрукты».

Условия применения данного критерия. Объем выборки должен быть от 3 до 5000 наблюдений. Данные должны быть числовыми и не содержать пропусков. Применяется к одномерным выборкам.

Нулевая гипотеза (H_0): распределение нормально.

$$f(x) = N(\mu, \sigma^2)$$

Альтернативная гипотеза (H_1): распределение не нормально.

$$f(x) \neq N(\mu, \sigma^2)$$

Команда в Python:

```
scipy.stats.shapiro(x)
```

где:

x - массив с данными.

Нет дополнительных параметров, возвращает statistic (W) и pvalue (p-значение).

Интерпретация результатов заключается в следующем, если p-значение > 0.05 , то нет оснований отвергать H_0 : данные можно считать нормально распределенными. Если p-значение ≤ 0.05 , то H_0 отвергается: данные не являются нормально распределенными.

Критерий Манна–Уитни. Используется для проверки гипотезы о том, что две независимые выборки принадлежат однородным генеральным совокупностям. Это непараметрический тест, который не требует нормальности данных и устойчив к выбросам. Критерий будет применен в датасете «Овощи» к выборкам «Плодовые овощи» и «Корнеплоды», в датасете «Фрукты» к выборкам «Цитрусовые» и «Косточковые».

Условия применения данного критерия. Две независимые выборки. Данные должны быть порядковыми или числовыми. Не требуется нормальность распределения. Объем выборок может быть разным, но желательно $n \geq 4$.

Нулевая гипотеза (H_0). Распределения двух выборок одинаковы (одинаковое положение и форма):

$$F_X(x) = F_Y(x)$$

Альтернативная гипотеза (H_1). Распределения двух выборок различаются (по положению или форме):

$$F_X(x) \neq F_Y(x)$$

Команда в Python:

```
scipy.stats.mannwhitneyu(x, y, use_continuity=True,  
alternative='two-sided', method='auto', *,  
nan_policy='propagate',
```

Интерпретация результатов заключается в следующем, если p-значение > 0.05 , то нет оснований отвергать H_0 : распределения можно считать одинаковыми. Если p-значение ≤ 0.05 , то H_0 отвергается: распределения существенно различаются.

Медианный критерий. Используется для проверки гипотезы о том, что несколько независимых выборок принадлежат однородным генеральным совокупностям. Критерий будет применен только в датасете «Овощи» к выборкам «Плодовые овощи» и «Корнеплоды».

Условия применения данного критерия. Выборки независимы, не требуется нормальность распределения и количество наблюдений больше двух.

Нулевая гипотеза (H_0):

$$med_{x1} = med_{x2}$$

Альтернативная гипотеза (H_1). {хотя бы одно равенство не выполняется}.

Команда в Python:

```
scipy.stats.median_test(*samples, ties='below',  
correction=True, lambda_=1, nan_policy='propagate').
```

Интерпретация результатов заключается в следующем, если хотя бы одна из выборок имеет отличную от других медиану, то критерий отвергнет нулевую гипотезу.

Критерий Краскела–Уоллиса. Используется для проверки гипотезы о том, что несколько независимых выборок принадлежат однородным генеральным совокупностям. Критерий будет применен для датасета «Овощи» к выборкам «Плодовые овощи», «Корнеплоды» и «Листовые овощи».

Условия применения данного критерия. Выборки независимы, не требуется нормальность распределения и количество наблюдений может быть разное.

Нулевая гипотеза (H_0):

$$F_{(x1)}(x) = F_{(x2)}(x) = F_{(x3)}(x)$$

Альтернативная гипотеза (H_1). {хотя бы одно равенство не выполняется}.

Команда в Python:

```
scipy.stats.kruskal(*samples, nan_policy='propagate')
```

Интерпретация результатов заключается в следующем, если хотя бы одна из выборок имеет отличную от других медиану, то критерий отвергнет нулевую гипотезу.

Критерий Спирмена. Используется для проверки гипотезы о наличии зависимости между двумя признаками X и Y. Критерий будет применен ко всему датасету «Овощи» и ко всему датасету «Фрукты».

Условия применения данного критерия. Не требуется нормальность распределения и количество наблюдений может быть маленьким.

Нулевая гипотеза (H_0):

$$\rho_s = 0$$

Альтернативная гипотеза (H_1):

$$\rho_s > 0$$

Команда в Python:

```
scipy.stats.spearmanr(a, b, axis=0, nan_policy='propagate',  
alternative='two-sided')
```

Функция возвращает коэффициент корреляции в диапазоне от -1 до 1 и pvalue для интерпретации значимости коэффициента.

Критерий Фишера. Используется для проверки гипотезы о равенстве дисперсий двух нормально распределенных совокупностей. Критерий будет применен к датасету «Фрукты» к выборкам «Цитрусовые» и «Косточковые».

Условия применения данного критерия. Две независимые выборки. Данные должны быть числовыми и не содержать пропусков. Предполагается, что данные имеют нормальное распределение. Объем выборок может быть разным, но желательно не слишком малым ($n \geq 4$).

Нулевая гипотеза (H_0):

$$\sigma_x^2 = \sigma_y^2$$

Альтернативная гипотеза (H_1):

$$\sigma_x^2 \neq \sigma_y^2$$

Команда в Python не встроена напрямую, но вычисляется вручную:

```
var1 = np.var(x, ddof=1);  
var2 = np.var(y, ddof=1);  
f_stat = max(var1, var2) / min(var1, var2);  
p = 2 * (1 - stats.f.cdf(f_stat, dfn, dfd))
```

где:

```
dfn = n_max-1,  
dfd = n_min-1.
```

Интерпретация результатов заключается в следующем, если p -значение > 0.05 , то нет оснований отвергать H_0 : дисперсии можно считать равными. Если p -значение ≤ 0.05 , то H_0 отвергается: дисперсии существенно различаются.

t-критерий Стьюдента. Используется для проверки гипотезы о равенстве математических ожиданий двух нормально распределенных совокупностей. Критерий будет применен к датасету «Фрукты» к выборкам «Цитрусовые» и «Косточковые».

Условия применения данного критерия. Две независимые выборки. Данные должны быть числовыми и не содержать пропусков. Распределение данных должно быть нормальным. Дисперсии выборок должны быть равны.

Нулевая гипотеза (H_0):

$$\mu_x = \mu_y$$

Альтернативная гипотеза (H_1):

$$\mu_x \neq \mu_y$$

Команда в Python:

```
scipy.stats.ttest_ind(x, y, equal_var=True,  
alternative='two-sided', nan_policy='propagate')
```

Параметры: `equal_var=True` (равные дисперсии), `alternative='two-sided'` (двусторонняя), `nan_policy='propagate'` (игнорировать NaN).

Интерпретация результатов заключается в следующем, если p -значение > 0.05 , то нет оснований отвергать H_0 : средние значения можно считать равными. Если p -значение ≤ 0.05 , то H_0 отвергается: средние значения существенно различаются.

4 Описание результатов проверки для каждого критерия с интерпретацией

Критерий Шапиро–Уилка применяется для проверки гипотезы о нормальности распределения данных в выборке. Нормальность распределения является ключевым условием для применения параметрических методов статистического анализа, таких как t -критерий Стьюдента или F -критерий Фишера. В контексте анализа сельскохозяйственных данных нормальность может свидетельствовать о стабильности производственных процессов, равномерности распределения объёмов или долей продукции, а также об отсутствии существенных выбросов или сезонных колебаний.

Проверка критерия была применена к выборкам «Плодовые овощи», «Корнеплоды», «Листовые овощи» из датасета «Овощи» и к выборкам «Цитрусовые», «Косточковые» из датасета «Фрукты». Полученные результаты для «Плодовые овощи»: p -value равно 0.0000, значит распределение ненормальное. Полученные результаты для «Корнеплоды»: p -value равно 0.0054, значит распределение ненормальное. Полученные результаты для «Листовые овощи»: p -value равно 0.0009, значит распределение ненормальное. Полученные результаты для «Цитрусовые»: p -value равно 0.9135, значит распределение нормальное. Полученные результаты для «Косточковые»: p -value равно 0.8214, значит распределение нормальное.

Во всех категориях овощей (плодовые, корнеплоды, листовые) p -value значительно ниже порогового значения 0.05, что свидетельствует о статистически значимом отклонении от нормального распределения. Это может быть обусловлено сезонными колебаниями в производстве, наличием выбросов (например, резких скачков в объёме), групповой структурой данных (разные регионы, технологии)

Можно сделать вывод, что для анализа этих данных необходимо использовать непараметрические критерии, устойчивые к форме распределения, такие как Манна–Уитни, медианный критерий, Краскелл–Уоллис и Спирмен.

Для категорий цитрусовых и косточковых фруктов p -value значительно превышает 0.05, что не даёт оснований отвергать гипотезу о нормальности. Это может

свидетельствовать о стабильности производственных процессов, отсутствии выраженных выбросов, однородности условий выращивания

Можно сделать вывод, что данные по фруктам соответствуют нормальному распределению, что позволяет применять параметрические методы анализа, включая F-критерий и t-критерий.

В аграрной статистике нормальность может отражать стабильность производства, так как нормальное распределение часто возникает при равномерных условиях. Также гетерогенность среды, потому что ненормальность может сигнализировать о влиянии внешних факторов (почва, климат, технологии). Проверка критерия важна для выбора методов анализа: от нормальности зависит корректность статистических выводов.

Таким образом, результаты критерия Шапиро–Уилка позволяют не только выбрать адекватные методы анализа, но и сделать предварительные выводы о природе и устойчивости производственных процессов в разных категориях сельскохозяйственной продукции.

U-критерий Манна–Уитни применяется для проверки гипотезы о равенстве распределений двух независимых выборок, когда отсутствует уверенность в нормальности данных. Это непараметрический аналог t-критерия Стьюдента, устойчивый к выбросам и асимметрии. В контексте анализа сельскохозяйственной продукции данный критерий позволяет определить, различаются ли статистически значимо показатели (например, объёмы или доли производства) между двумя категориями продукции, даже если распределения не являются нормальными.

Проверка критерия была применена к выборкам «Плодовые овощи» и «Корнеплоды» из датасета «Овощи» и к выборкам «Цитрусовые» и «Косточковые» из датасета «Фрукты». Полученные результаты для «Плодовые овощи» и «Корнеплоды»: p-value равно 0.4936, значит распределения одинаковы. Полученные результаты для «Цитрусовые» и «Косточковые»: p-value равно 0.1714, значит распределения одинаковы.

Интерпретация полученных результатов для «Плодовые овощи» и «Корнеплоды». Значение p-value = 0.4936 значительно превышает уровень значимости $\alpha = 0.05$. Это означает, что нет статистических оснований отвергать нулевую гипотезу. Следовательно, распределения показателей (например, объёма или доли производства) для плодовых овощей и корнеплодов не различаются статистически значимо. В аграрном контексте это может указывать на схожесть условий производства, одинаковую степень стабильности или сравнимую рыночную структуру этих подкатегорий.

Интерпретация полученных результатов для «Цитрусовые» и «Косточковые». Значение p-value = 0.1714 также превышает порог 0.05. Это свидетельствует о

статистическом сходстве распределений между цитрусовыми и косточковыми фруктами. Несмотря на различия в биологических характеристиках, объёмы или доли производства этих групп демонстрируют схожую структуру распределения, что может быть связано с едиными агротехническими подходами, сезонностью или рыночным спросом.

Результаты U-критерия Манна–Уитни подтверждают, что внутри групп овощей и фруктов отсутствуют статистически значимые различия между подкатегориями. Это позволяет рассматривать каждую группу (овощи, фрукты) как относительно однородную в рамках дальнейшего анализа. Также это может служить основанием для агрегирования данных или применения единых стратегий агропроизводства и маркетинга внутри каждой группы.

Медианный критерий относится к классу непараметрических критериев однородности и применяется для проверки гипотезы о равенстве медиан двух или более независимых выборок. В отличие от параметрических методов, он не требует нормальности распределения и устойчив к выбросам и асимметрии. В контексте аграрной статистики медианный критерий позволяет определить, различаются ли центральные тенденции (медианы) между категориями продукции, даже если распределения имеют разную форму или содержат экстремальные значения.

Критерий был применен к выборкам «Плодовые овощи» и «Корнеплоды» из датасета «Овощи». Полученные результаты: p -value равно 0.6171, общая медиана равна 288.550, значит медианы равны.

Значение p -value = 0.6171 значительно превышает пороговое значение $\alpha = 0.05$, что не даёт оснований отвергать нулевую гипотезу. Это означает, что центральные значения (медианы) показателей (например, объёма или доли производства) для плодовых овощей и корнеплодов статистически не различаются. Общая медиана по объединённым данным составляет 288.550, и обе категории демонстрируют сходную концентрацию значений вокруг этой точки.

Если подойти к интерпретации полученных результатов проверки критерия с аграрной точки зрения, то сравнение медиан позволяет оценить типичное значение показателя, не зависящее от выбросов. Равенство медиан между плодовыми и корнеплодами может свидетельствовать о схожем уровне производственной активности, сравнимой рыночной значимости и единых условиях выращивания и распределения.

Критерий Краскела–Уоллиса относится к классу непараметрических критериев однородности и применяется для проверки гипотезы о равенстве распределений в трёх и более независимых выборках. Он является обобщением U-критерия Манна–Уитни и используется, когда данные не соответствуют нормальному распределению или содержат

выбросы. В контексте аграрной статистики данный критерий позволяет определить, существуют ли статистически значимые различия между категориями сельскохозяйственной продукции (например, по объёму или доле производства), даже если распределения имеют разную форму.

Проверка критерия была применена к выборкам «Плодовые овощи», «Корнеплоды», «Листовые» из датасета «Овощи». Полученные результаты: $p\text{-value}$ равно 0.0261, значит группы различаются.

Значение $p\text{-value} = 0.0261$ меньше порогового уровня значимости ($\alpha = 0.05$), что даёт основание отвергнуть нулевую гипотезу. Это означает, что по крайней мере одна из категорий овощей (плодовые, корнеплоды, листовые) имеет распределение, статистически отличающееся от других. Критерий не указывает, какие именно группы различаются, но подтверждает наличие гетерогенности внутри группы «овощи».

Аграрная интерпретация говорит о том, что различия могут быть обусловлены биологическими особенностями культур (разные циклы роста, урожайность), технологией производства (открытый/закрытый грунт, механизация), сезонностью и климатическими условиями, рыночным спросом и логистикой.

Критерий Спирмена применяется для оценки степени монотонной связи между двумя переменными. В отличие от параметрического критерия Пирсона, он не требует нормальности распределения и устойчив к выбросам. Это делает его особенно полезным в аграрной статистике, где данные могут быть асимметричными, сезонными или содержать экстремальные значения. В данном исследовании критерий Спирмена используется для анализа взаимосвязи между объёмом производства и долей продукции в общей структуре по категориям овощей и фруктов.

Проверка критерия была применена ко всему датасету «Овощи», а именно объём и доля, и ко всему датасету «Фрукты», а именно объём и доля. Полученные результаты для датасета «Овощи»: $p\text{-value}$ равно 0.0000, коэффициент ρ равен 0.961, значит есть сильная положительная связь. Полученные результаты для датасета «Фрукты»: $p\text{-value}$ равно 0.0000, коэффициент ρ равен 1.000, значит есть абсолютная положительная связь.

Для датасета «Овощи» коэффициент Спирмена $\rho = 0.961$ указывает на очень сильную положительную монотонную связь между объёмом производства и долей продукции. Это означает, что по мере увеличения объёма производства возрастает и доля данной категории в общей структуре, и наоборот. Значение $p = 0.0000$ подтверждает статистическую значимость связи. В аграрном контексте это может свидетельствовать о пропорциональном распределении ресурсов между категориями, стабильной рыночной позиции овощных подкатегорий, отсутствии перекосов или дисбаланса в производственной структуре.

Для датасета «Фрукты» коэффициент $\rho = 1.000$ указывает на идеальную положительную монотонную связь, а именно все ранги объёма и доли совпадают. Это означает, что распределение фруктов по объёму полностью соответствует их долевым участию в общей структуре. Значение $p = 0.0000$ подтверждает высокую статистическую значимость. Такая связь может быть обусловлена целенаправленным планированием производства, жёсткой структурной привязкой объёмов к рыночной доле, высокой управляемостью и прогнозируемостью фруктового сегмента.

В обеих группах (овощи и фрукты) выявлена статистически значимая положительная связь между объёмом и долей продукции. Это подтверждает, что увеличение объёма сопровождается ростом доли, что важно для оптимизации производственных стратегий, прогнозирования рыночной динамики, оценки эффективности распределения ресурсов.

F-критерий Фишера применяется для проверки гипотезы о равенстве дисперсий двух независимых выборок, предполагая, что данные в каждой выборке имеют нормальное распределение. Он используется как предварительный этап перед применением t-критерия Стьюдента, поскольку классический t-тест требует равенства дисперсий. В аграрной статистике дисперсия отражает степень разброса значений. Например, объёмов или долей производства внутри категории продукции. Сравнение дисперсий позволяет оценить стабильность и однородность производственных процессов между группами.

Проверка критерия была применена к выборкам «Цитрусовые» и «Косточковые» из датасета «Фрукты». Полученные результаты: p-value равно 0.0013, F-статистика равна 39.586, значит дисперсии не равны.

Значение $F = 39.586$ указывает на существенное различие в дисперсиях между двумя группами фруктов. Значение $p = 0.0013$ значительно ниже порогового уровня значимости ($\alpha = 0.05$), что даёт основание отвергнуть нулевую гипотезу. Это означает, что разброс значений внутри категорий цитрусовых и косточковых фруктов статистически различается. Аграрная интерпретация говорит о том, что более высокая дисперсия может свидетельствовать о гетерогенности условий производства (разные регионы, сорта, технологии), сезонных колебаниях урожайности, разной степени рыночной стабильности. Например, если косточковые фрукты имеют более высокую дисперсию, это может означать, что их производство менее предсказуемо, чем у цитрусовых.

t-критерий Стьюдента применяется для проверки гипотезы о равенстве математических ожиданий (средних значений) двух независимых выборок. Он используется при условии, что данные в выборках имеют нормальное распределение. В классической форме t-критерий предполагает равенство дисперсий, однако при их неравенстве применяется поправка Уэлча, которая корректирует степени свободы и делает тест

устойчивым к гетерогенности. В аграрной статистике сравнение средних значений между категориями продукции (например, цитрусовыми и косточковыми фруктами) позволяет оценить типичный уровень объёма или доли производства, выявить структурные различия и определить приоритетные направления развития.

Проверка критерия была применена к выборкам «Цитрусовые» и «Косточковые» из датасета «Фрукты». Полученные результаты: $p\text{-value}$ равно 0.0983, значит средние значения не различаются статистически.

Значение $p\text{-value} = 0.0983$ превышает пороговое значение $\alpha = 0.05$, что не даёт оснований отвергать нулевую гипотезу. Это означает, что средние значения анализируемого признака (например, объёма или доли производства) для цитрусовых и косточковых фруктов статистически не различаются. Применение поправки Уэлча было обосновано ранее выявленным неравенством дисперсий (по результатам F-критерия), что делает выводы более надёжными. Аграрная интерпретация говорит о том, что равенство средних значений может свидетельствовать о сравнимом уровне производственной активности между категориями, схожей рыночной значимости цитрусовых и косточковых фруктов, сбалансированной структуре распределения ресурсов в фруктовом сегменте. Несмотря на различие в дисперсиях, типичные значения показателя остаются сопоставимыми, что важно для планирования, прогнозирования и оценки эффективности.

Результаты t-критерия с поправкой Уэлча подтверждают, что средние значения признака для цитрусовых и косточковых фруктов не различаются статистически значимо. Это позволяет рассматривать обе категории как однородные по уровню типичных значений, несмотря на различие в разбросе. Выводы могут быть использованы для агрегирования данных в рамках общей модели, сравнительного анализа эффективности, оптимизации производственных стратегий.

Для ознакомления со всеми вычислениями можно перейти по ссылке:

https://github.com/FrogFromTheSwamp/information_processing

По данной ссылке расположен репозиторий «information_processing», в папке «КР-2» расположен файл «КР-2.ipynb» с вычислениями. Также в данной папке расположены все файлы таблицами.

Выводы

Проведенное статистическое исследование объемов производства овощной и фруктовой продукции в Турции за 2025 год с использованием комплекса параметрических и непараметрических критериев подтвердило ряд ключевых структурных и динамических особенностей аграрного сектора.

Анализ распределения и применимости критериев. Предварительная проверка критерием Шапиро–Уилка выявила принципиальное различие в распределении данных: объемы производства овощей подчиняются ненормальному распределению, что обусловлено высокой асимметрией и доминирующим влиянием отдельных культур, тогда как данные по фруктам демонстрируют нормальное распределение, свидетельствующее о более стабильных и однородных производственных процессах в этом сегменте. Данное различие обосновало необходимость использования смешанного методологического подхода, включающего непараметрические критерии для овощей и параметрические для фруктов.

Оценка различий между категориями. Применение непараметрического критерия Краскела–Уоллиса к категориям овощей («Плодовые», «Корнеплоды», «Листовые») установило статистически значимые различия в распределении объемов, что указывает на высокую гетерогенность овощеводческого сектора и требует дифференцированного планирования. В то же время, критерии Манна–Уитни и Медианный не выявили статистически значимых различий между плодовыми овощами и корнеплодами, позволяя рассматривать эти две подгруппы как однородные по типичному уровню производственной активности.

В сегменте фруктов, несмотря на подтвержденное F-критерием Фишера существенное неравенство дисперсий (различную степень волатильности или предсказуемости производства цитрусовых и косточковых), t-критерий Стьюдента с поправкой Уэлча показал, что средние значения объемов статистически не различаются. Это свидетельствует о сопоставимом уровне производственной активности и рыночной значимости этих двух категорий при разном уровне риска, связанного с разбросом показателей.

Анализ взаимосвязи. Коэффициент корреляции Спирмена выявил сильную положительную связь (0.961 для овощей) и идеальную положительную связь (1.000 для фруктов) между объемом производства и долей продукции. Это подтверждает, что в аграрной экономике Турции увеличение физического объема производства прямо пропорционально ведет к росту рыночной доли.

Полученные результаты позволяют сделать вывод о необходимости применения дифференцированных стратегий управления рисками. Для овощного сегмента, характеризующегося гетерогенностью, требуется таргетированная поддержка по категориям, а для фруктового сегмента ключевым направлением является снижение волатильности в группе с большей дисперсией. Обнаруженная устойчивая и сильная прямая зависимость между объемом производства и долей рынка указывает на то, что

стратегические инвестиции, направленные на увеличение объемов, являются наиболее эффективным рычагом для повышения конкурентоспособности и укрепления рыночных позиций в обеих отраслях.