

ЛАБОРАТОРНАЯ РАБОТА 3.

КРИТЕРИИ ЗНАЧИМОСТИ

ЗАДАНИЕ. Решить задачи, используя критерии значимости, предназначенные для проверки гипотез о значениях параметров нормального распределения. Уровень значимости принять $\alpha = 0,05$.

Во всех задачах считать, что исследуемые признаки имеют нормальное распределение.

Контрольные вопросы

1. В чем заключается основная задача при проверке статистической гипотезы?
2. Что называется статистической гипотезой?
3. В каком случае статистическая гипотеза называется простой? сложной?
4. В чем разница между нулевой и альтернативной гипотезами?
5. В каком случае статистическая гипотеза называется параметрической? непараметрической?
6. Что называется критерием значимости? Что называется критерием согласия?
7. Что называется уровнем значимости статистического критерия?
8. Как видоизменяется критерий проверки гипотезы в случае односторонней альтернативы?
9. Что характеризует выборочное среднее? Что характеризует выборочная дисперсия?
10. Как рассчитать выборочное среднее?
11. Как рассчитать несмещенную оценку дисперсии?
12. Какие критерии используются для проверки гипотез о математических ожиданиях одной и двух независимых нормальных выборок?
13. Какие критерии используются для проверки гипотез о дисперсиях одной и двух независимых нормальных выборок?
14. Что такое однородность дисперсий и как она проверяется?
15. Для проверки каких гипотез используется критерий Фишера?
16. Как используется критерий Фишера для проверки однородности нескольких дисперсий?
17. Для проверки каких гипотез используется критерий χ^2 ?
18. Чем отличается процедура проверки гипотезы о равенстве математического ожидания заданному значению для случаев известной и неизвестной дисперсии?
19. Для проверки каких гипотез используется критерий Стьюдента?
20. Чем отличается процедура проверки гипотезы о равенстве средних двух зависимых и независимых нормальных выборок?
21. В чем заключается процедура проверки гипотезы о равенстве средних в случае парных (зависимых) выборок?
22. Как учитывается предположение о равенстве дисперсий при сравнении средних?

Пример и методические указания по выполнению лабораторной работы в Excel

Задача 1. Проводилось исследование предела прочности на разрыв различных по химической структуре твердых смол. При этом требовалось выяснить, вносят ли действия оператора какое-нибудь смещение в результаты наблюдений. Было взято по два образца каждой смолы. Двум операторам А и В предложили испытать по одному образцу смол каждого типа. Можно ли говорить, что наблюдаются различия между результатами наблюдений двух операторов?

Смола	127	135	138	139	146	152
Оператор А	5250	4975	5050	5075	4795	5190
Оператор В	5230	4980	5020	5085	4750	5120

Решение. Исследуемый признак (СВ) – предел прочности на разрыв смол.

Поскольку два оператора испытывали одни и те же образцы смол, имеем задачу сравнения средних в случае зависимых выборок.

Проверяем гипотезу H_0 о том, что в среднем результаты наблюдений одинаковы

$$H_0: a_{\Delta x} = 0$$

при альтернативе \overline{H} , согласно которой есть различия в результатах наблюдений:

$$\overline{H}: a_{\Delta x} \neq 0.$$

Дисперсия σ^2 **неизвестна** и будет оцениваться по выборке. Критерием проверки гипотезы является статистика

$$t_{\text{расч}} = \frac{|\overline{\Delta x}|}{\sqrt{s^2 / n}} < t_{\text{табл}} = t_{\alpha; n-1}.$$

Рассчитаем выборку значений $\Delta x_i = x_{1i} - x_{2i}$ разностей результатов наблюдений двух операторов.

Смола	127	135	138	139	146	152
$\Delta x_i = x_{1i} - x_{2i}$	20	-5	30	-10	45	70

Объем выборки $n = 6$. Рассчитаем несмещенные оценки среднего и дисперсии:

$$\overline{\Delta x} = \frac{1}{6} \cdot (20 - 5 + 30 - 10 + 45 + 70) = 25;$$

$$D_{\text{в}} = \frac{1}{6} \cdot ((20 - 25)^2 + (-5 - 25)^2 + (30 - 25)^2 + (-10 - 25)^2 + (45 - 25)^2 + (70 - 25)^2) = \\ = \frac{1}{6} \cdot (25 + 900 + 25 + 1225 + 400 + 2025) = \frac{1}{6} \cdot 4600 = \frac{2300}{3};$$

$$s^2 = \frac{6}{5} \cdot \frac{4600}{6} = 920.$$

Поскольку $t_{\text{расч}} = \frac{25}{\sqrt{920 / 6}} \approx 2,02 < t_{\text{табл}} = t_{0,05; 5} = 2,57$, то делаем вывод: на уровне

значимости 0,05 можно утверждать, что гипотеза H_0 не противоречит экспериментальным данным, т. е. **различия между результатами измерений у двух операторов следует признать незначительными**.

Ниже приведен фрагмент рабочего листа по выполнению расчетов для решения задачи в Excel с помощью встроенных статистических функций и с помощью Пакета анализа.

Пакет анализа – набор средств анализа данных в Microsoft Excel, предназначенный для решения сложных статистических и инженерных задач. Средства, которые включены в Пакет анализа данных, доступны через команду меню **Данные**→**Анализ данных**.

Если этой команды нет в меню, необходимо загрузить надстройку **Пакет анализа**: 1) открыть вкладку **Файл**→**Параметры**→**Надстройки**; 2) в раскрывающемся списке **Управление** выбрать пункт **Надстройки Excel** и нажать кнопку **Перейти**; 3) в диалоговом окне **Надстройки** установить флажок **Пакет анализа** и нажать кнопку **ОК**.

Лаб3_пример Анализ данных.xls [Режим совместимости] - Excel

Файл Главная Вставка Рисование Разметка страницы Формулы Данные Рецензирование Вид Разработчик Справка Foxit PDF Что вы хотите сделать?

Получить данные Обновить все Запросы и подключения Свойства Изменить связи Запросы и подключения Сортировка и фильтр Фильтр Очистить Повторить Дополнительно Анализ "что если" Лист прогноза Структура Анализ

Получить и преобразовать данные

А9 Парный двухвыборочный t-тест для средних

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Задача 1	alpha=	0,05													
2	Смола	127	135	138	139	146	152									
3	Оператор A	5250	4975	5050	5075	4795	5190									
4	Оператор B	5230	4980	5020	5085	4750	5120									
5	delta x _i	20	-5	30	-10	45	70									
6	n=	6	x-mean=	25	s2=	920	f=	5								
7	t-rasch=	2,0189321	t-tabl=	2,570582												
8	Парный двухвыборочный t-тест для средних															
9	Оператор A Оператор B															
10	Среднее	5055,8333	5030,8333													
11	Дисперсия	26134,167	26424,167													
12	Наблюдения	6	6													
13	Корреляция Пирсона	0,9825106														
14	Гипотетическая разность	0														
15	df	5														
16	t-статистика	2,0189321														
17	P(T<=t) одностороннее	0,0497529														
18	t критическое одностороннее	2,0150484														
19	P(T<=t) двухстороннее	0,0995057														
20	t критическое двухстороннее	2,5705818														

Парный двухвыборочный t-тест для средних

Входные данные

Интервал переменной 1: \$A\$3:\$G\$3

Интервал переменной 2: \$A\$4:\$G\$4

Гипотетическая средняя разность:

☒ Метки

Альфа: 0,05

Параметры вывода

☒ Выходной интервал: \$A\$9

☐ Новый рабочий дист.

☐ Новая рабочая книга

OK Отмена Справка

Использование Пакета анализа

Вызовите Данные→Анализ данных→Парный двухвыборочный t-тест для средних.

Сравните результаты:

$t\text{-статистика} = t_{\text{расч}}$;

t критическое двухстороннее = $t_{\text{табл}}$ (для случая односторонней альтернативы имеется t критическое одностороннее);

$P(T \leq t)$ двухстороннее – это пороговое значение уровня значимости, показывающее, что при $\alpha > 0,0995$ гипотеза H_0 при альтернативе \bar{H} отвергается.

Задача 2. Для того чтобы определить, сокращается ли время сварки на отливках, если при литье вместо сырой формовочной смеси использовать сухую смесь, было проведено специальное исследование. Стоимость литья в случае сухой формовочной смеси выше, но есть мнение, что это может быть оправдано, если время сварки значимо уменьшится. В таблице приведены значения времени сварки в минутах. Можно ли сказать, что имеет место значимое уменьшение времени сварки при использовании сухой формовочной смеси?

	1	2	3	4	5
Сырая смесь	19	28	14	29	15
Сухая смесь	21	15	11	12	21

Решение. Исследуемый признак (СВ) – время сварки на отливках.

Имеем задачу сравнения средних в случае независимых выборок.

Проверяем при уровне значимости $\alpha = 0,05$ нулевую гипотезу H_0 о том, что в среднем время сварки одинаковое для двух технологий:

$$H_0: a_1 = a_2$$

при альтернативе \bar{H} , согласно которой для второй технологии время в среднем сокращается (односторонняя альтернатива):

$$\bar{H}: a_1 > a_2.$$

При сравнении двух средних нужно учитывать, **однородны ли дисперсии** двух выборок, поэтому предварительно проверяется **вспомогательная гипотеза** о равенстве дисперсий:

$$H'_0: \sigma_1^2 = \sigma_2^2;$$

$$\bar{H}': \sigma_1^2 \neq \sigma_2^2.$$

Проверка проводится по критерию Фишера. Гипотеза H'_0 при альтернативе \bar{H}' на уровне значимости α принимается, если

$$F_{\text{расч}} = \frac{s_{\text{max}}^2}{s_{\text{min}}^2} < F_{\text{табл}} = F_{\alpha/2; f_1; f_2},$$

где f_1 и f_2 – числа степеней свободы большей и меньшей оценок дисперсий соответственно.

Рассчитаем статистические характеристики выборок.

	n_i	\bar{x}_i	s_i^2	f_i
Сырая смесь	5	21	50,5	4
Сухая смесь	5	16	23	4

Объемы выборок $n_1 = n_2 = 5$. Число степеней свободы каждой оценки дисперсии равно числу наблюдений, по которым она рассчитана, минус 1:

$$f_1 = n_1 - 1 = 5 - 1 = 4; \quad f_2 = n_2 - 1 = 5 - 1 = 4.$$

Расчетное значение критерия Фишера (нужно разделить большую оценку дисперсии на меньшую) равно

$$F_{\text{расч}} = \frac{50,5}{23} \approx 2,2;$$

табличное значение $F_{\text{табл}} = F_{0,025; 4; 4} = 9,6$. Поскольку $F_{\text{расч}} = 2,2 < F_{\text{табл}} = 9,6$, то делаем вывод: **на уровне значимости 0,05 можно считать дисперсии однородными.**

Если дисперсии однородны, то гипотеза H_0 при *односторонней* альтернативе \bar{H} на уровне значимости α принимается, если $\bar{x}_1 < \bar{x}_2$ (среднее время для первой технологии оказалось меньше среднего времени для второй) либо

$$t_{\text{расч}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < t_{\text{табл}} = t_{2\alpha; f},$$

$$\text{где } s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad f = n_1 + n_2 - 2.$$

Вычисляем общую средневзвешенную оценку дисперсии и ее число степеней свободы:

$$s^2 = \frac{(5-1) \cdot 50,5 + (5-1) \cdot 23}{5+5-2} = 36,75; \quad f = 5+5-2 = 8;$$

сравниваем расчетное и критическое значения критерия Стьюдента:

$$t_{\text{расч}} = \frac{21-16}{\sqrt{36,75 \cdot \left(\frac{1}{5} + \frac{1}{5}\right)}} \approx 1,3 < t_{\text{табл}} = t_{0,1;8} = 1,86,$$

делаем **вывод**: на уровне значимости 0,05 можно утверждать, что гипотеза H_0 не противоречит экспериментальным данным, т. е. **нет оснований утверждать, что при использовании сухой смеси время литья значимо уменьшится.**

Фрагмент рабочего листа в Excel.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Задача 2	alpha=	0,05											
2		1	2	3	4	5	n	x-mean	s2	f				
3	Сырая смесь	19	28	14	29	15	5	21	50,5	4				
4	Сухая смесь	21	15	11	12	21	5	16	23	4				
5	F-rasch=	2,195652	F-tabl=	9,60453										
6	t-rasch=	1,304101	t-tabl=	1,859548					36,75	8				
8	Двухвыборочный F-тест для дисперсии				Двухвыборочный t-тест с одинаковыми дисперсиями									
10	Сырая смесь				Сухая смесь									
11	Среднее	21	16	Среднее	21	16								
12	Дисперсия	50,5	23	Дисперсия	50,5	23								
13	Наблюдения	5	5	Наблюдения	5	5								
14	df	4	4	Объединенные	36,75									
15	F	2,195652		Гипотетическая	0									
16	P(F<=f) односторонне	0,232482		df	8									
17	F критическое	9,60453		t-статистика	1,304101									
18				P(T<=t) односторонне	0,114234									
19				t критическое	1,859548									
20				P(T<=t) двусторонне	0,228468									
21				t критическое	2,306004									

The dialog box "Двухвыборочный t-тест с одинаковыми дисперсиями" is open, showing the following settings:

- Входные данные: Интервал переменной 1: \$A\$3:\$F\$3, Интервал переменной 2: \$A\$4:\$F\$4
- Гипотетическая средняя разность: (empty)
- Метки: ☒ Метки
- Альфа: 0,05
- Параметры вывода: ☒ Выходной интервал: \$E\$8
- ☐ Новый рабочий лист:
- ☐ Новая рабочая книга

Задача 3. При производстве синтетического волокна для уменьшения последующей усадки продукция, движущаяся непрерывным потоком, подвергается термической обработке. Даны результаты измерений величины усадки в процентах для волокон после обработки при двух температурах: 120°С и 140°С. До начала эксперимента предполагалось, что дисперсии усадки при рассмотренных температурах не равны между собой. Требуется проверить, будет ли усадка при 140°С больше, чем при 120°С.

120°С	3,45	3,62	3,6	3,49	3,64	3,56	3,52	3,53	3,57	3,44	3,56	3,43
140°С	3,72	4,01	3,54	3,67	4,03	3,4	3,96	3,6	3,76	3,91		

Решение. Исследуемый признак (СВ) – величина усадки волокон.

Требуется проверить, будут ли значения усадки во второй выборке в среднем больше, чем в первой, или усадка в среднем одинакова в обеих выборках.

Имеем **задачу сравнения средних в случае независимых выборок**.

Проверяем при уровне значимости $\alpha = 0,05$ нулевую гипотезу H_0 о том, что в среднем усадка одинакова в обеих выборках:

$$H_0: a_1 = a_2$$

при альтернативе \bar{H} , согласно которой усадка при 140°C больше, чем при 120°C (односторонняя альтернатива):

$$\bar{H}: a_1 < a_2.$$

При сравнении двух средних нужно учитывать, **однородны ли дисперсии** двух выборок. Предварительно проверяется вспомогательная гипотеза о равенстве дисперсий:

$$H'_0: \sigma_1^2 = \sigma_2^2;$$

$$\bar{H}': \sigma_1^2 \neq \sigma_2^2.$$

Проверка проводится по критерию Фишера. Гипотеза H'_0 при альтернативе \bar{H}' на уровне значимости α принимается, если

$$F_{\text{расч}} = \frac{s_{\text{max}}^2}{s_{\text{min}}^2} < F_{\text{табл}} = F_{\alpha/2; f_1; f_2},$$

где f_1 и f_2 – числа степеней свободы большей и меньшей оценок дисперсий соответственно.

Рассчитаем статистические характеристики выборок.

	n_i	\bar{x}_i	s_i^2	f_i
120°C	12	3,53	0,005	11
140°C	10	3,76	0,046	9

Объемы выборок $n_1 = 12$; $n_2 = 10$. Число степеней свободы каждой оценки дисперсии равно числу наблюдений, по которым она рассчитана, минус 1:

$$f_1 = n_1 - 1 = 12 - 1 = 11; \quad f_2 = n_2 - 1 = 10 - 1 = 9.$$

Расчетное значение критерия Фишера (нужно разделить большую оценку дисперсии на меньшую) равно

$$F_{\text{расч}} = \frac{0,046}{0,005} \approx 9,2;$$

табличное значение $F_{\text{табл}} = F_{0,025; 9; 11} = 3,59$. Поскольку $F_{\text{расч}} = 9,2 > F_{\text{табл}} = 3,59$, то делаем вывод: **на уровне значимости 0,05 дисперсии следует признать неоднородными**.

Если **дисперсии неоднородны**, то гипотеза H_0 при **односторонней альтернативе \bar{H}** на уровне значимости α принимается, если $\bar{x}_1 > \bar{x}_2$ (среднее первой выборки оказалось больше второго среднего) либо

где

Сравнивая расчетное и критическое значения критерия Стьюдента:

делаем **вывод**: на уровне значимости 0,05 гипотеза H_0 отвергается, т. е. **усадка при 140° C больше, чем при 120° C**.

Фрагмент рабочего листа Excel.

Задача 4. Для проверки точности двух станков проведены измерения некоторого признака выпускаемых ими однотипных деталей. Можно ли на основании этих данных при уровне значимости $\alpha = 0,05$ сделать вывод о том, что точность первого станка выше точности второго?

Станок	\bar{x}_i	s_i^2	n_i
1-й	120,3	12,25	25
2-й	118,9	28,4	18

Решение. Исследуемый признак (СВ) – точность станка характеризуется разбросом значений признака выпускаемых изделий, произведенных станком: чем меньше разброс значений (т. е. дисперсия значений измеренного признака выпускаемых изделий), тем выше точность станка.

Имеем задачу сравнения двух дисперсий: проверяем при уровне значимости $\alpha = 0,05$ нулевую гипотезу H_0 о том, что разброс значений измеренного признака выпускаемых изделий одинаков для двух станков

$$H_0: \sigma_1^2 = \sigma_2^2$$

при альтернативе \bar{H} , согласно которой для первого станка разброс значений меньше, чем для второго (односторонняя альтернатива):

$$\bar{H}: \sigma_1^2 < \sigma_2^2.$$

Проверка проводится по критерию Фишера. Гипотеза H_0 при односторонней альтернативе \bar{H} на уровне значимости α принимается, если $s_1^2 > s_2^2$ (оценка дисперсии для первого станка оказалась меньше второй оценки дисперсии) либо

$$F_{\text{расч}} = \frac{s_{\text{max}}^2}{s_{\text{min}}^2} < F_{\text{табл}} = F_{2 \cdot \alpha / 2; f_1; f_2},$$

где f_1 и f_2 – числа степеней свободы большей и меньшей оценок дисперсий соответственно.

Объемы выборок $n_1 = 25; n_2 = 18$. Число степеней свободы каждой оценки дисперсии равно числу наблюдений, по которым она рассчитана, минус 1:

$$f_1 = n_1 - 1 = 25 - 1 = 24; \quad f_2 = n_2 - 1 = 18 - 1 = 17.$$

Расчетное значение критерия Фишера (нужно разделить большую оценку дисперсии на меньшую) равно

$$F_{\text{расч}} = \frac{28,4}{12,25} \approx 2,32;$$

табличное значение $F_{\text{табл}} = F_{0,05; 17; 24} = 2,07$ (уровень значимости делится на 2 в соответствии с формулой критерия Фишера, но удваивается в случае односторонней альтернативы; числа степеней свободы берутся в порядке, соответствующем порядку оценок дисперсий – сначала число степеней большей оценки дисперсии, затем меньшей).

Поскольку $F_{\text{расч}} = 2,32 > F_{\text{табл}} = 2,07$, то на уровне значимости 0,05 гипотеза H_0 о равенстве дисперсий должна быть отвергнута. Таким образом, по результатам экспериментальных данных на уровне значимости 0,05 точность первого станка признаем выше точности второго.

Задача 5. Измерялось сопротивление проволок трех типов. Утверждается, что между проволоками разных типов в среднем нет различий. Можно ли принять гипотезу об одинаковом среднем значении сопротивления для проволок трех типов?

A	121	120	124	121	120	124	126	120
B	120	119	126	128	126	124	122	127
C	129	132	137	139	130	132	137	136

Решение.

1 способ. Имеем задачу сравнения нескольких средних в случае независимых выборок. Решим задачу методом **однофакторного дисперсионного анализа**.

Проверяется гипотеза о равенстве средних трех выборок:

$$H_0 : a_1 = a_2 = a_3$$

при альтернативе \bar{H} : не все средние равны между собой.

Процедуре сравнения средних предшествует проверка однородности дисперсий:

$$H'_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2;$$

\bar{H}' : не все дисперсии равны между собой.

Проверку однородности дисперсий можно провести по критерию Фишера, сравнив наибольшую и наименьшую из оценок дисперсий.

Рассчитаем статистические характеристики выборок.

	n_i	\bar{x}_i	s_i^2	f_i
A	8	122	5,43	7
B	8	124	11,14	7
C	8	134	13,71	7

Сравнивая расчетное значение критерия Фишера (нужно разделить большую оценку дисперсии на меньшую) с табличным

$$F_{\text{расч}} = \frac{13,71}{5,43} \approx 2,53 < F_{\text{табл}} = F_{0,025; 7; 7} = 4,99,$$

делаем вывод, что на уровне значимости 0,05 дисперсии можно считать однородными. Это позволяет использовать дисперсионный анализ для проверки однородности средних.

Имеем $N = 3$ выборки объемами $n_1 = n_2 = n_3 = 8$. Объединяя все выборки в одну, вычисляем среднее объединенной выборки

$$\bar{\bar{x}} = \frac{1}{24} \cdot (121 + 120 + 124 + \dots + 137 + 136) \approx 126,67.$$

Рассчитаем межгрупповую дисперсию по формуле

$$s_{\text{факт}}^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - \bar{\bar{x}})^2 n_i =$$

$$= \frac{1}{3} \cdot ((122 - 126,67)^2 \cdot 8 + (124 - 126,67)^2 \cdot 8 + (134 - 126,67)^2 \cdot 8) = 330,67,$$

ее число степеней свободы равно $f_{\text{факт}} = N - 1 = 2$. Остаточная дисперсия представляет собой взвешенное среднее оценок дисперсий и рассчитывается по формуле

$$s_{\text{ост}}^2 = \frac{f_1 s_1^2 + f_2 s_2^2 + f_3 s_3^2}{f_1 + f_2 + f_3} = \frac{7 \cdot 5,43 + 7 \cdot 11,14 + 7 \cdot 13,71}{7 + 7 + 7} = 10,095,$$

$$f_{\text{ост}} = f_1 + f_2 + f_3 = 21.$$

Гипотеза H_0 при заданном уровне значимости α принимается (не противоречит экспериментальным данным), если

$$F_{\text{расч}} = \frac{s_{\text{факт}}^2}{s_{\text{ост}}^2} < F_{\text{табл}} = F_{\alpha; f_{\text{факт}}; f_{\text{ост}}},$$

где $F_{\alpha; f_{\text{факт}}; f_{\text{ост}}}$ определяется по таблице квантилей распределения Фишера.

Поскольку $F_{\text{расч}} = \frac{330,67}{10,095} \approx 32,75 > F_{\text{табл}} = F_{0,05; 2; 21} = 3,47$, то делаем вывод: на уровне значимости 0,05 гипотеза о равенстве трех средних отклоняется.

На фрагменте рабочего листа Excel представлены расчеты по решению задачи методом однофакторного дисперсионного анализа с помощью встроенных статистических функций и с помощью Пакета анализа.

The screenshot shows an Excel spreadsheet titled 'Лаб3_пример Анализ данных.xls'. The 'Данные' (Data) ribbon is active. The spreadsheet contains data for a one-way ANOVA analysis. The data is organized into columns A through N and rows 1 through 23. The data includes group names (A, B, C), sample sizes (n), means (x-mean), variances (s2), and F-statistics (f). A dialog box for 'Однофакторный дисперсионный анализ' (One-way ANOVA) is open, showing input ranges and options for grouping and output.

Группы	Счет	Сумма	Среднее	Дисперсия
A	8	976	122	5,428571
B	8	992	124	11,14286
C	8	1072	134	13,71429

Источники вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	661,3333	2	330,6667	32,75472	3,50055E-07	3,466800112
Внутри групп	212	21	10,09524			
Итого	873,3333	23				

The dialog box 'Однофакторный дисперсионный анализ' (One-way ANOVA) is open. It shows the following settings:

- Входные данные: \$A\$2:\$I\$4
- Группирование: ☒ по строкам
- Метки в первом столбце: ☒
- Альфа: 0,05
- Параметры вывода: ☒ Выходной интервал: \$A\$8

Е Введите исходные данные (всю таблицу) на Лист 5 в ячейки A2:I4; введите значение α , равное 0,05, – в ячейку C1.

Х *Использование статистических функций*
Проведите необходимые расчеты с помощью встроенных в Excel статистических

С Использование Пакета анализа

Е Вызовите Данные→Анализ данных→Однофакторный дисперсионный анализ. В диалоговом окне задаем:

Л Входные данные
Входной интервал \$A\$2:\$I\$4
Группирование выбираем по строкам
Поставьте галочку Метки в первом столбце
Альфа 0,05
Параметры вывода
Выходной интервал: \$A\$8
ОК

Е Сравните результаты.

Х Предварительные итоги по выборкам: объем выборки (Счет), сумма элементов в выборке (Сумма), среднее, дисперсия.

С Результаты дисперсионного анализа:

Е $df = f$ (числа степеней свободы факторной и остаточной дисперсий);

MS – факторная и остаточная дисперсии;

SS – суммы ($MS=SS/df$);

$F = F_{\text{расч}}$;

Л F критическое = $F_{\text{табл}}$;

P -Значение – это пороговое значение уровня значимости, показывающее, что при $\alpha > 3,5 \cdot 10^{-7}$ гипотеза H_0 при альтернативе \bar{H} отвергается.

2 способ. Имеем задачу сравнения нескольких средних в случае независимых выборок: проверяется гипотеза о равенстве средних трех выборок:

$$H_0: a_1 = a_2 = a_3$$

при альтернативе \bar{H} : не все средние равны между собой.

Решим задачу с помощью критерия Стьюдента, проверив гипотезу о равенстве наибольшего и наименьшего средних.

Рассчитаем статистические характеристики выборок.

	n_i	\bar{x}_i	s_i^2	f_i
A	8	122	5,43	7
B	8	124	11,14	7
C	8	134	13,71	7

Итак, проверим при уровне значимости $\alpha = 0,05$ нулевую гипотезу H'_0 о равенстве наибольшего и наименьшего средних (средних выборок А и С):

$$H'_0: a_1 = a_3$$

$$\bar{H}': a_1 \neq a_3.$$

При сравнении двух средних нужно учитывать, однородны ли дисперсии двух выборок, поэтому предварительно проверяется вспомогательная гипотеза о равенстве дисперсий:

$$H_0^{//} : \sigma_1^2 = \sigma_3^2;$$

$$\bar{H}^{//} : \sigma_1^2 \neq \sigma_3^2.$$

Проверка проводится по критерию Фишера: так как

$$F_{\text{расч}} = \frac{13,71}{5,43} \approx 2,53 < F_{\text{табл}} = F_{0,025; 7; 7} = 4,99,$$

то на уровне значимости 0,05 дисперсии признаем однородными.

Если дисперсии однородны, то гипотеза H_0' при альтернативе \bar{H}' на уровне значимости α принимается, если

$$t_{\text{расч}} = \frac{|\bar{x}_1 - \bar{x}_3|}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_3} \right)}} < t_{\text{табл}} = t_{\alpha; f},$$

где

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_3 - 1)s_3^2}{n_1 + n_3 - 2} = \frac{7 \cdot 5,43 + 7 \cdot 13,71}{8 + 8 - 2} \approx 9,57; \quad f = n_1 + n_3 - 2 = 8 + 8 - 2 = 14.$$

Сравнивая расчетное и критическое значения критерия Стьюдента:

$$t_{\text{расч}} = \frac{|122 - 134|}{\sqrt{9,57 \cdot \left(\frac{1}{8} + \frac{1}{8} \right)}} \approx 7,76 > t_{\text{табл}} = t_{0,05; 14} = 2,14,$$

заключаем: на уровне значимости 0,05 можно утверждать, что гипотеза H_0' о равенстве наименьшего и наибольшего средних отвергается. Следовательно, гипотеза о равенстве трех средних также отвергается.