



А.А. Чалганова

**Выполнение задачи однофакторного
дисперсионного анализа
с использованием табличного
процессора Excel**

**Учебное пособие
по дисциплине «Статистика»**

УДК 338.5

ББК 65.05

Ч-16

Чалганова, Алла Анатольевна.

Выполнение задачи однофакторного дисперсионного анализа с использованием табличного процессора Excel. Учебное пособие по дисциплине «Статистика» /

А.А. Чалганова. – [Текст : электронный]. – Санкт-Петербург : РГГМУ, 2022. – 52 с.

В учебном пособии по выполнению задачи однофакторного дисперсионного анализа с использованием табличного процессора Excel по дисциплине «Статистика» особое внимание уделяется изучению связей методов дисперсионного анализа в вычислительной среде табличного процессора Excel. При экономических и социальных исследованиях возникают вопросы исследования влияния одной или нескольких качественных переменных на одну зависимую количественную переменную. Данная задача решается с применением однофакторного дисперсионного анализа для изучения взаимосвязей. Изучение связей является одной из важнейших аналитических задач, решаемых статистикой.

Простейшим случаем дисперсионного анализа является однофакторный анализ. Большой объем вычислений, который требуется для проведения анализа, предполагает использования вычислительной техники и соответствующего программного обеспечения.

Учебное пособие предназначено для студентов всех форм обучения по направлению подготовки 38.03.01 «Экономика».

Учебное пособие по дисциплине «Статистика» одобрено на заседании кафедры экономики предприятия природопользования и учетных систем от 15 июня 2022 г., протокол № 12.

© Чалганова А.А., 2022

© Российский государственный

гидрометеорологический университет (РГГМУ), 2022

Введение

Среди аналитических задач, решаемых статистикой, важнейшей является изучение связей. При экономических и социальных исследованиях возникают вопросы о причинах, которые привели к получению данного результата, о факторах, которые определили его появление. Ответы на эти вопросы можно получить с помощью дисперсионного и регрессионного анализа.

Если необходимо оценить различие между характеристиками нескольких групп, используют дисперсионный анализ, который, однако, предполагает существенный объем вычислений, что требует использования вычислительной техники и соответствующего программного обеспечения. Удобным инструментом для решения многих статистических и эконометрических задач является табличный процессор Excel. Будучи универсальной вычислительной средой, Excel позволяет реализовать алгоритм решения либо путем программирования как арифметических, так и логических выражений в ячейках электронной таблицы, либо путем обращения к стандартным функциям и модулям.

Основной целью данного пособия является изложение численной методики решения задачи дисперсионного анализа в вычислительной среде табличного процессора Excel - установления значимости различия между средними значениями зависимой количественной переменной по группам фактора. Достигается это с помощью разложения общей дисперсии зависимой переменной на составляющие: дисперсию за счет разбиения на группы и дисперсию за счет остальных факторов. Анализируя эти компоненты дисперсии, можно оценить долю воздействия каждого индикатора на исследуемую величину.

1. Правило сложения дисперсий

Для того, чтобы оценить влияние отдельных факторов на колеблемость индивидуальных значений результативного признака y , используют дисперсионный анализ. При этом кроме общей средней (среднего значения признака для всей статистической совокупности \bar{y}_0) исчисляют средние значения признака по отдельным группам этой совокупности (групповые или частные средние \bar{y}_j) и три показателя дисперсии, называемые: общая дисперсия; межгрупповая дисперсия; средняя внутригрупповая дисперсия.

Величина общей дисперсии (σ_0^2) характеризует вариацию признака под влиянием всех факторов, формирующих уровень признака у единиц данной совокупности, и определяется по формуле:

$$\sigma_0^2 = \frac{\sum (y - \bar{y}_0)^2 \cdot f}{\sum f} = \overline{y^2} - (\bar{y}_0)^2 \quad (1)$$

где \bar{y}_0 - общая средняя арифметическая для всей изучаемой совокупности.

Межгрупповая дисперсия (дисперсия групповых средних - δ^2) отражает систематическую вариацию, т.е. различия в величине изучаемого признака, которые возникают под влиянием фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum (\bar{y}_j - \bar{y}_0)^2 n_j}{\sum n_j} \quad (2)$$

где \bar{y}_j - средняя арифметическая по отдельной группе;

n_j - число единиц в определенной группе.

Средняя внутригрупповая дисперсия ($\overline{\sigma^2}$) характеризует случайную вариацию, возникающую под влиянием других, неучтенных факторов, и не зависит от признака-фактора, положенного в основу группировки.

$$\overline{\sigma^2} = \frac{\sum \sigma_j^2 \cdot n_j}{\sum n_j} \quad (3)$$

где σ_j^2 - дисперсия по отдельной группе.

Дисперсия по отдельной группе рассчитывается по формуле:

$$\sigma_j^2 = \frac{\sum (y - \bar{y}_j)^2 \cdot f}{\sum f} \quad (4)$$

Указанные дисперсии взаимосвязаны. Их связь между собой выражена равенством, которое называется правилом сложения дисперсий: величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой. Математическое выражение для правила сложения дисперсий представлено ниже:

$$\sigma_o^2 = \delta^2 + \overline{\sigma^2} \quad (5)$$

Опираясь на это правило, можно определить, какая часть (доля) общей дисперсии складывается под влиянием признака-фактора, положенного в основу группировки (δ^2).

2. Дисперсионный анализ

Процедура дисперсионного анализа состоит в определении соотношения систематической (межгрупповой) дисперсии к случайной (внутригрупповой)

дисперсии в измеряемых данных. В качестве показателя изменчивости используется сумма квадратов отклонения значений признака от среднего: SS (от английского *Sum of Squares*). Можно показать, что общая сумма квадратов SS_{total} раскладывается на межгрупповую сумму квадратов SS_{bg} и внутригрупповую сумму квадратов SS_{wg} .

Дисперсионный анализ применяется для сравнения условных математических ожиданий (или средних величин) признака-результата при условии того или иного значения одного или нескольких признаков-факторов. Дисперсионный анализ используют для изучения взаимосвязей, исследования влияния одной или нескольких качественных переменных на одну зависимую количественную переменную. Основной целью подобного исследования является установление значимости различия между средними значениями зависимой количественной переменной по группам фактора. Достигается это с помощью разложения общей дисперсии зависимой переменной на составляющие: дисперсию за счет разбиения на группы и дисперсию за счет остальных факторов. Анализируя эти компоненты дисперсии, можно оценить долю воздействия каждого индикатора (фактора) на исследуемую величину.

В зависимости от количества факторов, определяющих вариацию признака-результата, дисперсионный анализ подразделяют на однофакторный и многофакторный.

Простейшим случаем дисперсионного анализа является однофакторный анализ, в котором n единиц совокупности распределены на m групп по значениям одного фактора.

Основная идея однофакторного дисперсионного анализа заключается в сравнении дисперсии исследуемого признака, вызванной действием фактора, с дисперсией ошибок измерения этого признака. Если различие между ними значимо, то фактор оказывает существенное влияние на исследуемый признак.

Общая вариация SSE (*the Sum of Squared estimate of Errors*) при этом подразделяется на часть, объясняющую различия между группами SSB (*between groups*) – межгрупповую сумму квадратов отклонений и часть, объясняющую различия между

единицами совокупности внутри группы SSW (*within groups*) – *внутригрупповую сумму квадратов отклонений*.

Любая сумма квадратов отклонений связана с числом степеней свободы df (*degrees of freedom*), т.е. с числом независимого варьирования переменных (числом свободы независимого варьирования признака).

Число степеней свободы связано с числом единиц совокупности n и с числом определяемых по ней констант. Применительно к исследуемой проблеме число степеней свободы должно показать, сколько независимых отклонений из n возможных $[(y_1 - \bar{y}), (y_2 - \bar{y}), \dots, (y_n - \bar{y})]$ требуется для образования данной суммы квадратов. Так, для общей суммы квадратов $\sum (y - \bar{y})^2$ необходимо $(n - 1)$ независимых отклонений, ибо по совокупности из n единиц после расчета среднего уровня свободно варьируют лишь отклонения, число которых на единицу меньше, чем общее число отклонений (или единиц совокупности) n , и составляет $(n - 1)$. Например, имеем ряд значений y : 1,2,3,4,5. Среднее из них равно 3, тогда n отклонений от среднего составят: -2; -1; 0; 1; 2. Поскольку $\sum (y - \bar{y}) = 0$, то свободно варьируют лишь четыре отклонения, а пятое отклонение может быть определено, если четыре предыдущие известны.

Общее число степеней свободы df_{total} также раскладывается на межгрупповое число степеней свободы df_{bg} (*between groups*) и внутригрупповое число степеней свободы df_{wg} (*within groups*):

$$df_{total} = df_{bg} + df_{wg}, \quad (6)$$

где $df_{total} = n - 1;$

$$df_{bg} = m - 1;$$

$$df_{wg} = n - m;$$

n – число наблюдений;

m – число групп.

Все эти величины объединяются в таблицу дисперсионного анализа, содержание которой иллюстрирует таблица 1.

Таблица 1

Однофакторный дисперсионный анализ

Источник вариации	df	SS	MS	F	Значимость F
Межгрупповая (<i>between groups</i>)	$m-1$	SSB	$SSB/(m-1)$	$\frac{SSB / (m-1)}{SSW / (n-m)}$	
Внутригрупповая (<i>within groups</i>)	$n-m$	SSW	$SSW/(n-m)$		
Общая (<i>total</i>)	$n-1$	$SSB+SSW$			

В таблице 1 использованы следующие обозначения:

df – число степеней свободы (*degrees of freedom*);

SS – суммы квадратов отклонений (*the Sum of Squared*);

MS – средняя сумма квадратов отклонений, приходящаяся на одну степень свободы (от англ. *Mean Squared* – средний квадрат);

F – F-критерий Фишера.

n - число единиц анализируемой совокупности;

m – число групп, по которым распределены единицы совокупности.

В основе дисперсионного анализа лежит разделение дисперсии на компоненты. Связь между общей дисперсией σ_o^2 , средней из внутригрупповых дисперсий $\overline{\sigma^2}$ и межгрупповой дисперсией δ^2 отражает правило сложения дисперсий:

$$\sigma_o^2 = \delta^2 + \overline{\sigma^2} \quad (7)$$

$$\text{Где } \sigma_o^2 = \frac{SSE}{n}; \quad \overline{\sigma^2} = \frac{SSW}{n}; \quad \delta^2 = \frac{SSB}{n}.$$

Или можно записать:

$$SSE = SSW + SSB, \quad (8)$$

Общая сумма квадратов отклонений *SSE* (*the Sum of Squared estimate of Errors*) рассчитывается по формуле:

$$SSE = \sum (y - \bar{y}_o)^2, \quad (9)$$

где \bar{y}_o - общая средняя арифметическая для всей изучаемой совокупности.

Межгрупповая сумма квадратов отклонений *SSB* (*between groups*) – рассчитывается по формуле:

$$SSB = \sum (\bar{y}_j - \bar{y}_o)^2, \quad (10)$$

где \bar{y}_j - средняя арифметическая по отдельной группе;

Внутригрупповая сумма квадратов отклонений *SSW* (*within groups*) рассчитывается следующим образом:

$$SSW = \sum_j \sum_i (y_{ij} - \bar{y}_j)^2 \quad (11)$$

Далее рассчитывают межгрупповой средний квадрат отклонений (на одну степень свободы) *MSB*, а также внутригрупповой средний квадрат отклонений (на одну степень свободы) *MSW*:

$$MSB = \frac{SSB}{(m-1)}; \quad (12)$$

$$MSW = \frac{SSW}{(n-m)}; \quad (13)$$

Если средние по группам, которые образованы в соответствии со значениями фактора x , отличаются в пределах случайных различий, то межгрупповой средний квадрат отклонений на одну степень свободы и внутригрупповой средний квадрат отклонений на одну степень свободы будут одинаковыми, т.е. $MSB = MSW$.

Введя обозначения S_1^2 для MSB и S_2^2 для MSW получим:

$$S_1^2 = \frac{SSB}{(m-1)} = MSB \quad (14)$$

$$S_2^2 = \frac{SSW}{n-m} = MSW \quad (15)$$

Таким образом, если различия внутригрупповых средних носят случайный характер, т.е. не связаны с изменением факторного признака, то $S_1^2 = S_2^2$.

Это положение соответствует нулевой гипотезе $H_0: S_1^2 = S_2^2$.

Статистической гипотезой называется предположение о свойстве генеральной совокупности, которое можно проверить, опираясь на данные выборки. Обозначается гипотезе буквой H (от лат. *hypothesis*).

Гипотеза о том, что две совокупности, сравниваемые по одному или нескольким признакам, не отличаются, называется *нулевой гипотезой* (или *нуль-гипотезой*). Она обозначается H_0 . При этом предполагается, что действительное различие сравниваемых величин равно нулю, а выявленное по данным выборки отличие от нуля носит случайный характер. Например, $H_0: \mu_1 = \mu_2$ (или же для нашей задачи $S_1^2 = S_2^2$).

Нулевая гипотеза отклоняется тогда, когда по выборке получается результат, который при истинности выдвинутой нулевой гипотезы маловероятен. Границей невозможного или маловероятного обычно считают $\alpha=0,05$ (т.е. 5%), или 0,01 (т.е. 1%), или 0,001 (т.е. 0,1%).

Как правило, значения критерия в статистическо-математических таблицах рассчитаны для вероятностей ошибки 0,05 и 0,01.

Статистическим критерием называют определенное правило, устанавливающее условия, при которых проверяемую нулевую гипотезу следует либо отклонить, либо не отклонять. Критерий проверки статистической гипотезы определяет, противоречит ли выдвинутая гипотеза фактическим данным или нет.

Проверка статистических гипотез складывается из следующих этапов:

- формулируется в виде статистической гипотезы задача исследования;
- выбирается статистическая характеристика гипотезы;
- выбираются испытываемая и альтернативная гипотезы на основе анализа возможных ошибочных решений и их последствий;
- определяются область допустимых значений, критическая область, а также критическое значение статистического критерия (t , F , χ^2) по соответствующей таблице;
- вычисляется фактическое значение статистического критерия;
- проверяется испытываемая гипотеза на основе сравнения фактического и критического значений критерия и в зависимости от результатов проверки гипотеза либо отклоняется, либо не отклоняется.

При испытании гипотезы возможны две ошибки.

Первая – отклонение H_0 , когда она фактически верна. Такая ошибка называется ошибкой I рода, вероятность такой ошибки обозначается α . Вторая – принятие H_1 , когда фактически она ложна. Эта ошибка называется ошибкой II рода и вероятность ее обозначается β . Виды ошибок при испытании гипотез представлены в таблице 2.

Виды ошибок при испытании гипотез

Решение	Правильная гипотеза	
	H_0	H_1
Принять	Нет ошибки	Ошибка II рода
Отклонить	Ошибка I рода	Нет ошибки

Критической областью называется область, попадание значения статистического критерия в которую приводит к отклонению. Вероятность попадания значения критерия в эту область равна принятому уровню значимости.

Область допустимых значений дополняет критическую область. Если значение критерия попадает в область допустимых значений, это свидетельствует о том, что выдвинутая гипотеза не противоречит фактическим данным (H_0 не отклоняется).

Точки, разделяющие критическую область и область допустимых значений, называют *критическими точками* или *границами критической области*.

Если вычисленное значение критерия попадает в критическую область, нулевая гипотеза отклоняется (не принимается), поскольку она противоречит фактическим данным.

Необходимо проверить гипотезу H_0 , т.е. предположение о случайности различий внутригрупповых средних или (что то же самое) о независимости этих различий от фактора (отсутствии связи различий средних по группам с изменением факторного признака).

Правилом проверки нулевой гипотезы H_0 выступает F -критерий Фишера:

$$F = \frac{S_1^2}{S_2^2} \quad (16)$$

где $S_1^2 \geq S_2^2$.

При заданном уровне значимости α и числе степеней свободы межгрупповой df_1 и внутригрупповой df_2 вариации определяется критическое значение F -критерия по таблице значений F -критерия Фишера.

Необходимыми допущениями для корректного применения F -критерия являются *нормальное распределение* значений зависимой (результативной) переменной в каждой группе и *равенство дисперсий* в группах.

3. Пример задачи однофакторного дисперсионного анализа

Рассмотрим пример однофакторного дисперсионного анализа.

Задача 1. Служба качества организации, имеющей сеть пунктов обслуживания и продажи мобильных телефонов и аксессуаров к ним, провела исследование времени работы устройств разных производителей от максимально заряженной батареи без подзарядки при предельно возможной нагрузке. Было взято по пять устройств каждого из четырех производителей. Испытания проводились на стенде (специальном устройстве), где обеспечивалась максимальная нагрузка на устройства и измерялось время до прекращения устройством работы из-за разрядки батареи ниже допустимого уровня. Результаты испытаний мобильных устройств при определении максимального времени работы без подзарядки батареи представлены в таблице 3.

Таблица 3

Время работы батарей мобильных телефонов разных производителей, час

Порядковый номер мобильного устройства	Производитель	Производитель	Производитель	Производитель
	1	2	3	4
1	21,2	25,4	22,4	26
2	19,5	24,3	24	21,7
3	23,1	25,1	21,6	19
4	19,9	24,1	21,5	22,1
5	20,5	24,7	22,6	24,2
Итого	104,2	123,6	112,1	113,0
Среднее арифметическое	20,84	24,72	22,42	22,6

Очевидно, что между выборочными средними времени работы без подзарядки батареи мобильных устройств разных производителей наблюдаются не-большие различия. Можно ли считать их статистически значимыми?

В данной задаче фирма-производитель выступает в качестве признака-фактора (x), влияние которого на надежность работы мобильных устройств (y), т.е. время работы без подзарядки, проверяется.

Для построения таблицы дисперсионного анализа необходимо:

1. Определить выборочные средние времени работы без подзарядки батареи для каждой группы. В последней строке таблицы 3 приведены результаты расчетов средних.
2. Определить среднее время работы без подзарядки батареи для всех мобильных телефонов (по формуле средней арифметической).

Для этого надо просуммировать все 20 значений и разделить сумму на общее количество наблюдений ($n=20$).

$$\bar{y} = \frac{1}{n} \sum_{j=1}^4 \sum_{i=1}^5 y_{ij} = \frac{452.9}{20} = 22.645 \text{ час}$$

Полученное значение – общая средняя.

3. Вычислить суммы квадратов отклонений:

- Межгрупповая сумма квадратов отклонений:

$$SSB = \sum_{j=1}^4 n_j (\bar{y}_j - \bar{y})^2$$

$$SSB = 5 \cdot (20.84 - 22.645)^2 + 5 \cdot (24.72 - 22.645)^2 + 5 \cdot (22.42 - 22.645)^2 + 5 \cdot (22.6 - 22.645)^2 = 38.0815$$

- Внутригрупповая сумма квадратов отклонений:

$$SSW = \sum_{j=1}^4 \sum_{i=1}^5 (y_{ij} - \bar{y}_j)^2$$

$$SSW = (21.4 - 20.84)^2 + (19.6 - 20.84)^2 + \dots + (20.3 - 20.84)^2 + (25.1 - 24.72)^2 + (24.4 - 24.72)^2 + \dots + (24.9 - 24.72)^2 + (22.3 - 22.42)^2 + (24.1 - 22.42)^2 + \dots + (22.6 - 22.42)^2 + (26.1 - 22.6)^2 + (21.5 - 22.6)^2 + \dots + (24.2 - 22.6)^2 = 41.388$$

- Общая сумма квадратов отклонений:

$$SSE = SSB + SSW = \sum_{j=1}^4 \sum_{i=1}^5 (y_{ij} - \bar{y})^2$$

$$SSE = (21.4 - 22.645)^2 + \dots + (20.3 - 22.645)^2 + (25.1 - 22.645)^2 + \dots + (24.9 - 22.645)^2 + (22.3 - 22.645)^2 + \dots + (22.6 - 22.645)^2 + (26.1 - 22.645)^2 + \dots + (24.2 - 22.645)^2 = 79.4695$$

4. Вычислить суммы квадратов отклонений, приходящихся на одну степень свободы:

- межгрупповая сумма квадратов отклонений:

$$S_1^2 = MSB = \frac{SSB}{(m-1)} = \frac{38,0815}{4-1} = \frac{38,0815}{3} = 12,6938$$

- внутригрупповая сумма квадратов отклонений:

$$S_2^2 = MSW = \frac{SSW}{n-m} = \frac{41,388}{20-4} = \frac{41,388}{16} = 2,58675$$

5. Вычислить F-критерий.

$$F = \frac{S_1^2}{S_2^2} = \frac{MSB}{MSW} = \frac{12,6938}{2,58675} = 4,907252$$

Примем уровень значимости α равным 0,05. При заданном уровне значимости и числе степеней свободы межгрупповой и внутригрупповой вариации определим критическое значение F -критерия по таблице значений F -критерия Фишера (Приложение 1).

Для рассматриваемой задачи числитель (MSB) имеет три степени свободы $df_1 = 4 - 1 = 3$, а знаменатель $df_2 = 20 - 4 = 16$. Тогда при уровне значимости 0,05, т.е. при вероятности ошибочного решения 5%, $df_1 = 3$, $df_2 = 16$, верхнее критическое значение F -критерия равно 3,24 (пересечение столбца $df_1 = 3$ и строки $df_2 = 16$).

Вычисленное значение F -критерия равно 4,907252, а табличное $F_{табл}=3,24$, что меньше расчетного (вычисленного). Следовательно, различия между средними значениями времени работы мобильных телефонов разных производителей без подзарядки батареи статистически значимы, т.е. нулевая гипотеза H_0 отклоняется. Заполненная таблица однофакторного дисперсионного анализа для рассматриваемых данных представлена в таблице 4.

Однофакторный дисперсионный анализ

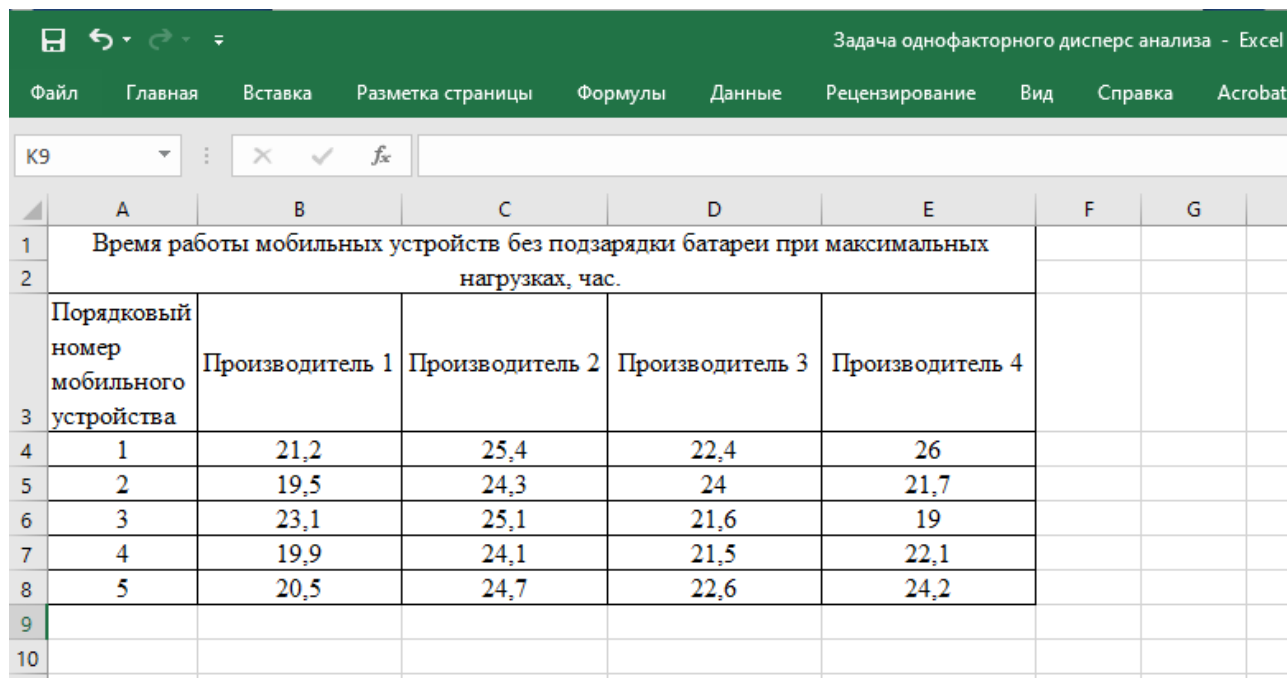
Источник вариации	df	SS	MS	F	Критическое значение $F(\alpha, df_1, df_2)$
Межгрупповая (<i>between groups</i>)	3	38,0815	12,69383	4,90725	3.24
Внутригрупповая (<i>within groups</i>)	16	41,388	2,58675		
Общая (<i>total</i>)	19	79,4695			

Поскольку расчетное значение F -критерия больше табличного, нулевая гипотеза о случайном характере различий групповых средних отклоняется.

4. Решение задачи однофакторного дисперсионного анализа с использованием стандартного табличного процессора Microsoft Excel

Рассмотрим пример решения задачи однофакторного дисперсионного анализа с использованием Microsoft Excel для исходных данных, приведенных в условии уже рассмотренной ранее **задачи 1**.

Сформируем таблицу с исходными данными для анализа на листе Excel. Скриншот листа Excel с расположенной на нем таблицей исходных данных для дисперсионного анализа представлен на рис. 1.



Порядковый номер мобильного устройства	Производитель 1	Производитель 2	Производитель 3	Производитель 4
1	21,2	25,4	22,4	26
2	19,5	24,3	24	21,7
3	23,1	25,1	21,6	19
4	19,9	24,1	21,5	22,1
5	20,5	24,7	22,6	24,2

Рис. 1. Скриншот листа Excel с расположенной на нем таблицей исходных данных для решения задачи.

Для решения задачи дисперсионного анализа воспользуемся функцией «Однофакторный дисперсионный анализ» пакета анализа из надстроек табличного процессора Microsoft Excel.

Следует заметить, что пакет анализа должен быть установлен на компьютере. Если же это не было сделано ранее, нужно его установить. Для этого можно

зайти во вкладку **Надстройки** (**Файл – Параметры – Надстройки – Надстройки Excel - Перейти**), после чего требуется установить флажок для пункта **Пакет анализа**.

Чтобы выполнить однофакторный дисперсионный анализ, надо выполнить следующие действия:

1. Выбрать **Данные** → **Анализ данных** → **Однофакторный дисперсионный анализ**.

Скриншот листа Excel с расположенной на нем таблицей исходных данных для решения задачи и выделенными пунктами меню «Данные» и «Анализ данных» представлен на рисунке 2.

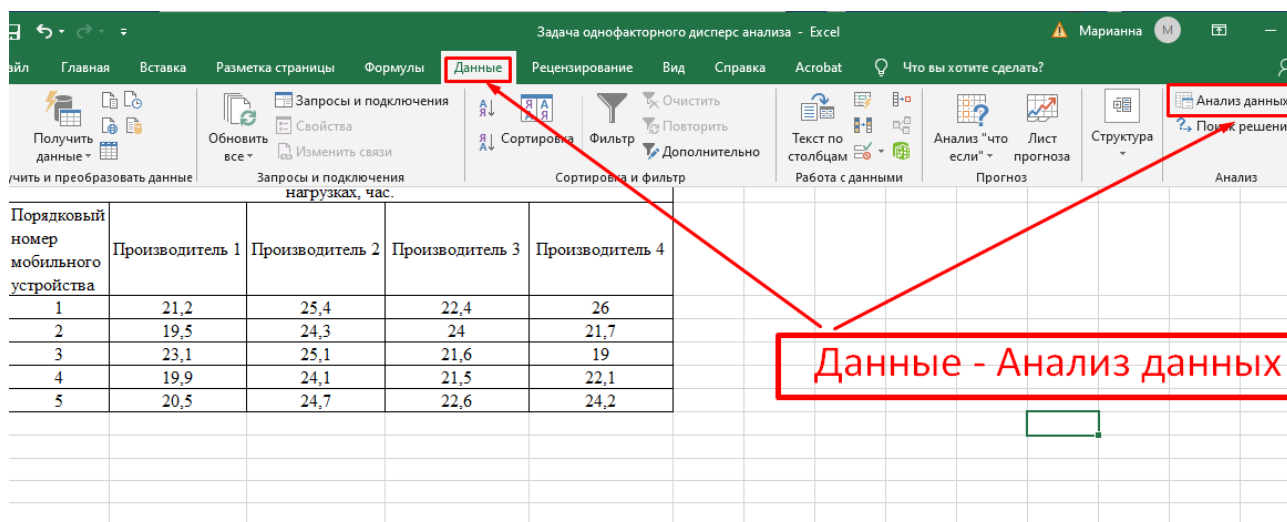


Рис. 2. Скриншот листа Excel с расположенной на нем таблицей исходных данных для решения задачи и выделенными пунктами меню «Данные» и «Анализ данных».

Скриншот окна *Анализ данных* представлен на рисунке 3, где выделен режим *Однофакторный дисперсионный анализ*.

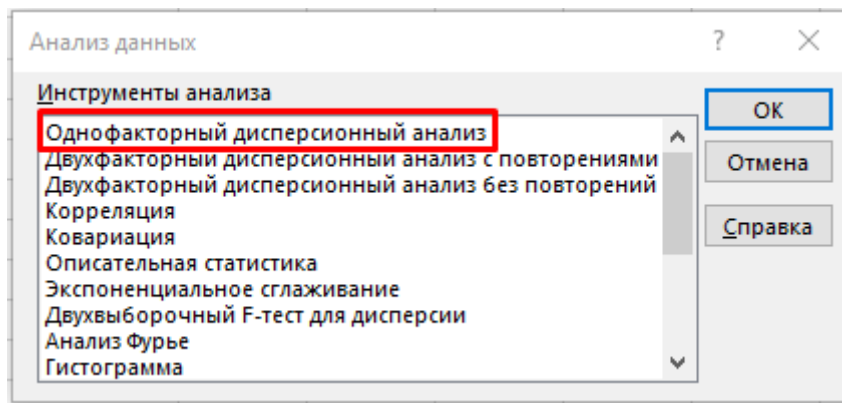


Рис. 3. Окно модуля *Анализ данных*.

После вызова режима *Однофакторный дисперсионный анализ*., который нужно выбрать кликом кнопки **ОК**, на экране появляется диалоговое окно. Вид диалогового окна *Однофакторный дисперсионный анализ* представлен на рис. 4.

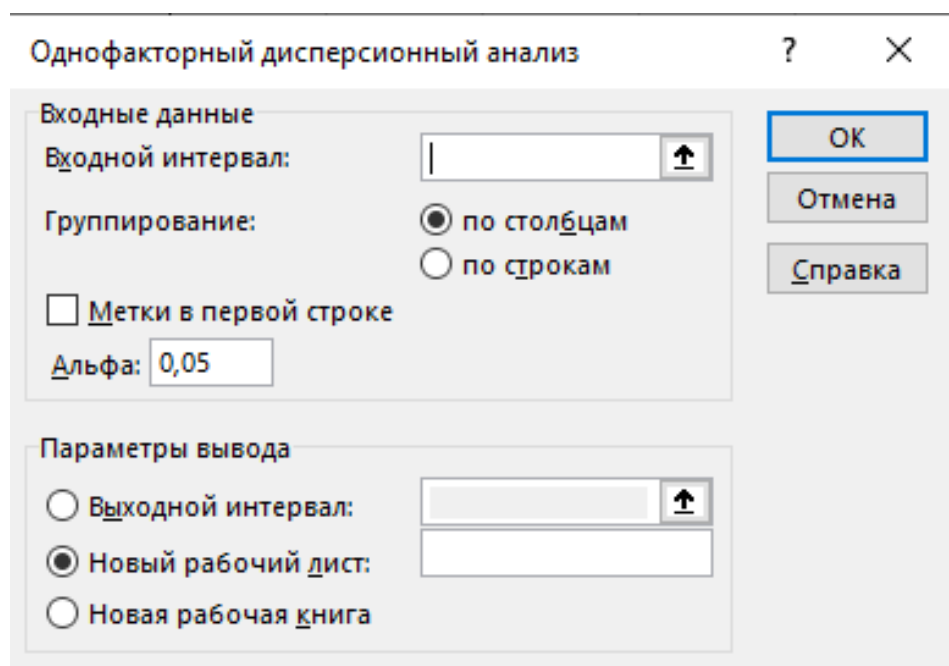
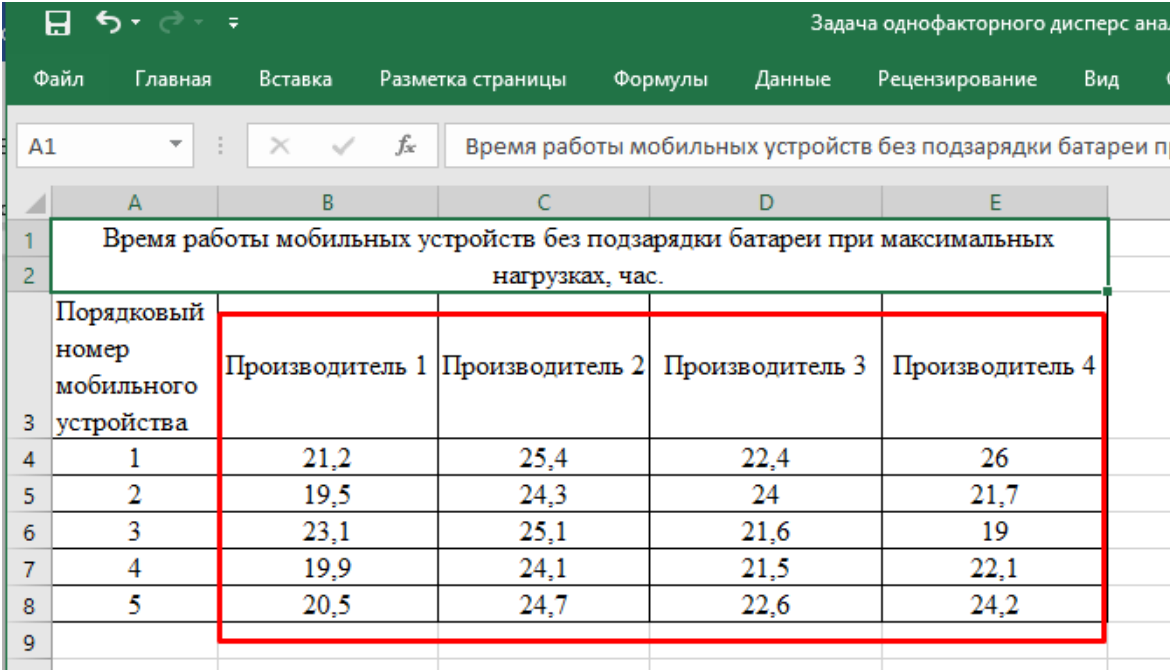


Рис. 4. Диалоговое окно *Однофакторный дисперсионный анализ*.

2. В диалоговом окне **Однофакторный дисперсионный анализ** нужно выполнить следующие действия:

- В окне редактирования **Входной интервал** ввести диапазон переменной;
- Установить переключатель **Группирование** в положение по столбцам, поскольку в таблице исходных данных группирование в соответствии с фактором выполнено по столбцам;
- Установить флажок **Метки**, если в первой строке входного диапазона данных содержится заголовок, чтобы программа не трактовала его как числовые данные.

На рисунке 5 представлен скриншот листа Excel с выделенным диапазоном ячеек исходной таблицы, включающим строку заголовка, и требующего поэтому установления флажка **Метки** во включенное состояние.

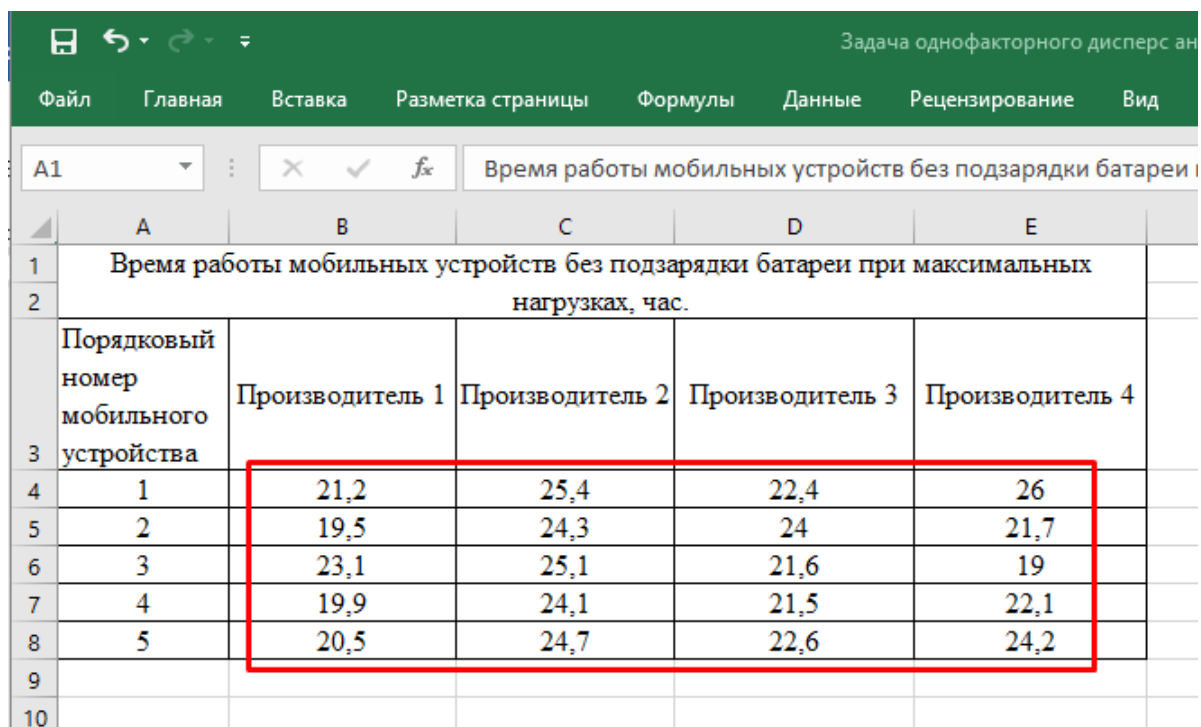


	A	B	C	D	E
1	Время работы мобильных устройств без подзарядки батареи при максимальных				
2	нагрузках, час.				
3	Порядковый номер мобильного устройства	Производитель 1	Производитель 2	Производитель 3	Производитель 4
4	1	21,2	25,4	22,4	26
5	2	19,5	24,3	24	21,7
6	3	23,1	25,1	21,6	19
7	4	19,9	24,1	21,5	22,1
8	5	20,5	24,7	22,6	24,2
9					

Рис. 5. Скриншот листа Excel с выделенным диапазоном ячеек таблицы исходных данных, содержащим строку заголовка.

В случае введения входного интервала, который не содержит заголовка, не требуется включать флажок **Метки**. На рисунке 6 представлен скриншот листа

Excel с выделенным диапазоном ячеек исходной таблицы, не содержащим заголовка, поэтому флажок **Метки** должен быть выключен.



	А	В	С	Д	Е
1	Время работы мобильных устройств без подзарядки батареи при максимальных				
2	нагрузках, час.				
3	Порядковый номер мобильного устройства	Производитель 1	Производитель 2	Производитель 3	Производитель 4
4	1	21,2	25,4	22,4	26
5	2	19,5	24,3	24	21,7
6	3	23,1	25,1	21,6	19
7	4	19,9	24,1	21,5	22,1
8	5	20,5	24,7	22,6	24,2
9					
10					

Рис. 6. Скриншот листа Excel с выделенным диапазоном ячеек таблицы исходных данных, не содержащим строку заголовка.

- В графе редактирования **Альфа** ввести число 0,05;
- В окне редактирования **Выходной интервал** при включении активизируется поле, куда необходимо ввести адрес левой верхней ячейки выходного диапазона, который будет содержать таблицы с результатами вычислений режима **Однофакторный дисперсионный анализ**.

При включении параметра **Новый рабочий лист** открывается новый лист, в который, начиная с ячейки A1, вставляются результаты работы режима **Однофакторный дисперсионный анализ**.

При включении параметра **Новая рабочая книга** открывается новая книга, на первом листе которой, начиная с ячейки A1, вставляются результаты работы режима **Однофакторный дисперсионный анализ**.

- Нажать кнопку **ОК**.

На рисунке 7 представлен скриншот диалогового окна **Однофакторный дисперсионный анализ**, где введен диапазон исходных данных, включающий строку заголовка, поэтому флажок **Метки** включен, выходной интервал задан ячейкой на листе с исходными данными, куда будут вставлены результаты работы режима **Однофакторный дисперсионный анализ**.

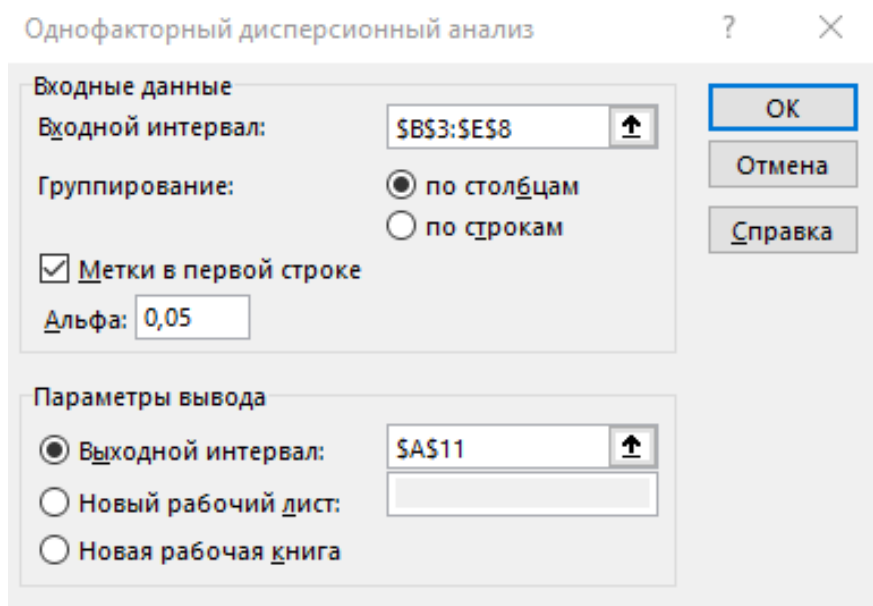


Рис. 7. Скриншот диалогового окна **Однофакторный дисперсионный анализ**.

Результаты дисперсионного анализа, полученные с помощью Microsoft Office Excel, будут совпадать с представленными в таблице 4. Скриншот листа Excel с таблицей однофакторного дисперсионного анализа средних значений времени работы без подзарядки батареи для мобильных телефонов разных производителей представлен на рисунке 8.

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Производитель 1	5	104,2	20,84	2,008		
Производитель 2	5	123,6	24,72	0,292		
Производитель 3	5	112,1	22,42	1,012		
Производитель 4	5	113	22,6	7,035		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	38,0815	3	12,693833	4,907251699	0,013227	3,238871517
Внутри групп	41,388	16	2,58675			
Итого	79,4695	19				

Рис. 8. Скриншот листа Excel с таблицей однофакторного дисперсионного анализа.

В последнем столбце таблицы однофакторного дисперсионного анализа на рисунке 8 показано табличное (критическое) значение F-критерия (α , df_1 , df_2), на основе сравнения с которым делается вывод о статистической значимости изучаемой связи. Проведенный дисперсионный анализ показал наличие статистически значимой связи между средним временем работы мобильного телефона при максимальной нагрузке без подзарядки батареи от фирмы-производителя телефона.

Эту связь можно измерить с помощью эмпирического корреляционного отношения, которое рассчитывается как корень квадратный из отношения факторной дисперсии к общей дисперсии результативного признака:

$$\eta_{yx} = \sqrt{\frac{SSB}{SSE}} \quad (17)$$

Этот показатель принимает значения в интервале от 0 до 1. Связь тем теснее, чем ближе значение η к 1, и наоборот.

Для рассматриваемой задачи:

$$\eta_{yx} = \sqrt{\frac{SSB}{SSE}} = \sqrt{\frac{38,0815}{79,4695}} = \sqrt{0,479196} = 0,69$$

Для вывода (заключения) о тесноте связи можно воспользоваться следующими рекомендациями, сведенными в таблицу 5:

Таблица 5.

Значение эмпирического корреляционного отношения η
и оценка степени тесноты связи результата с фактором

Значение эмпирического корреляционного отношения η	Теснота связи
0,3 и меньше	слабая
0,3 – 0,5	средняя
0,5 – 0,7	умеренно тесная
0,7 и больше	тесная

На основании значения эмпирического корреляционного отношения для нашей задачи $\eta_{yx} = 0,69$ можно сделать вывод, что между средним временем работы мобильного телефона при максимальной нагрузке без подзарядки батареи и фирмой-производителем телефона существует достаточно тесная связь.

Нередко при решении конкретных задач предпочтение отдается показателю η^2_{yx} , который называется коэффициентом детерминации. Этот показатель отражает долю межгрупповой колеблемости в общей и поэтому часто выражается в процентах. По данным рассмотренной задачи можно заключить, что на

48% среднее время работы мобильного телефона при максимальной нагрузке без подзарядки батареи зависит от фирмы-производителя, а на 52% - от других факторов (они в задаче не рассматривались).

Контрольные вопросы

1. Для чего применяют однофакторный дисперсионный анализ. Каковы его особенности?
2. Докажите основное тождество однофакторного дисперсионного анализа.
3. Объясните, каковы процедуры однофакторного дисперсионного анализа?
4. Почему для проверки нулевых гипотез в дисперсионном анализе используется отношение дисперсий?
5. Какие предположения о случайной величине x используются в дисперсионном анализе?
6. Как вычисляется статистика Фишера при проверке гипотезы о влиянии фактора A ?
7. От чего зависит критическое значение статистики Фишера?

Задания для самостоятельной работы

Задача 1. Служба качества организации, имеющей сеть пунктов обслуживания и продажи мобильных телефонов и аксессуаров к ним, провела исследование надежности динамиков мобильных устройств разных производителей. Было взято по пять телефонных аппаратов каждой из четырех выбранных фирм-производителей. Испытания проводились на стенде (специальном устройстве), где обеспечивалась максимальная нагрузка на устройства. Чем дольше работает динамик до выхода из строя, тем он надежнее. Время наработки на отказ измерялось в условных единицах. Результаты испытаний мобильных устройств при определении максимального времени работы до выхода из строя представлены в таблице 6.

Таблица 6

Время работы батарей мобильных телефонов разных производителей, час

Порядковый номер мобильного устройства	Производитель 1	Производитель 2	Производитель 3	Производитель 4
1	21,4	25,1	22,3	26,1
2	19,6	24,4	24,1	21,5
3	23,1	25,3	21,7	18,9
4	19,8	23,9	21,4	22,3
5	20,3	24,9	22,6	24,2
Итого	104,2	123,6	112,1	113,0
Среднее арифметическое	20,84	24,72	22,42	22,6

Очевидно, что между выборочными средними времени работы динамиков разных фирм наблюдаются небольшие различия. Можно ли считать их статистически значимыми (принять уровень значимости $\alpha = 0,05$)?

В этой задаче фирма-производитель выступает в качестве фактора (x), влияние которого на надежность динамиков мобильных телефонов (y) проверяется.

Задача 2.

Группой из пяти экспертов проведена оценка качества объекта при трех значениях уровня фактора, влияние которого на качество объекта изучается. Значения уровня фактора во время оценки экспертам не известны. При уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о влиянии фактора на качество объекта. Экспертные оценки уровня качества объекта приведены в таблицы 7. Задачу решить с применением табличного процессора Excel.

Таблица 7

№	Уровень фактора (Φ)		
	$\Phi 1$	$\Phi 2$	$\Phi 3$
1	18	24	36
2	28	36	12
3	12	28	22
4	14	40	45
5	32	16	40

Задача 3.

Используя анализ однофакторной модели, проверить гипотезу о влиянии методик на оценку качества признака при уровне значимости $\alpha = 0,05$.

Исследуется эффективность обучения тремя различными методами. Студентам дается задание изучить тему «Денежная система в Англии». Для этого 10 студентов конспектируют первоисточник, 10 изучают ее по электронному учебнику, 10 – с помощью обучающих компьютерных программ. По окончании их уровень знаний проверяется с помощью теста, состоящего из 100 вопросов. Результаты представлены в таблице 8.

Задачу решить с применением табличного процессора Excel.

Таблица 8

№	Баллы по тесту для трех методик		
	Первоисточник	Учебник	Компьютер
1	28	39	41
2	33	52	49
3	42	53	56
4	47	54	62
5	48	56	63
6	50	58	64
7	50	59	65
8	51	63	72
9	60	64	77
10	71	77	87

Как в результате проведенного однофакторного дисперсионного анализа можно ответить на следующие вопросы:

1) Влияет ли методика изучения темы на результат? Есть ли значимые различия между тремя выборками по уровню усвоения материала?

2) Есть ли статистически значимая тенденция возрастания показателей в порядке «первоисточник» - «учебник» - «компьютер»?

Задача 4.

В течение шести лет в хозяйстве использовались четыре различных технологии по выращиванию сельскохозяйственной культуры. Необходимо установить при уровне значимости $\alpha = 0,05$ влияние различных технологий на урожайность культуры. Данные наблюдений представлены в таблице 9.

Задачу решить с применением табличного процессора Excel.

Таблица 9

Год	Технология (фактор А)			
	А1	А2	А3	А4
1	0,9	1	1,3	1,8
2	0,8	0,7	1,5	1,9
3	1	1,3	1,6	1,1
4	1,3	1	1,4	1,4
5	1,4	1,3	1,2	1,3
6	0,8	1,3	1,1	1,9

Задача 5.

Требуется оценить влияние квалификации наладчиков (фактор К) на рассеяние диаметров шариков, которые они изготавливают, при уровне значимости $\alpha = 0,05$. Замеры отклонения диаметра от номинала для каждого из пяти наладчиков проводились по 6 раз. Результаты представлены в таблице 10.

Задачу решить с применением табличного процессора Excel.

Таблица 10

№	Квалификация наладчика				
	К1	К2	К3	К4	К5
1	1,2	0,6	0,9	1,7	1,0
2	1,1	1,1	0,6	1,4	1,4
3	1,0	0,8	0,8	1,3	1,1
4	1,3	0,7	1,0	1,6	0,9
5	1,1	0,7	1,0	1,2	1,2
6	0,8	0,9	1,1	1,3	1,5

Проверяется нулевая гипотеза о равенстве математических ожиданий отклонения для всех пяти наладчиков, то есть предполагается, что квалификация наладчика не влияет на точность изготовления шариков.

Задача 6.

Требуется оценить влияние квалификации наладчиков (фактор К) на рассеяние диаметров шариков при уровне значимости $\alpha = 0,05$. Замеры отклонения диаметра от номинала для каждого из шести наладчиков проводились по 6 раз. Результаты представлены в таблице 11.

Задачу решить с применением табличного процессора Excel.

Таблица 11

№	Квалификация наладчика					
	К1	К2	К3	К4	К5	К6
1	1,2	0,6	0,9	1,7	1,0	1,4
2	1,1	1,1	0,6	1,4	1,4	1,3
3	1,0	0,8	0,8	1,3	1,1	1,1
4	1,3	0,7	1,0	1,6	0,9	1,6
5	1,1	0,7	1,0	1,2	1,2	1,3
6	0,8	0,9	1,1	1,3	1,5	1,5

Проверяется нулевая гипотеза о равенстве математических ожиданий отклонения для всех шести наладчиков, то есть предполагается, что квалификация наладчика не влияет на точность изготовления шариков.

Задача 7.

Компания, работающая на рынке, хочет оценить влияние обучения персонала на уровень продаж. Кадровая служба компании осуществляет обучение сотрудников отдела продаж в группах (групповое обучение) и организует индивидуальное обучение. Результаты продаж восьми сотрудников после группового и индивидуального обучения приведены в таблице 12. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли обучения персонала на уровень продаж.

Задачу решить с применением табличного процессора Excel. [4]

Таблица 12

Влияние обучения персонала на уровень продаж

Объем продаж за различные периоды			
Статистический период	Объем продаж до обучения (млн. руб.)	Объем продаж после группового обучения	Объем продаж после индивидуального обучения
206	203	217	213
226	223	230	231
208	206	221	221
224	205	227	222
206	234	218	229
229	204	231	235
204	219	224	213
228	210	225	228
209	233	211	214
221	223	229	225

Задача 8.

Компания, производящая спортивные товары, желает сравнить расстояние, которое пролетают мячи, изготовленные по 4 разным технологиям. По каждой технологии произведено по 10 мячей. Мячи переданы для испытания в спортивный клуб, где испытаны в течение короткого промежутка времени при одинаковых погодных условиях. Результаты испытаний в м приведены в таблице 13. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли технология производства на расстояние, которое пролетает мяч.

Задачу решить с применением табличного процессора Excel.

Таблица 13

Технология			
1	2	3	4
206	203	217	213
226	223	230	231
208	206	221	221
224	205	227	222
206	234	218	229
229	204	231	235
204	219	224	213
228	210	225	228
209	233	211	214
221	223	229	225

Задача 9.

Используя анализ однофакторной модели, проверить гипотезу о влиянии технологии чистовой обработки детали на точность ее изготовления при уровне значимости $\alpha = 0,05$.

Возможно применение трех видов технологии чистовой обработки детали. Исследуется влияние технологии чистовой обработки детали на точность ее изготовления. Проводятся замеры точности обработки детали, по 4 замера отклонения размера детали от номинала в мкм (для каждого вида технологии). Результаты представлены в таблице 14.

Задачу решить с применением табличного процессора Excel.

Таблица 14

Номер замера	Вид технологии		
	1	2	3
1	1	2	3
2	2	1	2
3	2	3	2
4	1	2	3

Задача 10.

Четыре группы студентов на уборке картофеля за четыре дня выполнили объем работы, приведенный в таблице 15. При уровне значимости $\alpha = 0,05$ требуется выяснить, зависит ли объем выполненных работ от работающей группы.

Задачу решить с применением табличного процессора Excel.

Дата	Количество убранных рядков картофеля за рабочий день			
	Группа 1	Группа 2	Группа 3	Группа 4
20 августа	140	150	148	150
21 августа	144	149	149	155
22 августа	142	152	146	154
23 августа	145	150	147	152

Задача 11.

Достоверны ли при уровне значимости $\alpha = 0,05$ различия приростов бычков в четырех экспериментальных группах по пять животных, данные о которых приведены в таблице 16?

Задачу решить с применением табличного процессора Excel.

Таблица 16

Приросты массы (в фунтах) в четырех группах животных

Группа	№ животного				
	1	2	3	4	5
1	40	24	46	20	35
2	29	27	20	39	45
3	11	31	17	37	39
4	17	21	38	33	21

Задача 12.

На одной из опытных станций испытывались шесть местных сортов пшеницы. Результаты учета урожайности приведены в таблице 17. Используя однофакторный дисперсионный анализ, выясните, достоверны ли различия в урожайности между сортами при уровне значимости $\alpha = 0,05$.

Задачу решить с применением табличного процессора Excel.

Таблица 17

Урожайность сортов пшеницы, ц/га

№ сорта	Повторение			
	1	2	3	4
1	26,1	29,2	30,0	27,3
2	25,0	24,3	28,5	29,0
3	27,2	26,4	31,0	26,4
4	23,6	27,2	25,2	24,8
5	30,0	33,0	36,0	29,8
6	23,0	26,0	26,0	24,8

Задача 13.

Результаты оценки влияния различных норм минеральных удобрений на урожайность яровой пшеницы в полевом опыте приведены в таблице 18. Используя однофакторный дисперсионный анализ, выясните достоверность и степень влияние минеральных удобрений на урожайность при уровне значимости $\alpha = 0,05$.

Задачу решить с применением табличного процессора Excel.

Таблица 18

Урожайность яровой пшеницы при разных нормах удобрений, ц/га

Норма удобрений, кг д. в./га	Повторение			
	1	2	3	4
0 (Без удобрений)	16,2	15,8	14,0	15,1
N ₂₀ P ₂₀	17,5	18,9	16,7	18,2
N ₄₀ P ₂₀	18,9	18,7	19,3	20,6
N ₆₀ P ₂₀	26,3	24,5	24,3	25,5
N ₈₀ P ₂₀	23,0	23,9	24,2	23,9

Задача 14.

Изучается, как влияет тип магазина на товарооборот. Данные о зафиксированном товарообороте (в млн руб.) в магазинах трех типов за каждые 8 месяцев работы приведены в таблице 19. Используя однофакторный дисперсионный анализ, выясните достоверность влияния тип магазина на товарооборот при уровне значимости $\alpha = 0,05$.

Задачу решить с применением табличного процессора Excel.

Таблица 19

Магазин 1	Магазин 2	Магазин 3
19	20	16
23	20	15
26	32	18
18	37	26
20	40	19
20	24	17
18	22	19
35	18	18

Задача 15.

Изучается, как влияет тип магазина на товарооборот. Данные о зафиксированном товарообороте (в млн руб.) в магазинах трех типов за каждые 8 месяцев работы приведены в таблице 20. Используя однофакторный дисперсионный анализ, нужно выяснить достоверность влияния тип магазина на товарооборот при уровне значимости $\alpha = 0,05$.

Задачу решить с применением табличного процессора Excel.

Таблица 20

Магазин 1	Магазин 2	Магазин 3
19	20	19
23	20	23
26	32	18
18	37	26
20	40	20
20	24	20
18	22	18
35	18	35

Задача 16.

В течение шести лет использовались четыре разных технологии (Т) по выращиванию сельскохозяйственной культуры. Необходимо установить влияние различных технологий на урожайность культуры при уровне значимости $\alpha = 0,05$ по данным таблицы 21.

Задачу решить с применением табличного процессора Excel.

Урожайность яровой пшеницы при разных нормах удобрений, ц/га

Год	Технология (фактор Т)			
	T1	T2	T3	T4
1	0,9	1,0	1,3	1,8
2	0,8	0,7	1,5	1,9
3	1,0	1,3	1,6	1,1
4	1,3	1,0	1,4	1,4
5	1,4	1,3	1,2	1,3
6	0,8	1,3	1,1	1,9

Задача 17.

Деталь может изготавливаться из трех видов материалов. Производитель проверяет влияние на износостойкость детали материала, из которого она изготовлена. Получены данные по износостойкости пяти деталей для каждого материала: время работы детали до износа, тыс. час. Результаты испытаний приведены в таблице 22. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы проверить влияние на износостойкость детали материала (три вида), из которого она изготовлена.

Задачу решить с применением табличного процессора Excel.

Таблица 22

Материал 1	1,25	1,32	1,28	1,26	1,29
Материал 2	1,12	1,15	1,26	1,19	1,21
Материал 3	1,32	1,33	1,34	1,29	1,30

Задача 18.

Менеджер по продажам в сети супермаркетов хочет знать, влияет ли расположение рекламных щитов на объем продаж товара. Для каждого из трех видов щитов отобрано случайным образом по 6 магазинов, расположенных в соответствующем районе. Объемы продаж за месяц (в млн руб.) приведены в таблице 23. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли расположение рекламных щитов на объем продаж товара.

Задачу решить с применением табличного процессора Excel.

Таблица 23

Расположение щитов		
1	2	3
8,1	3,9	4,8
7,5	4,4	6,0
5,8	2,8	4,4
6,6	3,4	5,8
5,9	4,8	6,2
4,8	3,6	4,9

Задача 19.

Крупной организации для принятия решения о закупках лицензионного программного обеспечения (ПО) необходимо выяснить значимость выбора

поставщика с точки зрения возможности сокращения при этом издержек. Требуется выяснить, насколько выбор поставщика комплекта лицензионного ПО влияет на издержки предприятия. В шести классах градации фактора П (1-й поставщик, 2-й поставщик и т.д.) собраны данные о потенциальных затратах на покупку 100 единиц продукции в тысячах рублей в течение 4 лет. Данные о затратах за год (в тыс. руб.) приведены в таблице 24. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,05$, чтобы выяснить, влияет ли выбор поставщика на размер потенциальных затрат. [5]

Задачу решить с применением табличного процессора Excel.

Таблица 24

Данные о потенциальных затратах, тыс. руб.

Поставщик	2018	2019	2020	2021
1-й	72,1	75,5	72,9	76,0
2-й	78,9	82,7	73,9	81,8
3-й	72,5	70,1	73,7	75,3
4-й	77,5	74,1	72,7	74,2
5-й	77,0	82,8	85,5	86,0
6-й	75,6	80,5	80,7	78,4

Задача 20.

Организация, которая занимается разработкой конструкционных сплавов, проводит круговые межлабораторные испытания на растяжение образцов типового полуфабриката конструкционного сплава. В качестве характеристики конструкционного материала выбран предел прочности при растяжении – одна из

основных характеристик конструкционного материала. Каждая лаборатория испытывала серию из 10 стандартных образцов. В круговых испытаниях принимали участие 5 различных лабораторий. Необходимо определить, являются ли расхождения в средних оценках предела прочности при растяжении, полученные в разных лабораториях, значимыми. Данные результатов испытаний на растяжение для участвовавших лабораторий приведены в таблице 25. Используйте однофакторный дисперсионный анализ, задав уровень значимости $\alpha = 0,01$, чтобы выяснить, влияет ли выбор лаборатории (фактор L) на полученное среднее значение предела прочности при растяжении. [6]

Задачу решить с применением табличного процессора Excel.

Таблица 25

Результаты испытаний на растяжение

Условный номер образца ($i = 1, \dots, 10$)	Значения предела прочности при растяжении, полученные j -м участником ($j = 1, \dots, 5$)				
	L ₁	L ₂	L ₃	L ₄	L ₅
1	465	460	470	465	445
2	470	460	470	465	420
3	470	460	465	465	395
4	465	460	470	465	395
5	470	460	455	465	440
6	455	445	455	450	410
7	460	445	455	455	395
8	460	445	455	450	435
9	455	445	455	455	445
10	460	445	455	455	395

Задача 21.

Требуется оценить эффективность нескольких различных способов действий, направленных на достижение одной цели – получение максимально дешевого кредита на покупку автомобиля. Для этого рассмотрим данные о годовой процентной ставке на кредиты для новых автомобилей в девяти самых крупных банках шести городов США. Данные представлены в таблице 26. Проверяется нулевая гипотеза о независимости в среднем ставки от города при уровне значимости $\alpha = 0,05$.

Задачу решить с применением табличного процессора Excel.

Таблица 26

Ставки по кредиту в городах США

№	Данные					
	Atlanta	Chicago	Houston	Memphis	New York	Philadelphia
1	13,75	14,25	14	15	14,5	13,5
2	13,75	13	14	14	14	12,25
3	13,5	12,75	13,51	13,75	14	12,25
4	13,5	12,5	13,5	13,59	13,9	12
5	13	12,5	13,5	13,25	13,75	12
6	13	12,4	13,25	12,97	13,25	12
7	13	12,3	13	12,5	13	12
8	12,75	11,9	12,5	12,25	12,5	11,9
9	12,5	11,9	12,5	11,89	12,45	11,9

Проверяется нулевая гипотеза о независимости в среднем ставки от города.

Задача 22.

Четыре бригады выполняли работы по посадке декоративных кустарников. Объем работы, выполненный каждой бригадой за каждый из четырех рабочих дней, приведен в таблице 27. При уровне значимости $\alpha = 0,05$ требуется выяснить, зависит ли объем выполненных работ от работающей бригады?

Задачу решить с применением табличного процессора Excel.

Таблица 27

День	Бригада (фактор)			
	Бригада 1	Бригада 2	Бригада 3	Бригада 4
1-й	140	150	148	150
2-й	144	149	149	155
3-й	142	152	146	154
4-й	145	150	147	152

Список литературы

1. Валеев С. Г. Практикум по прикладной статистике: учебное пособие / С.Г. Валеев, В.Н. Клячкин. – Ульяновск: УлГТУ, 2008. – 129 с.: ил.
2. УЧЕБНОЕ ПОСОБИЕ ПО ДИСЦИПЛИНЕ "СТАТИСТИКА. СТАТИСТИЧЕСКИЙ АНАЛИЗ: РЕАЛИЗАЦИЯ В ПРОГРАММЕ R". Атчаде М.Н., Елисеева И.И. Санкт-Петербург, 2017.
3. Фрумин И.Л. Статистические методы и модели в агроэкологии. Дисперсионный анализ. Методические указания к выполнению лабораторных работ по дисциплине "Компьютерное моделирование экосистем" - Челябинск, 2007. Издательство: Челябинская государственная агроинженерная академия. – 31 с.
4. Слабнин А.Д. ОЦЕНКА ВЛИЯНИЯ ОБУЧЕНИЯ ПЕРСОНАЛА НА УРОВЕНЬ ПРОДАЖ МЕТОДОМ ДИСПЕРСИОННОГО АНАЛИЗА // В сборнике: Современное состояние, проблемы и перспективы развития отраслевой науки. Материалы Всероссийской конференции с международным участием. 2017. С. 209-213.
5. Емцев А.С., Алексеев В.В. ИСПОЛЬЗОВАНИЕ МЕТОДА ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА В СППР О ЗАКУПКАХ ЛИЦЕНЗИОННОГО ПО ДЛЯ КРУПНОЙ ОРГАНИЗАЦИИ. // В сборнике: Проблемы информатики в образовании, управлении, экономике и технике. Сборник статей XX Международной научно-технической конференции, посвященной 75-летию Победы в Великой Отечественной войне. 2020. С. 14-19.
6. Подживотов Н.Ю. ОЦЕНКА РЕЗУЛЬТАТОВ ИСПЫТАНИЙ С ПОМОЩЬЮ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА. // Труды ВИАМ. 2022. № 8 (114). С. 141-152.
7. Петрашин И.В., Сувориков Е.Е. ИСПОЛЬЗОВАНИЕ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ПРОВЕРКИ ГИПОТЕЗ // В сборнике: Фундаментальные и прикладные научные исследования: актуальные вопросы, достижения и инновации. сборник статей XVIII Международной научно-

практической конференции. В 2 частях. Ответственный редактор Г.Ю. Гуляев. 2018. С. 17-19.

8. Шарстнёв В.Л., Вардомацкая Е.Ю. ПРОГНОЗИРОВАНИЕ УРОВНЯ ПРОДАЖ НА ОСНОВЕ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА // В сборнике: Моделирование в технике и экономике. сборник материалов международной научно-практической конференции. Главный редактор: Ванкевич Е.В., 2016. С. 450-453.

9. Александров В.Д. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ СВОЙСТВ СОЛЕВЫХ РАСТВОРОВ - ХИМИЧЕСКИХ АНТИГОЛОЛЕДНЫХ РЕАГЕНТОВ / Александров В.Д., Александрова О.В., Соболев О.В., Соболев А.Ю. // Вестник Донбасской национальной академии строительства и архитектуры. 2020. № 1 (141). С. 115-120.

10. Сенкевич Д.В. ПРИМЕНЕНИЕ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ ВЛИЯНИЯ ПОГОДНЫХ УСЛОВИЙ НА РАЗВИТИЕ САР / В сборнике: Материалы конференции студентов и молодых ученых, посвященной памяти профессора Д. А. Маслакова. Гродно, 2012. С. 369-370.

11. Володько О.С., Дергунов А.В. ДИСПЕРСИОННЫЙ АНАЛИЗ ЗАВИСИМОСТИ КОНЦЕНТРАЦИИ ВЗВЕШЕННЫХ ЧАСТИЦ PM2.5 В АТМОСФЕРЕ Г. КРАСНОЯРСКА И МЕТЕОУСЛОВИЙ / В сборнике: РЕГИОНАЛЬНЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ. Материалы IX Международной научной конференции. Научный редактор Е.А. Ваганов, отв. редактор Г.М. Цибульский. Красноярск, 2022. С. 195-198.

12. Нешумов Е.В. МЕТОДЫ ДИСПЕРСИОННОГО АНАЛИЗА В ЭКОНОМИКЕ / В сборнике: Инновационный потенциал развития науки в современном мире. Сборник трудов по материалам IX Всероссийского конкурса научно-исследовательских работ. Уфа, 2022. С. 98-104.

13. Митяков Е.С., Гнатченко Д.И. О ВОЗМОЖНОСТИ И АКТУАЛЬНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ОЦЕНКИ ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ РЕГИОНА / В сборнике:

Актуальные вопросы экономики, менеджмента и инноваций. Материалы Международной научно-практической конференции ученых, специалистов, преподавателей вузов, аспирантов, студентов. 2017. С. 214-217.

14. Чеканова Е.В. РАЗВИТИЕ ЦИРКУЛЯРНОЙ ЭКОНОМИКИ НА ОСНОВЕ ИННОВАЦИЙ: СРАВНИТЕЛЬНЫЙ РЕГИОНАЛЬНЫЙ АСПЕКТ // Вестник Белгородского университета кооперации, экономики и права. 2021. № 4 (89). С. 162-171.

1.1. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$

$df_1 \backslash df_2$	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48

1	2	3	4	5	6	7	8	9	10	11
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

Оглавление

Введение.....	3
1. Правило сложения дисперсий.....	4
2. Дисперсионный анализ.....	5
3. Пример задачи однофакторного дисперсионного анализа	14
4. Решение задачи однофакторного дисперсионного анализа с использованием стандартного табличного процессора Microsoft Excel	19
Контрольные вопросы	27
Задания для самостоятельной работы.....	28
Список литературы	47
Приложение 1	50

Учебное издание

А.А. Чалганова

Выполнение задачи однофакторного дисперсионного анализа с использованием табличного процессора Excel

Учебное пособие по дисциплине «Статистика»

Публикуется в авторской редакции.

Подписано к публикации 30.12.2022.

Формат 60×90 1/16. Гарнитура Times New Roman.

Усл. печ. л. 3,25. Заказ № 1346.

РГГМУ, 192007, Санкт-Петербург, Воронежская ул., д. 79.