Проблема множественных сравнений

Чем больше статистических гипотез проверяется на одних и тех же данных, тем вероятнее ошибка первого рода – заключение о наличии различий между группами, тогда как на самом деле верна нулевая гипотеза об отсутствии различий

Пример. Исследуют влияние препаратов А и Б на уровень глюкозы плазмы. Исследования проводят на трех группах: получавших препарат А, получавших препарат Б и получавших плацебо В. С помощью критерия Стьюдента проводят три парных сравнения А и В, Б и В, А и Б.

Получив достаточно высокое значение t ***хотя бы в одном из сравнений***, делают вывод о статистической значимости различий ($\alpha<0{,}05$).

Но ошибка в 5% возможна ***в каждом из трех сравнений***, т.е. вероятность ошибки в целом будет превышать 5%.

Вероятность ошибиться хотя бы в одном из сравнений:

- $p=1-(1-0{,}05)^k$,

где k – число парных сравнений

- ~ $p=0{,}05 \cdot k$

- $k=3$; $p=0{,}05 \cdot 3=0{,}15$

7. Lee K. L. et al. Clinical judgment and statistics. Lessons from a simulated randomized trial in coronary artery disease / K. K. Lee, J. F. McNeer, C. F. Starmer et al. // Circulation. – 1980. – Vol. 61. – N 3. – P. 508–515.

- Симуляция изучения эффективности двух различных методов лечения ишемической болезни сердца.
- Две равные группы, ***одно и то же лечение***!
- Данные были обработаны так, как будто бы одной группе назначалось лечение А, а другой – лечение Б.
- При сравнении эффективности ≪двух видов лечения≫ различий обнаружено не было.
- Разбили каждую из групп пациентов еще на 6 по количеству пораженных коронарных артерий (1, 2 или 3 сосуда) и сократительной способности миокарда левого желудочка (выше или ниже определенного критического уровня).
- Результаты лечения не различались в пяти подгруппах, а в подгруппе пациентов с наиболее тяжелой формой заболевания лечение А было более эффективно ($p = 0,025$).
- Но в действительности обе группы получали **одно и то же лечение**!

Поправка Бонферрони

Если мы хотим обеспечить вероятность ошибки первого рода α, то в каждом из сравнений мы должны принять уровень значимости α/k, где k – число попарных сравнений

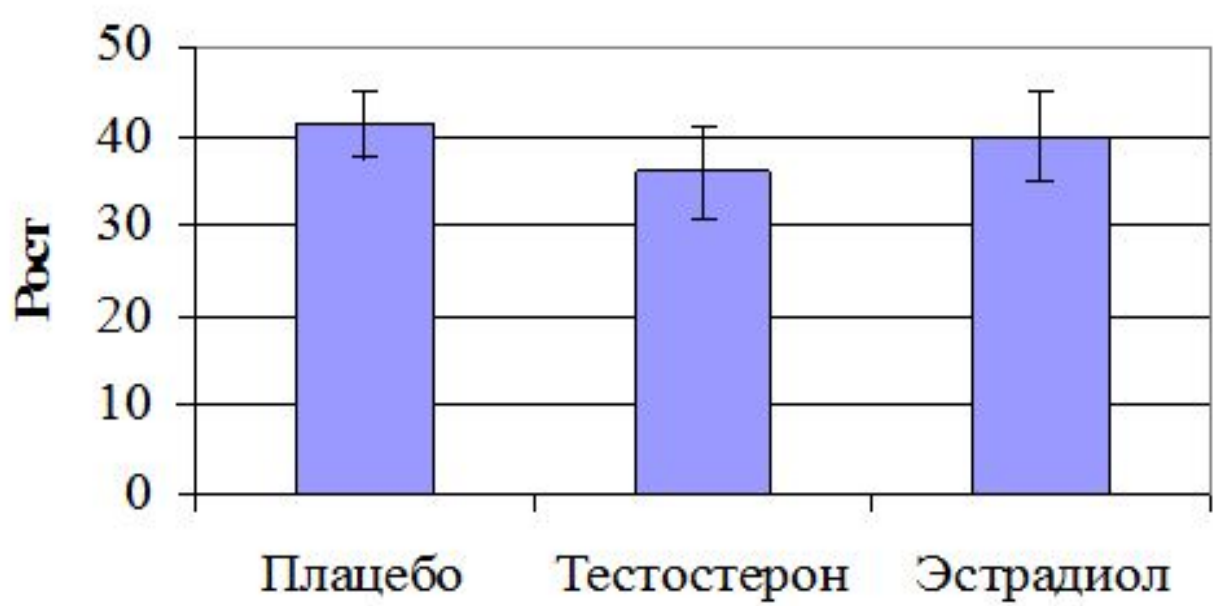
При сравнении нескольких групп с одной контрольной k=m-1, где m – количество групп.

Множественные парные сравнения групп и подгрупп обоснованы, если они запланированы в начале исследования, до начала сбора данных!

Три случайные выборки из одной совокупности:

- N=200, μ=40, σ=5

Рост	Частота	Рост	Частота	Рост	Частота
31	1	34	1	34	2
32	1	37	1	36	1
33	1	40	2	38	1
34	1	42	2	39	3
35	2	43	1	44	2
38	1	44	1	50	1
41	1	46	2		
42	1				
46	1				
$\overline{X_1} = 36$ $s_1 = 5$		$\overline{X_2} = 41{,}5$ $s_2 = 3{,}8$		$\overline{X_3} = 40$ $s_3 = 5$	



- плацебо-тестостерон $t=2,39$;
- плацебо - эстрадиол $t=0,93$;
- тестостерон - эстрадиол $t=1,34$.

$\nu = 10+10-2=18$, $t_{0,05;18}=2,101$.

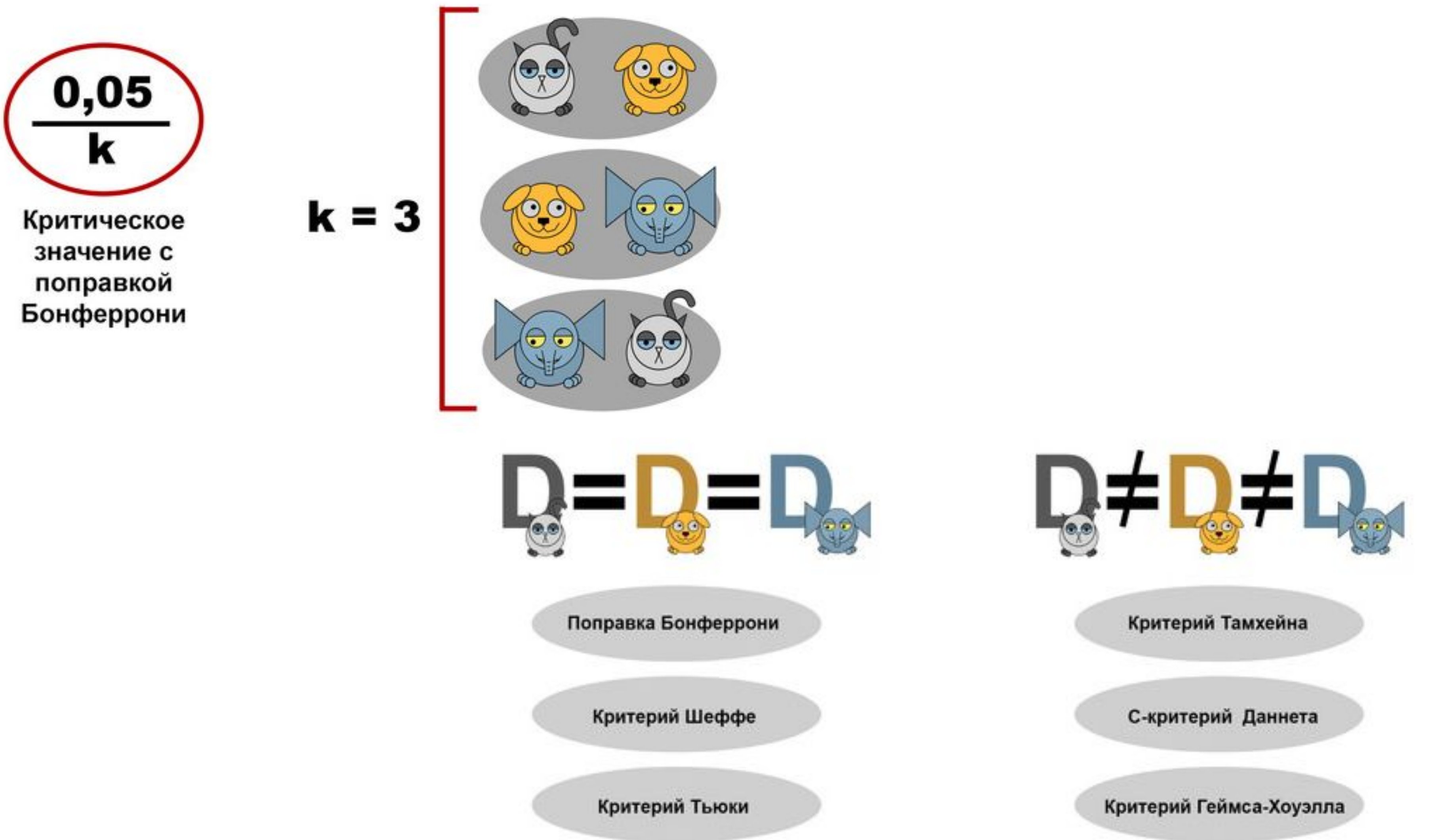
$k=3$, $\alpha=0,05/3=0,017$

$t_{0,02;18}=2,552 > 2,39$ → нет значимых различий!

В. Савельев «**СТАТИСТИКА И КОТИКИ**»

http://www.statcats.ru

https://lib.rus.ec/b/624980



Критерий Стьюдента для сравнения средних в двух взаимосвязанных выборках

(*Парный критерий Стьюдента, критерий Стьюдента для повторных измерений)*

Выборки называются ***независимыми (несвязанными),*** если процедура эксперимента и полученные результаты измерения некоторого признака у испытуемых одной выборки не оказывают влияния на особенности протекания этого же эксперимента и результаты измерения этого же признака у испытуемых другой выборки.

И, напротив, выборки называется ***зависимыми (связанными)*** если процедура эксперимента и полученные результаты измерения некоторого свойства, проведенные на одной выборке, оказывают влияние на другую.

В зависимых выборках одному случаю из первой выборки соответствует один случай из второй выборки и наоборот. <u>Примеры зависимых выборок</u>:

- пары близнецов;
- два измерения какого-либо признака до и после экспериментального воздействия,
- мужья и жёны
- родители и дети и т.д.

Зависимые выборки всегда имеют одинаковый объём, а объём независимых может отличаться

<u>Пример</u>. Некий исследователь выдвинул «гипотезу» о том, что люди выше, когда они в обуви, чем когда они босиком.

Схема эксперимента: в случайной выборке из 15 взрослых людей измерили рост каждого в обуви и без нее.

A	164	179	176	151	156	177	175	164	162	157	176	160	192	176	150
B	161	175	172	147	152	174	170	160	157	151	174	155	188	172	148
A-B	3	4	4	4	4	3	5	4	5	6	2	5	4	4	2

$\overline{X}_A = 167{,}7$; $s_A = 12{,}03$; $\overline{X}_B = 163{,}7$; $s_B = 12{,}7$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

- t = 0,89. Для уровня значимости α=0,05 и числа степеней свободы ν=28 критическое значение t равно 2,05. Рассчитанное значение меньше критического. Различия не являются статистически значимыми???

Причина: разность средних (равна 4) очень мала по сравнению с разбросом значений в каждой из выборок (стандартное отклонение 12,03 и 12,17)

На самом деле нас интересует только разница между двумя группами. Здесь есть только одна выборка D: разность между двумя измерениями.

- H_0 – среднее значение в выборке не отличается от 0
- H_1 – среднее значение в выборке отличается от 0

$$t = \frac{\overline{X_D} - 0}{s_{\overline{X_D}}} = \frac{\overline{X_D}}{s_D / \sqrt{n}}$$

Число степеней свободы v=n-1

$$\overline{X_D} = 3,93$$ s_D=1,1

$$s_{\overline{X_D}} = 1,1 / \sqrt{15} = 0,28$$

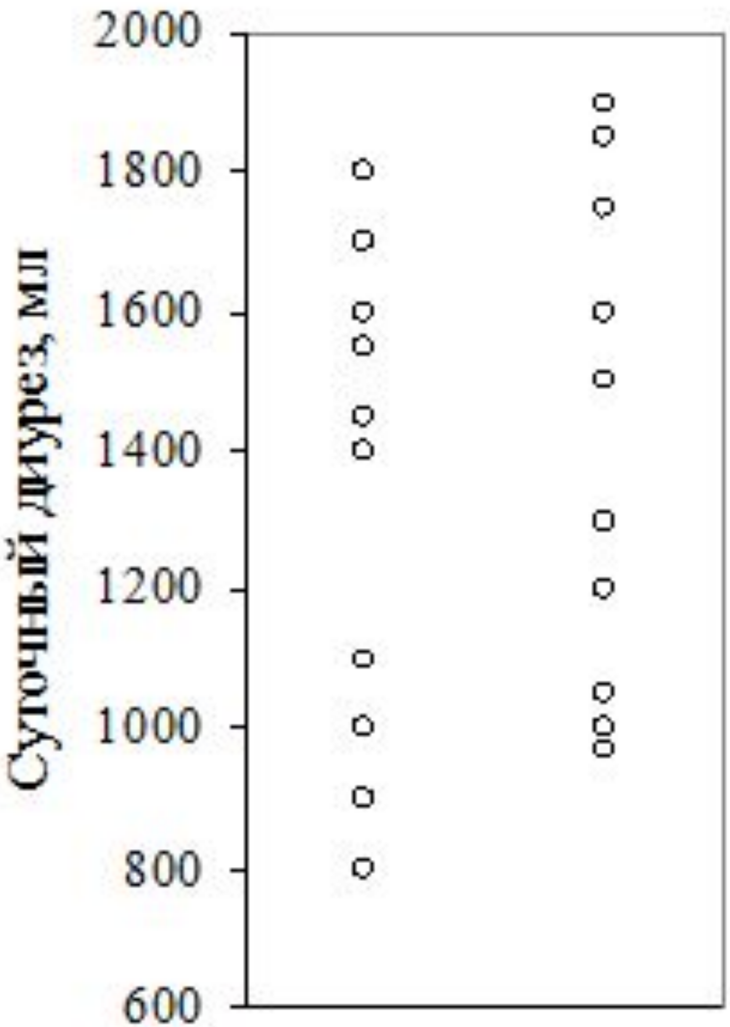
t=13,85; v=14; $t_{0,05}$= 2,145; $t_{0,001}$=4,14

Часто значительная часть внутригрупповой изменчивости (вариации) в обеих группах может быть объяснена индивидуальными различиями субъектов.

В случае независимых выборок нельзя определить (или «удалить») часть вариации, связанную с индивидуальными различиями субъектов.

Если та же самая выборка тестируется дважды, то можно легко исключить эту часть вариации.

Пример. Проводилось изучение суточного диуреза у 10 человек после приема препарата и у 10 после приема плацебо.



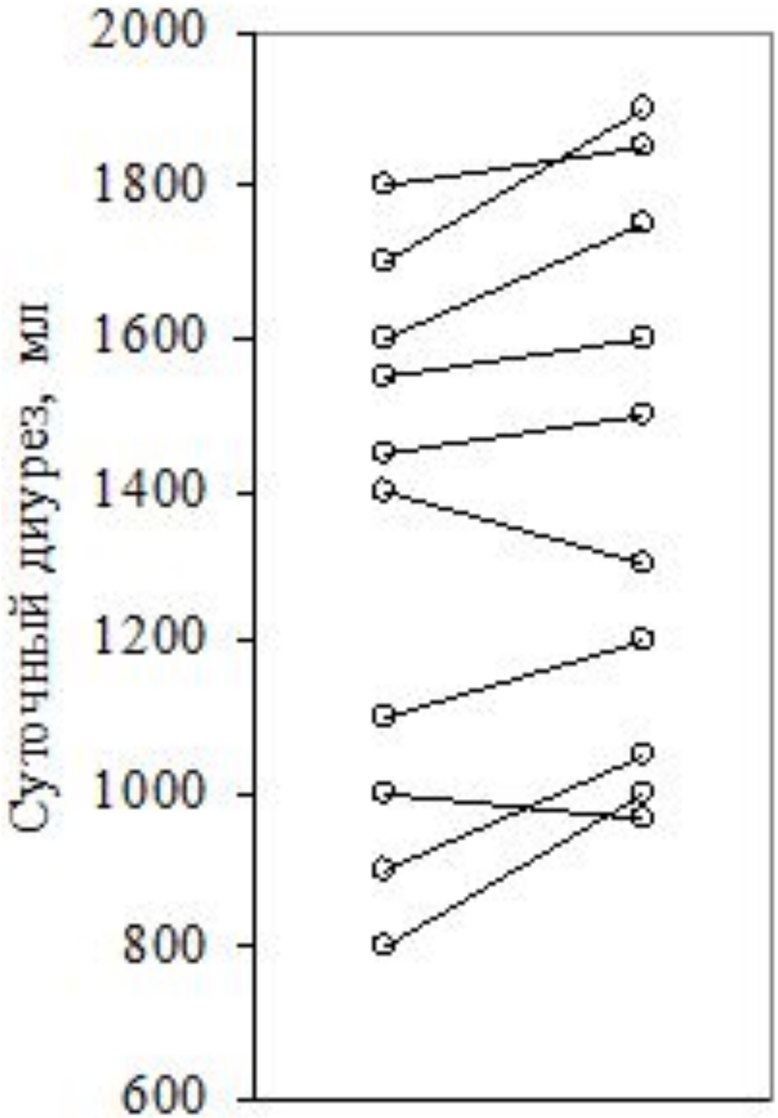
$\overline{X}_к = 1330$ мл

$s_к = 353{,}7$ мл

$\overline{X}_э = 1412$ мл

$s_э = 356{,}1$ мл

t=0,52 – нет значимых различий



$\overline{X_D} = 82,$

$s_D = 97{,}84$

t=2,65

Различия статистически значимы

<u>Условие применения</u>: нормальное распределение разности между парами значений

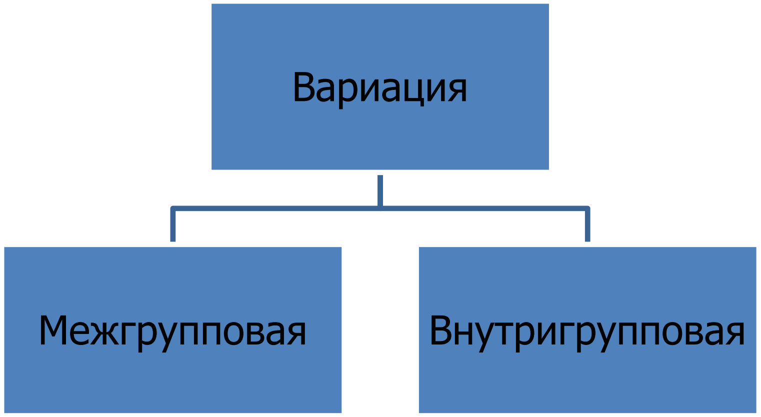
Если схема эксперимента предполагает не две, а три и более групп?

Попарные сравнения групп–

проблема множественных сравнений!!!

Дисперсионный анализ
(ANOVA – analysis of variance)

- Разработан в 20-х годах прошлого века английским математиком и генетиком Р.Фишером
- Выявляет статистически значимые различия между несколькими группами

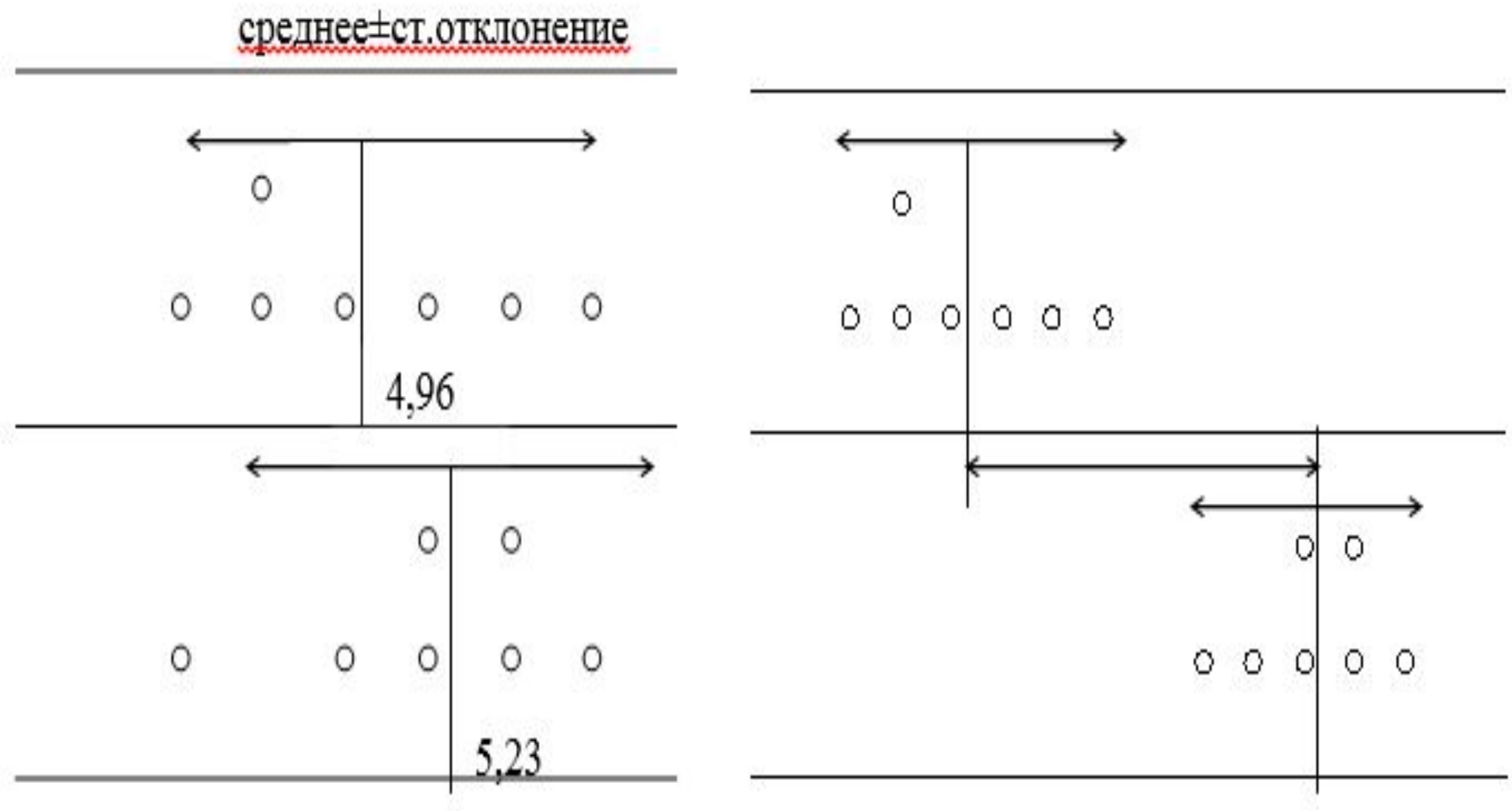


- Значение критерия - отношение межгрупповой вариации к внутригрупповой

Пример. Ученые исследовали влияние диеты на сердечный выброс. Случайным образом отобрали 28 человек и разделили их на 4 группы по 7 человек в каждой. Члены первой (контрольной) группы продолжали питаться как обычно, второй – ели преимущественно макароны, третьей – мясо, четвертой – фрукты. Через месяц у всех участников эксперимента измерили сердечный выброс.

	1	2	3	4
Среднее	4,96	5,23	4,93	4,8
Дисперсия	0,1	0,12	0,22	0,2
Стандартное отклонение	0,315	0,33	0,46	0,44

Нулевая гипотеза: ни одна из диет не влияет на сердечный выброс. Как убедиться в этом?



Оценка дисперсии совокупности:

1) на основании дисперсий в каждой группе. Такая оценка не зависит от различий групповых средних.

2) по разбросу выборочных средних. Такая оценка зависит от различий выборочных средних.

Если экспериментальные группы являются случайными выборками из одной и той же нормально распределенной совокупности, то обе оценки дисперсии дают примерно одинаковые результаты

Оценка по выборочным дисперсиям:

$$s^2_{\text{внутр}} = \frac{\sum s_i^2}{k}$$

k – количество групп,
s_i^2- выборочные оценки дисперсии в группах.

Оценка по выборочным средним

$$s^2_{\text{меж}} = n s^2_{\bar{X}},$$

где n – объем выборок,
$s^2_{\bar{X}}$ - дисперсия выборки, составленной из выборочных средних.

$$F = \frac{s^2_{меж}}{s^2_{внутр}}$$

$\nu_{меж}=k-1$, где k-число групп.

$\nu_{внутр}=k(n-1)$, где n – численность каждой группы.

Если рассчитанное значение F будет больше, чем табличное для соответствующего числа степеней свободы и уровня значимости, то нулевая гипотеза о равенстве выборочных средних отвергается – различия будут статистически значимыми.

Этапы дисперсионного анализа

1. Проверка нормальности в каждой из групп
2. Проверка гипотезы о равенстве дисперсий (тест Левена)

   Если условия 1-2 не выполняются, следует применить непараметрический аналог дисперсионного анализа!

3. Собственно анализ вариаций
4. Апостериорное сравнение групп с помощью специальных процедур

Примеры

1. Женщины с остеопорозом были распределены случайным образом по трем группам:

– лечение по стандартной методике,
– лечение по новой методике
– плацебо (контрольная группа).

Исследуемой переменной является изменение минеральной плотности костной ткани, по которому различаются группы. Результаты можно проанализировать с помощью однофакторного дисперсионного анализа.

2. В условиях предыдущего примера добавляем в качестве второй группирующей переменной возраст. Возраст классифицируется как одна из четырех порядковых категорий: от 30 до 40 лет, от 41 до 50, от 51 до 60, от 61 года и старше. Данные можно проанализировать с помощью двухфакторного дисперсионного анализа

3. В условиях предыдущего примера добавление новых категориальных переменных, таких как диета (вегетарианская или невегетарианская) и употребление алкоголя (менее 60 мл алкоголя в день, от 60 до 150 мл в день, более 150 мл в день), может превратить двухфакторный анализ в четырехфакторный или многофакторный дисперсионный

$$s^2_{внутр} = \frac{0,1+0,12+0,22+0,2}{4} = 0,16$$

$$\overline{X} = \frac{4,96+5,23+4,93+4,8}{4} = 4,98;$$

$$s^2_{\overline{X}} = \frac{(4,98-4,96)^2+(4,98-5,23)^2+(4,98-4,93)^2+(4,98-4,8)^2}{4-1} = 0,0326$$

$$s^2_{меж} = 7 \cdot 0,0326 = 0,23; \qquad F = 0,23/0,16 = 1,44$$

$\nu_{меж}=4-1=3$; $\nu_{внутр}=4(7-1)=24$;

$F_{0,05;3;24}=3,01.$

Диета из рассмотренного примера не влияет на сердечный выброс

Обобщение метода на случай неравной численности групп

- Имеется k групп, n_i – численность i-ой группы
- $\overline{X_i}$ - среднее в i-ой группе
- s_i^2 – дисперсия в i-ой группе
- $N = \sum n_i$ - общий объем исследования

$$F = \frac{S_{меж} / \nu_{меж}}{S_{внутр} / \nu_{внутр}}$$

$$S_{внутр} = \sum_{i=1}^{k} (n_i - 1) s_i^2, \quad \nu_{внутр} = N - k;$$

$$S_{меж} = \sum_{i=1}^{k} n_i \overline{X_i}^{\,2} - \frac{(\sum_{i=1}^{k} n_i \overline{X_i})^2}{N}, \quad \nu_{меж} = k - 1.$$

Курение считают основным фактором, предрасполагающим к хроническим обструктивным заболеваниям легких. Является ли таким фактором пассивное курение?

Для проверки данного предположения изучалась проходимость дыхательных путей у некурящих, активных и пассивных курильщиков. Измерялась максимальная объемная скорость середины вдоха (л/с) у некурящих, активных и пассивных курильщиков. Ее уменьшение свидетельствует о нарушении проходимости дыхательных путей.

Можно ли считать этот показатель одинаковым во всех группах? (Выборки считать извлеченными из нормально распределенной

Данные по максимальной объемной скорости середины вдоха (л/с)

Группа	Численность	Среднее значение	Стандартное отклонение
1) Некурящие, работающие в помещении, где не курят;	200	3,17	0,74
2) Некурящие, работающие в накуренном помещении;	201	2,72	0,71
3) Курящие небольшое количество сигарет;	199	2,63	0,73
4) Курящие среднее количество сигарет;	202	2,29	0,7
5) Курящие большое количество сигарет	198	2,12	0,72

Количество групп k=5, общая численность исследования N=1000 человек.

$$S_{меж} = \sum_{i=1}^{k} n_i \overline{X_i}^{\,2} - \frac{(\sum_{i=1}^{k} n_i \overline{X_i})^2}{N} = (200 \cdot 3{,}17^2 + 201 \cdot 2{,}72^2 + 199 \cdot 2{,}63^2 + 202 \cdot 2{,}29^2 + 200 \cdot 2{,}12^2) -$$

$$- \frac{(200 \cdot 3{,}17 + 201 \cdot 2{,}72 + 199 \cdot 2{,}63 + 202 \cdot 2{,}29 + 200 \cdot 2{,}12)^2}{1000} = 132{,}9$$

$$\nu_{меж} = k - 1 = 4.$$

$$S_{внутр} = \sum_{i=1}^{k} (n_i - 1) s_i^2 = (200-1) \cdot 0{,}74^2 + (201-1) \cdot 0{,}71^2 + (199-1) \cdot 0{,}73^2 + (202-1) \cdot 0{,}7^2 +$$

$$+ (198-1) \cdot 0{,}72^2 = 515{,}9$$

$$\nu_{внутр} = N - k = 995$$

$$F = \frac{S_{меж} / \nu_{меж}}{S_{внутр} / \nu_{внутр}} = \frac{132{,}9 / 4}{515{,}9 / 995} = 64{,}1$$

Таблица 2

Критические значения F для α=0,05 (верхняя строка) и α=0,01 (*нижняя строка*)

Число степеней свободы для знаменателя ($\nu_{внутр}$)	*Число степеней свободы для числителя ($\nu_{меж}$)* 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245
	4052	*4999*	*5404*	*5624*	*5764*	*5859*	*5928*	*5981*	*6022*	*6056*	*6083*	*6107*	*6126*	*6143*
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42
	98.5	*99.0*	*99.16*	*99.25*	*99.30*	*99.33*	*99.36*	*99.38*	*99.39*	*99.40*	*99.41*	*99.42*	*99.42*	*99.43*
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71
	34.12	*30.82*	*29.46*	*28.71*	*28.24*	*27.91*	*27.67*	*27.49*	*27.34*	*27.23*	*27.13*	*27.05*	*26.98*	*26.92*
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87
	21.20	*18.00*	*16.69*	*15.98*	*15.52*	*15.21*	*14.98*	*14.80*	*14.66*	*14.55*	*14.45*	*14.37*	*14.31*	*14.25*
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64
	16.26	*13.27*	*12.06*	*11.39*	*10.97*	*10.67*	*10.46*	*10.29*	*10.16*	*10.05*	*9.96*	*9.89*	*9.82*	*9.77*
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96
	13.75	*10.92*	*9.78*	*9.15*	*8.75*	*8.47*	*8.26*	*8.10*	*7.98*	*7.87*	*7.79*	*7.72*	*7.66*	*7.60*
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53
110	3.93	3.08	2.69	2.45	2.30	2.18	2.09	2.02	1.97	1.92	1.88	1.84	1.81	1.78
	6.87	*4.80*	*3.96*	*3.49*	*3.19*	*2.97*	*2.81*	*2.68*	*2.57*	*2.49*	*2.41*	*2.35*	*2.30*	*2.25*
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80	1.78
	6.85	*4.79*	*3.95*	*3.48*	*3.17*	*2.96*	*2.79*	*2.66*	*2.56*	*2.47*	*2.40*	*2.34*	*2.28*	*2.23*
130	3.91	3.07	2.67	2.44	2.28	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.80	1.77
	6.83	*4.77*	*3.94*	*3.47*	*3.16*	*2.94*	*2.78*	*2.65*	*2.55*	*2.46*	*2.39*	*2.32*	*2.27*	*2.22*
140	3.91	3.06	2.67	2.44	2.28	2.16	2.08	2.01	1.95	1.90	1.86	1.82	1.79	1.76
	6.82	*4.76*	*3.92*	*3.46*	*3.15*	*2.93*	*2.77*	*2.64*	*2.54*	*2.45*	*2.38*	*2.31*	*2.26*	*2.21*
160	3.90	3.05	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.81	1.78	1.75
	6.80	*4.74*	*3.91*	*3.44*	*3.13*	*2.92*	*2.75*	*2.62*	*2.52*	*2.43*	*2.36*	*2.30*	*2.24*	*2.20*
180	3.89	3.05	2.65	2.42	2.26	2.15	2.06	1.99	1.93	1.88	1.84	1.81	1.77	1.75
	6.78	*4.73*	*3.89*	*3.43*	*3.12*	*2.90*	*2.74*	*2.61*	*2.51*	*2.42*	*2.35*	*2.28*	*2.23*	*2.18*
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74
	6.76	*4.71*	*3.88*	*3.41*	*3.11*	*2.89*	*2.73*	*2.60*	*2.50*	*2.41*	*2.34*	*2.27*	*2.22*	*2.17*

Рассчитанное значение (64,1) больше табличного (3,41 для уровня 0,01).

Можем опровергнуть нулевую гипотезу с уровнем значимости 0,01 и утверждать, что максимальная объемная скорость середины вдоха в группах статистически значимо различается (вероятность ошибки менее 1%)

Критерий Стьюдента с точки зрения дисперсионного анализа

Критерий Стьюдента является вариантом дисперсионного анализа в случае сравнения двух групп, при этом выполняется равенство $F=t^2$.

Межгрупповое число степеней свободы будет равно $v_{меж}=k–1=2–1=1$;

внутригрупповое $v_{внутр}=k(n–1)=2(n–1)$

Средняя продолжительность госпитализации 36 больных пиелонефритом, получавших правильное (соответствующее официальным рекомендациям) лечение, составила 4,51 суток, а у 36 больных, получавших неправильное лечение – 6,28 суток. Стандартные отклонения для этих групп составили соответственно 1,98 суток и 2,54 суток. Можно ли считать эти различия случайными?

$$s^2 = \frac{s_1^2 + s_2^2}{2} = \frac{1{,}98^2 + 2{,}54^2}{2} = 5{,}18$$

$$t = \frac{4{,}51 - 6{,}28}{\sqrt{\frac{5{,}18}{36} + \frac{5{,}18}{36}}} = -3{,}3$$

Число степеней свободы ν = 2(n–1) = 2 (36 – 1)= 70. Для α = 0,01 и ν=70 $t_{крит}$=2,648 . Следовательно, различия в сроках госпитализации статистически значимы. Вероятность ошибки данного заключения составляет менее 1%.

$$s^2_{\text{внутр}} = \frac{(s_1^2 + s_2^2)}{2} = \frac{1{,}98^2 + 2{,}54^2}{2} = 5{,}19$$

$$\overline{X} = \frac{\overline{X_1} + \overline{X_2}}{2} = \frac{4{,}51 + 6{,}28}{2} = 5{,}4$$

Дисперсия для среднего по выборочным средним равна:

$$s^2_{\overline{X}} = \frac{(\overline{X_1} - \overline{X})^2 + (\overline{X_2} - \overline{X})^2}{k-1} = \frac{(4{,}51 - 5{,}4)^2 + (6{,}28 - 5{,}4)^2}{1} = 1{,}56$$

$$s^2_{меж} = ns^2_{\overline{X}} = 36 \cdot 1{,}56 = 56{,}26$$

$$F = \frac{s^2_{меж}}{s^2_{внутр}} = \frac{56{,}26}{5{,}19} = 10{,}84$$

$$\nu_{меж} = k - 1 = 1; \qquad \nu_{\text{внутр}} = k(n-1) = 2(36-1) = 70; \quad F_{0{,}01;1;70} = 7{,}01$$

$t^2 = (-3{,}3)^2 = 10{,}89 \approx F.$

Дисперсионный анализ повторных измерений

В дисперсионном анализе повторных измерений одна и та же группа последовательно подвергается действию изучаемого фактора или просто наблюдается в несколько последовательных моментов времени.

- **Вариация**
  - **Межиндивидуальная**
  - **Внутрииндивидуальная**
    - **Факторная**
    - **Остаточная**

$$F = \frac{S_{\Phi} / \nu_{\Phi}}{S_{ОСТ} / \nu_{ОСТ}}$$