Лекция 3

Дисперсионный анализ ANOVA

Для закрепления материала по прошлому занятию разберем конкретный пример использования t-критерия Стьюдента

Мы с вами, врачи-исследователи, изучаем действие какого-то препарата на заболевание. Мы произвели замеры показателя болезни у каждого пациента до приема препарата и после. Нам нужно доказать, что препарат действует, соответственно, показатель болезни у пациентов после приема препарата статистически значимо уменьшается.

H_0 (**нулевая гипотеза**, null hypothesis) – её мы собираемся опровергать, говорит, что нет различий, нет эффекта, нет изменений (препарат не действует).

H_1 (**альтернативная гипотеза**, alternative hypothesis) – её мы примем, если удастся отвергнуть H_0 .

Проведем анализ в программе «Паст»

1. Вбиваем данные в два столбца →
2. Проверяем данные в обоих столбцах на нормальность распределения

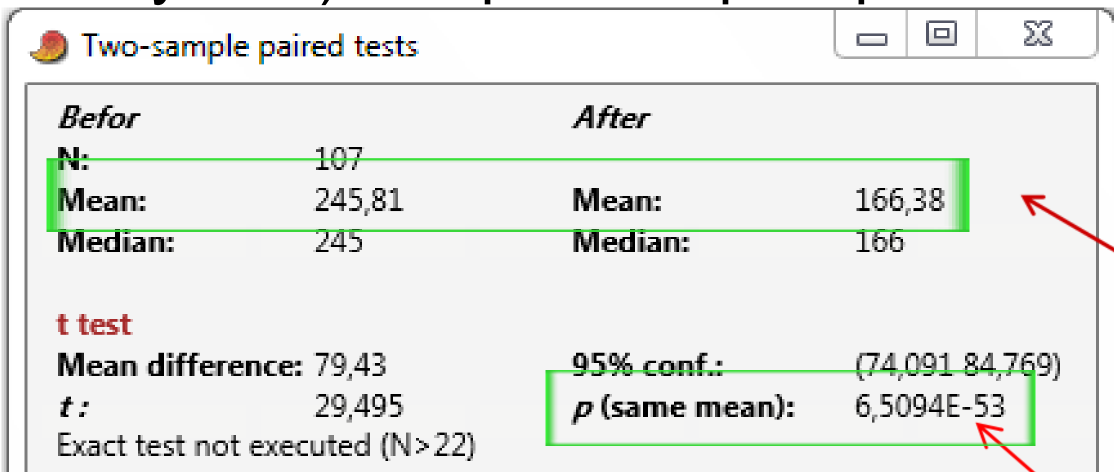
P>0,1 – распределение нормальное

Tests for normal distribution

	Befor	After
N	107	107
Shapiro-Wilk W	0,9874	0,9902
p(normal)	0,4164	0,6328

	Befor	After
1	195	125
2	208	164
3	254	152
4	226	144
5	290	212
6	239	171
7	216	164
8	286	200
9	243	190
10	217	130
11	245	170
12	257	182
13	199	153
14	277	204
15	249	174
16	197	160
17	279	205
18	226	159
19	262	170
20	231	180
21	234	161
22	170	139
23	242	159
24	186	114
25	223	134
26	220	166
27	277	170
28	235	136
29	216	134
30	197	138

3. Решаем какой t-критерий Стьюдента выбрать. Выборки зависимы (пациенты одни и те же в обоих случаях) – парный t-критерий



4. Выписываем значения t и p. Так как p намного меньше 0,05 заключаем, что имеются статистически значимые различия

5. Среднее значение (Mean) во втором столбике меньше, чем в первом

Общий вывод:

Препарат работает и приводит к снижению показателя заболевания

И мы с вами это ДОКАЗАЛИ !

ANOVA

Сравнение **ТРЕХ** И **БОЛЕЕ** групп

Дисперсионный анализ

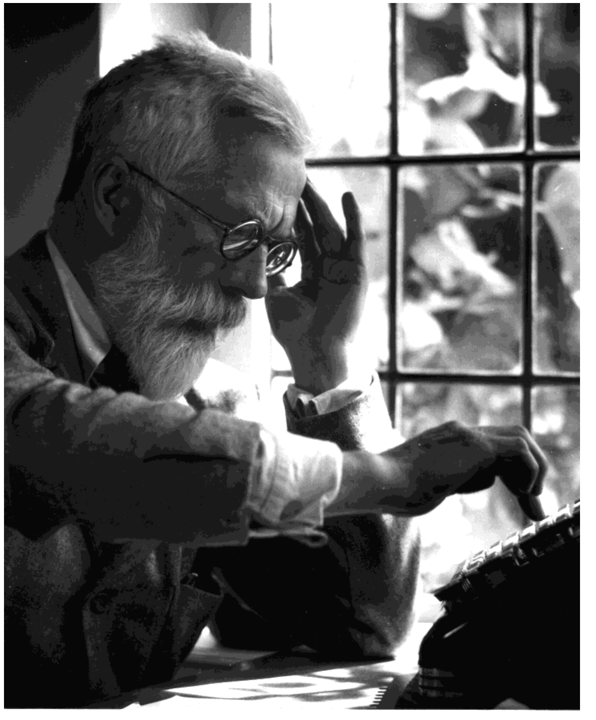
Дисперсионный анализ — метод, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях.

В отличие от t-критерия, позволяет сравнивать средние значения **трёх и более** групп.

В литературе также встречается обозначение ANOVA (от англ. ANalysis Of VAriance).

Дисперсионный анализ
ANOVA (analysis of variance)

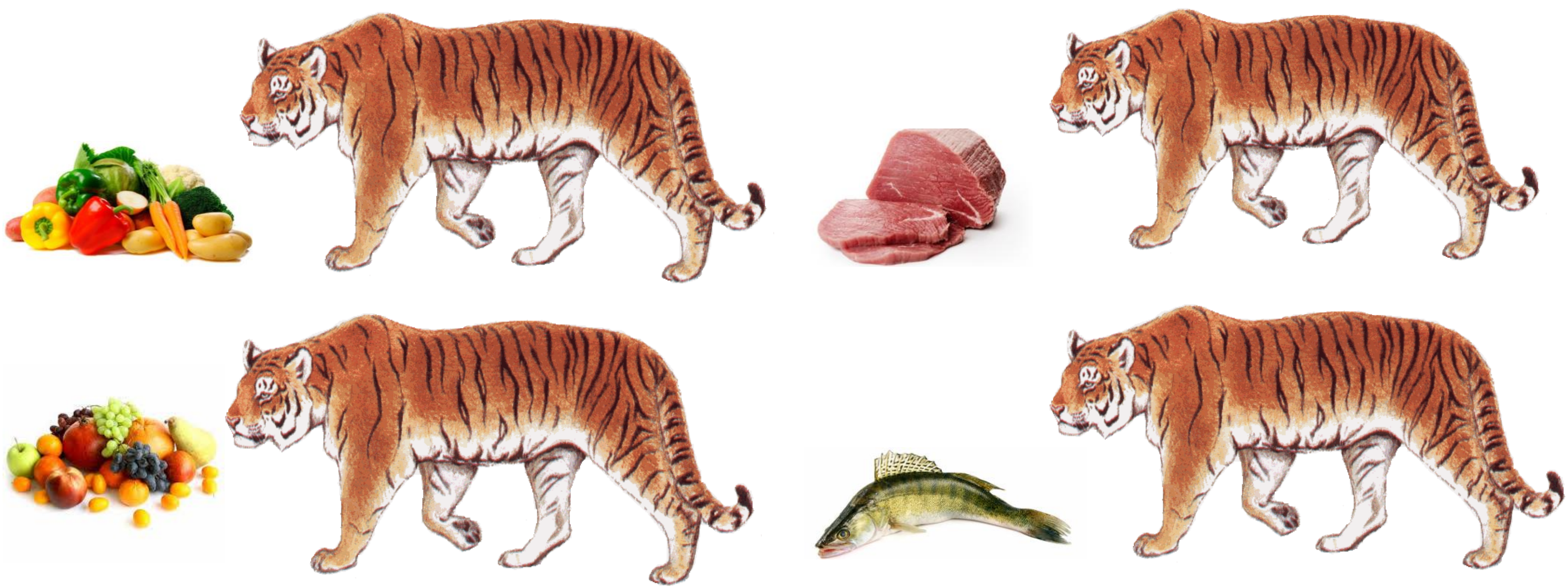
Разработан Роналдом Фишером для анализа результатов экспериментальных исследований



Сэр Ро́налд Э́йлмер Фи́шер (Sir Ronald Aylmer Fisher *1890-1962*)

Мы тестировали гипотезы о среднем значении для одной и двух выборок.
Теперь выборок **три или больше**.

Например, у нас 4 группы тигров, которых кормят по-разному. Различается ли средняя масса тигра в этих группах?



Одна зависимая переменная (variable): масса;
Одна независимая (группирующая, factor) – тип еды.

One-way ANOVA

однофакторный дисперсионный анализ

Формулируем гипотезу H_0:

Тигров кормили:

1. овощами;
2. фруктами;
3. рыбой;
4. мясом.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Это сложная гипотеза (omnibus hypothesis). Она включает в себя много маленьких гипотез :

$$\left.\begin{aligned} H_{01} &: \mu_1 = \mu_2 \\ H_{02} &: \mu_1 = \mu_4 \\ H_{03} &: \mu_1 = \mu_3 \\ H_{04} &: \mu_2 = \mu_3 \\ H_{05} &: \mu_2 = \mu_4 \\ H_{06} &: \mu_3 = \mu_4 \end{aligned}\right\}$$

Парные (pairwise) нулевые гипотезы

$$\left.\begin{aligned} H_{07} &: \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2} \\ H_{08} &: \mu_1 = \frac{\mu_2 + \mu_3 + \mu_4}{3} \\ &\dots \end{aligned}\right\}$$

Комплексные (complex) нулевые гипотезы

Формулируем альтернативную гипотезу:

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$ **?**

НЕВЕРНО!

H_1: $\mu_1 \neq \mu_2$ или $\mu_1 \neq \mu_3$ или $\mu_1 \neq \mu_4$ …

Мы отвергаем общую H_0 гипотезу если верна <u>хотя бы одна</u> из маленьких частных альтернативных гипотез (парных или комплексных)!

Какая именно – ANOVA не говорит.

Почему бы не сравнить группы попарно *t*-критерием? (Ошибка использования критерия Стьюдента)

1. мы таким образом проверяем не все гипотезы, которые содержатся в сложной гипотезе;
2. резко увеличивается вероятность найти различия, где их нет! (общая вероятность ошибки 1-го рода).

Эффект МНОЖЕСТВЕННЫХ СРАВНЕНИЙ

(при сравнении нескольких групп попарно).

Level of Significance, α, Used in the t Tests			
k	C	0.05	0.01
2	1	0.05	0.01
3	3	0.14	0.03
4	6	0.26	0.06
5	10	0.40	0.10
6	15	0.54	0.14
10	45	0.90	0.36
	∞	1.00	1.00

*There are $C = k(k-1)/2$ pairwise comparisons of k means.

В случае, если все H_0 верны, суммарная вероятность ошибки 1-го рода в эксперименте из C сравнений -

$$p = 1 - (1 - \alpha)^C$$

В случае 4-х групп ≈ **26,5%**

Zar, 2010

Общая логика ANOVA

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$ (т.е., средние в 4-х популяциях равны)

Формируем 4 независимых случайных выборки и считаем выборочные средние для каждой из них (они оценивают популяционные средние).

Если H_0 верна, выборочные средние должны быть **примерно** одинаковы.

Чем дальше друг от друга отстоят средние значения в группах, тем меньше вероятность, что верна H_0



Пусть все группы будут одинакового размера (для простоты объяснения).

Если H_0 верна, то 4 наших группы получены из ОДНОЙ популяции с конкретными средним μ и дисперсией σ^2.

Получим 2 независимые **точечные оценки σ^2 и сравним их!**

На этой идее основана АНОВА.

1. Оценим σ^2 на основе дисперсии групповых средних (как будто это выборочные средние)

2. Оценим σ^2 на основе дисперсий внутри групп

овощи	фрукты	мясо	рыба
151	108	147	130
135	94	138	128
137	84	143	140
118	87	135	142
132	82	153	139
135	79	137	145
131	74	148	144
137	73	140	140
121	67	144	141
140	78	146	140
152	63	151	142
133	90	145	137
151	81	146	148
132	96	147	142
139	83	150	143
96	89	144	140
133,7	***83***	***144,6***	***140,1***

1. Оценка общей дисперсии по разбросу МЕЖДУ группами

средние в каждой группе

общее среднее

$$MS_B = \boxed{s_{\bar{X}}^2 n} = \frac{\sum\left(\bar{X}_j - \bar{X}_G\right)^2}{k-1} n$$

$df_B = k\text{-}1$

число групп -1 (4 - 1 = 3)

размер группы

MS_B – **me**an **s**quare **b**etween groups, средний квадрат - оценка расстояния между средними в группах. (сокращение от mean squared deviation from the mean)

различия большие - H_0 не верна

MS_B = groups MS

овощи	фрукты	мясо	рыба
151	108	147	130
135	94	138	128
137	84	143	140
118	87	135	142
132	82	153	139
135	79	137	145
131	74	148	144
137	73	140	140
121	67	144	141
140	78	146	140
152	63	151	142
133	90	145	137
151	81	146	148
132	96	147	142
139	83	150	143
96	89	144	140
133,7	***83***	***144,6***	***140,1***

2. Оценка общей дисперсии по разбросу ВНУТРИ групп

сумма квадратов стандартных отклонений внутри групп

$$MS_W = \frac{\overbrace{s_1^2 + s_2^2 + s_3^2 + \ldots + s_k^2}}{k}$$

число групп

$$\boxed{df_W = n_G - k}$$

MS_w = error MS

Статистика критерия: F

$$F = \frac{\text{оценка дисперсии } \textbf{между} \text{ группами}}{\text{оценка дисперсии } \textbf{внутри} \text{ групп}} \qquad F = \frac{MS_B}{MS_W}$$

✓Не соответствует общей формуле

$$\text{Статистика} = \frac{\text{параметр } \textbf{выборки} - \text{параметр } \textbf{популяции}}{\text{стандартная } \textbf{ошибка} \text{ параметра выборки}}$$

✓Приводится как F_{df_B, df_W}, т.е., например, $F_{3,60}$

ANOVA table

источник изменчивости	***SS***	***df***	***MS***	***F***
между	SS_B	df_B	MS_B	F
внутри	SS_W	df_W	MS_W	
общее	SS_T	df_T		

$$SS_T = SS_W + SS_B$$

$$df_T = df_W + df_B$$

$df_W = n_G - k$ $df_B = k$-1

SS это суммы квадратов отклонений (sum of squared deviations):
SS_B - средних в группах от общего среднего = **Effect**;
SS_W – измерений от средних в группах = **Error**.

$$MS_W = \frac{SS_W}{df_W}$$

$$\boxed{F = \frac{MS_B}{MS_W}}$$

$$MS_B = \frac{SS_B}{df_B}$$

Связь с двухвыборочным t-критерием Стьюдента

В случае числа групп k=2 однофакторный дисперсионный анализ ANOVA эквивалентен t-критерию Стьюдента, причём

$$F = t^2$$

MS_w идентично s_{pooled}

Т.е., условия применения и выводы одинаковы, за исключением того, что ANOVA – всегда двусторонний тест.

ANOVA effect size

«Практическая значимость» результата:

1. $$\boxed{\eta^2 = \frac{SS_B}{SS_T}}$$

η = 0.0099 – маленький эффект,
η = 0.0588 – средний эффект,
η = 0.1379 – большой эффект.

2. $$f = \frac{s_{\overline{X}}}{\sqrt{MS_W}}$$

f = 0.1 – маленький эффект
f = 0.25 – средний эффект
f = 0.4 – большой эффект

SS_B - межгрупповая сумма квадратов

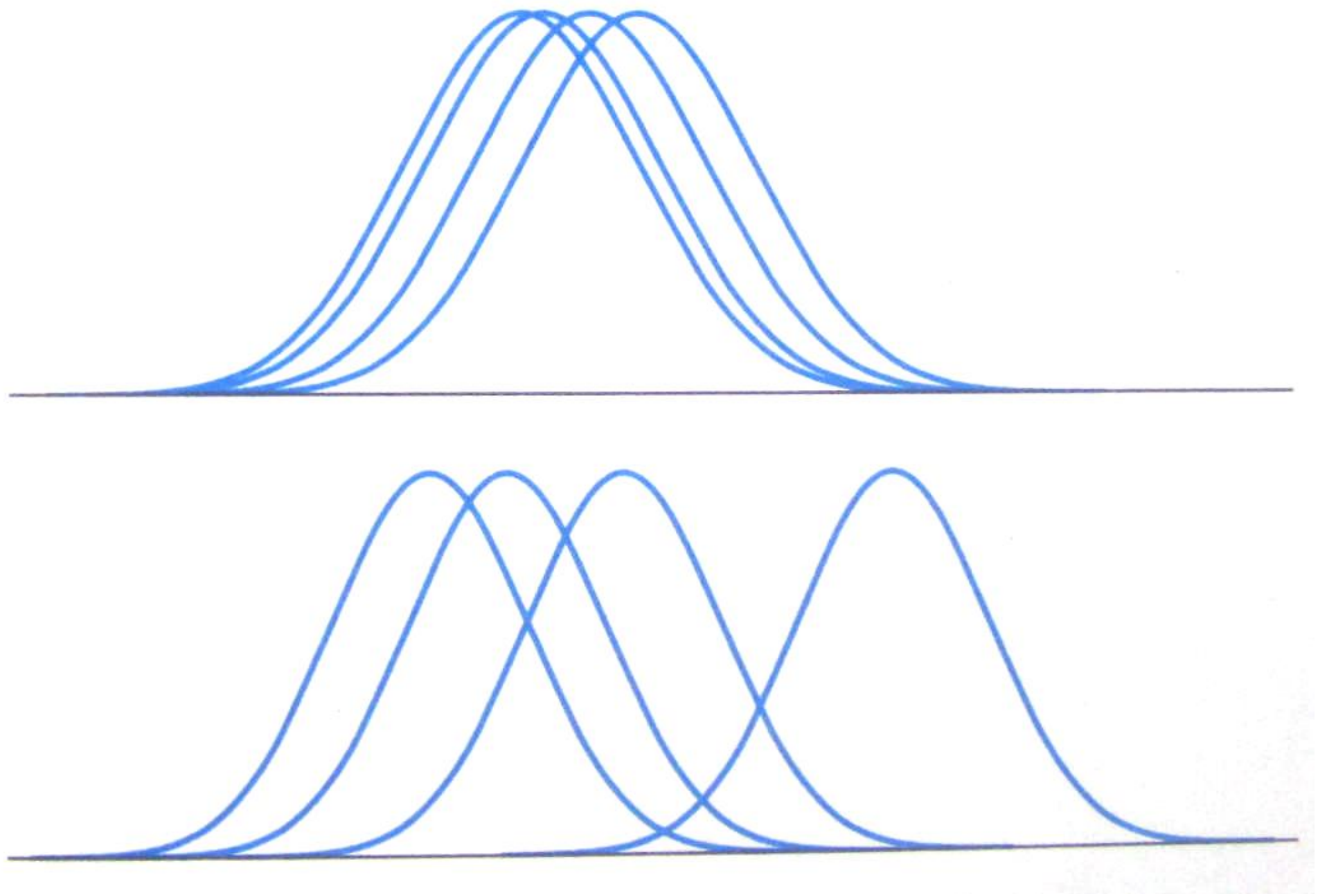
SS_T - общая сумма квадратов

S_x – стандартное отклонение

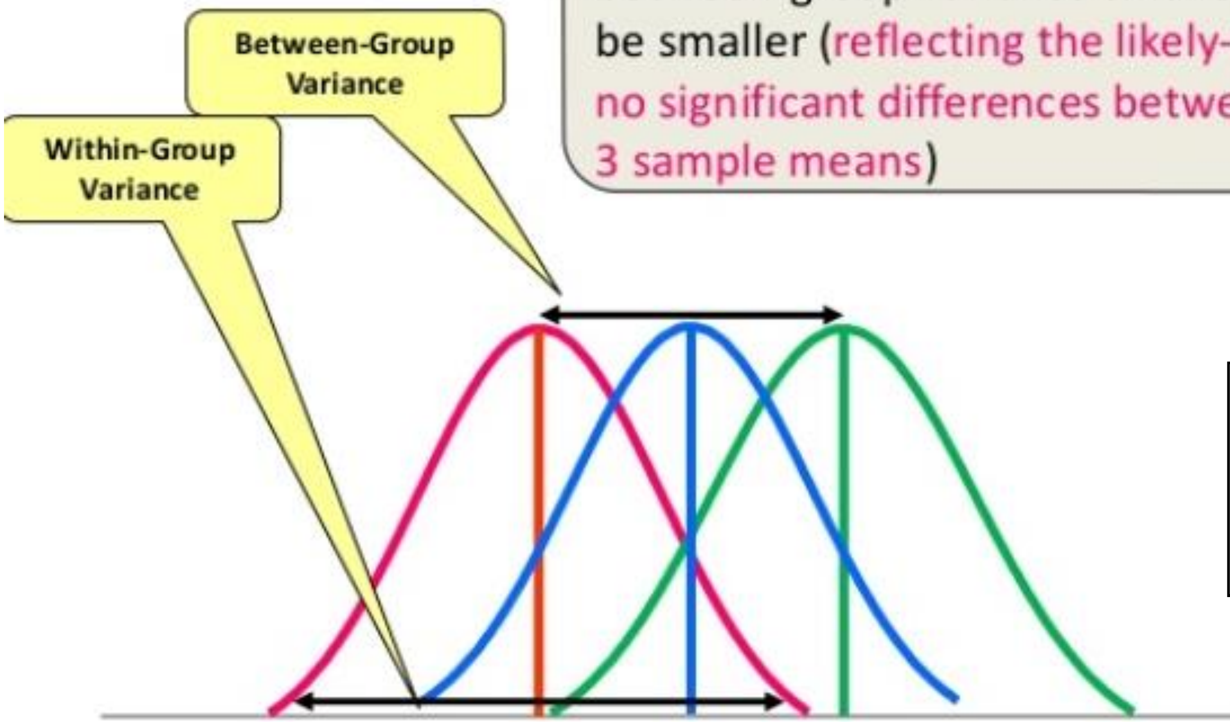
MS_W - средний квадрат отклонений внутри групп

Нет однозначных рекомендаций как считать размер эффекта для ANOVA

В каком случае значение F-статистики будет больше?

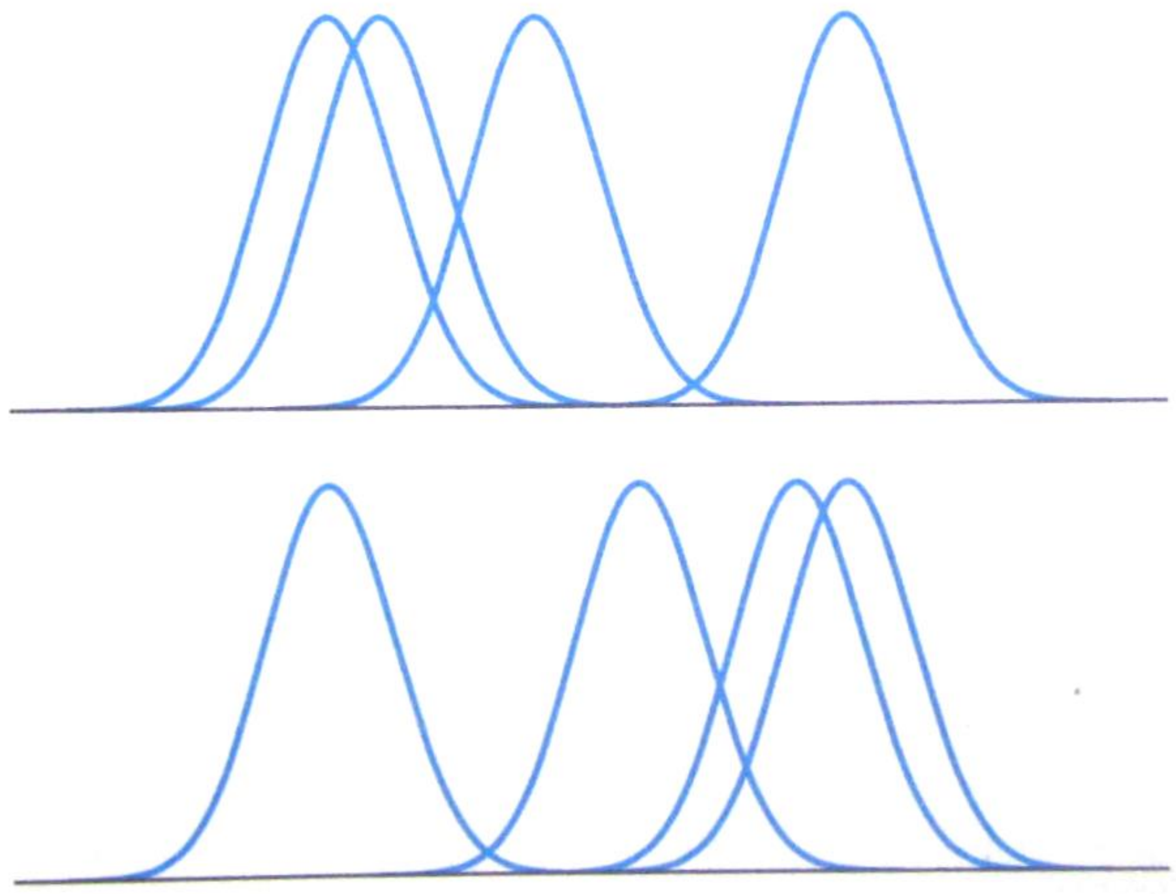


Within-group variance is larger, and the between-group variance smaller, so F will be smaller (reflecting the likely-hood of no significant differences between these 3 sample means)

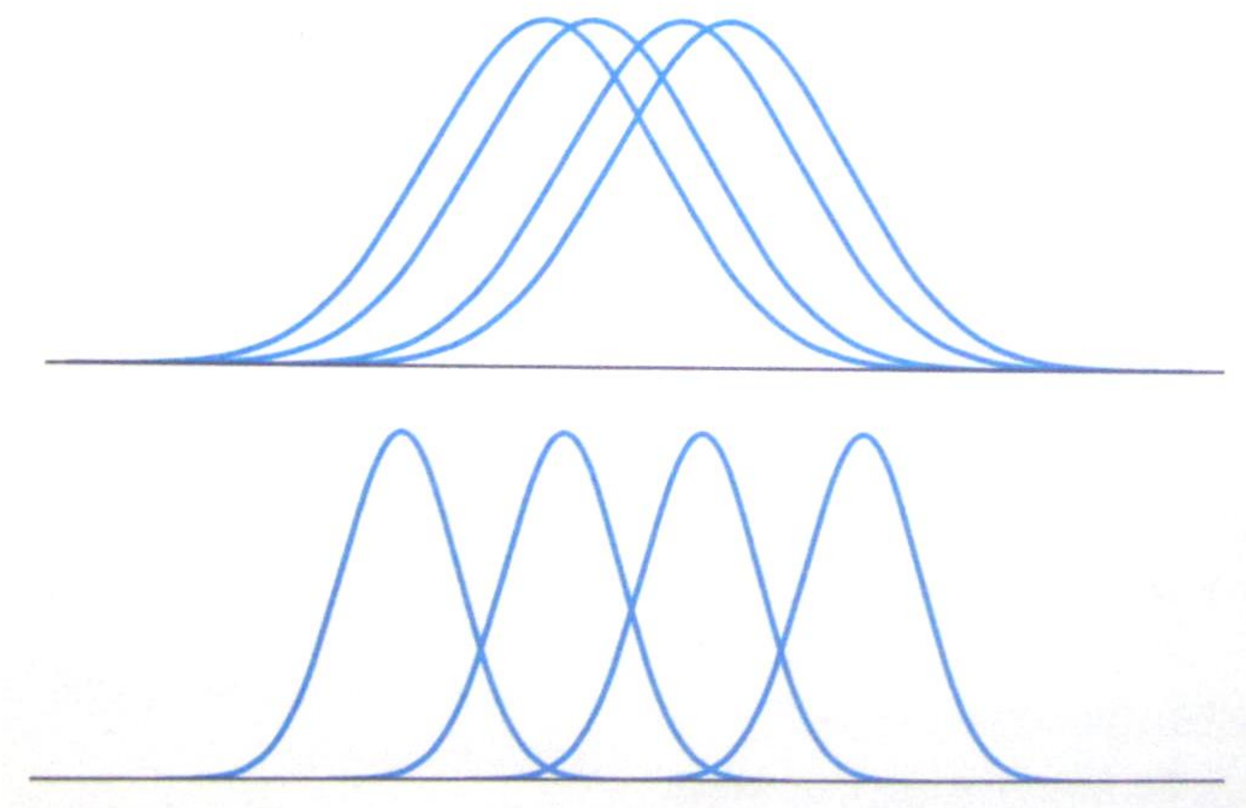


$$F = \frac{MS_B}{MS_W}$$

В каком случае значение F-статистики будет больше?

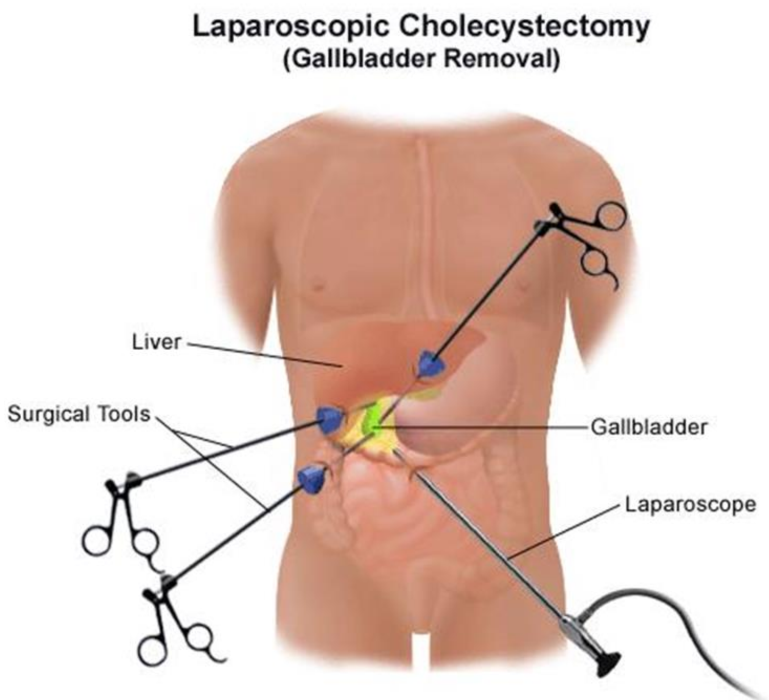


В каком случае значение F-статистики будет больше?



Пример 2 (для удобства восприятия на очень маленьких выборках)

В условиях клинической больницы было решено провести исследование по **оценке влияния возраста на длительность госпитализации после лапароскопической холецистэктомии**.



Лапароскопическая холецистэктомия (называемая также эндоскопическая холецистэктомия, или лапароскопия желчного пузыря) — это хирургическое вмешательство по **удалению желчного пузыря**, которое является наиболее эффективным методом лечения желчнокаменной болезни. Операция является малотравматичной, проводится эндоскопически, т.е. без больших разрезов. По статистике холецистэктомия — самая распространенная в мире лапароскопическая операция.

9 пациентов были разделены на 3 группы в зависимости от возраста:

Таблица

Длительность госпитализации после лапароскопической холецистектомии в зависимости от возраста, дни

Группа №1 Младше 45 лет	Группа №2 45-55 лет	Группа №3 Старше 55 лет
3	5	7
1	3	6
2	4	5
$\bar{x}=2$	$\bar{x}=4$	$\bar{x}=6$

Нужно сделать вывод о влиянии возраста на длительность госпитализации после операции

1. Постановка нулевой гипотезы

H_0 указывает на отсутствие различий между группами, иными словами все группы по возрасту относятся к одной генеральной совокупности и, соответственно, средние в них равны

$$\mu 1 = \mu 2 = \mu 3$$

H_1 Альтернативная гипотеза выдвигает предположение, что длительность госпитализации зависит от возраста, и средние в этих группах не равны

$$\mu 1 \neq \mu 2 \neq \mu 3$$

Группа №1 Младше 45 лет	Группа №2 45-55 лет	Группа №3 Старше 55 лет
3	5	7
1	3	6
2	4	5
$\bar{x}=2$	$\bar{x}=4$	$\bar{x}=6$

2. *Найдем <u>межгрупповую сумму квадратов</u>*

Для этого нам необходимо найти квадрат отклонения каждой из выборочных средних относительно общей средней:

$SS_B = 3(2-4)^2+3(4-4)^2+3(6-4)^2=24$ Sum of squares between groups

3. *Найдем <u>сумму квадратов внутри групп</u>*

последовательно вычитая из каждого значения в группе групповую среднюю:

$SS_W = (3-2)^2 + (1-2)^2 + (2-2)^2 + (5-4)^2 + (3-4)^2 + (4-4)^2 + (7-6)^2 + (6-6)^2 + (5-6)^2 =$

$2+2+2 = 6$ Sum of squares within groups

Группа №1 Младше 45 лет	Группа №2 45-55 лет	Группа №3 Старше 55 лет
3	5	7
1	3	6
2	4	5
$\bar{x}=2$	$\bar{x}=4$	$\bar{x}=6$

4. Найдем значение критерия Фишера,

исходя из **средних квадратов отклонений** внутри групп (MS_W)

и между ними (MS_B) и соответствующих **степеней свободы**:

$df_B = m - 1 = 3\text{-}1 = 2$

$df_W = n - m = 9 - 3 = 6$

$$MS_B = \frac{SS_B}{df_B}$$

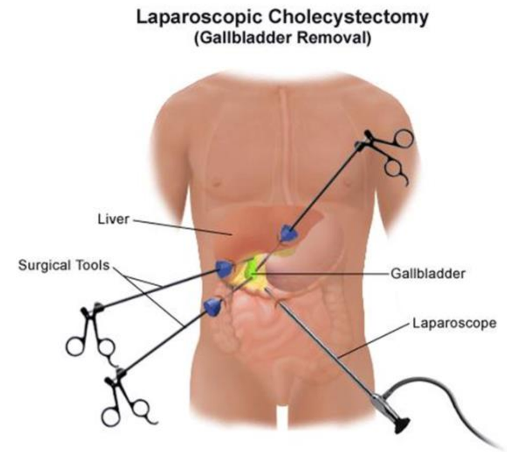
$$MS_W = \frac{SS_W}{df_W}$$

$$\boxed{F = \frac{MS_B}{MS_W}}$$

m – число групп

n – количество наблюдений в каждой из групп

Группа №1 Младше 45 лет	Группа №2 45-55 лет	Группа №3 Старше 55 лет
3	5	7
1	3	6
2	4	5
$\bar{x}=2$	$\bar{x}=4$	$\bar{x}=6$



F = 12, Fкрит. = 5,1 при **уровне значимости** **α = 0,05**

F > Fкрит.

Делаем вывод о наличии статистически значимых отличий между группами:
так как <u>наше значение F больше критического значения при заданном количестве наблюдений и количестве групп</u>

Возраст влияет на длительность госпитализации после холецистэктомии

Как найти **критическое** = **табличное** значение критерия F Фишера для сравнения ?

$$F = \frac{MS_B}{MS_W}$$

в таблице найти значение на пересечении числа степеней свободы для межгрупповой дисперсии и внутригрупповой дисперсии

$df_{BG} = m - 1 = 3\text{-}1 = \mathbf{2}$

$df_{WG} = n - m = 9 - 3 = \mathbf{6}$

Значения критерия F Фишера при уровне значимости $\alpha = 0.05$
(число степеней свободы указано для дисперсии знаменателя – в строке, для дисперсии числителя – в столбце)

df_2 \ df_1	1	2	3	4	5	6	7	8	9	10	15	20	30	∞
1	161	200	216	225	230	234	237	239	241	242	246	248	250	254
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.4
3	10.1	9.6	9.3	9.1	9.0	8.9	8.9	8.9	8.8	8.8	8.7	8.7	8.6	8.5
4	7.7	6.9	6.6	6.4	6.3	6.2	6.1	6.0	6.0	5.9	5.9	5.8	5.8	5.6
5	6.6	5.8	5.4	5.2	5.1	5.0	4.9	4.8	4.8	4.7	4.6	4.6	4.5	4.4
6	6.0	5.1	4.7	4.5	4.4	4.3	4.2	4.2	4.1	4.1	4.0	3.9	3.8	3.7
7	5.6	4.7	4.4	4.1	4.0	3.9	3.8	3.7	3.7	3.6	3.5	3.4	3.4	3.2
8	5.3	4.5	4.1	3.8	3.7	3.6	3.5	3.4	3.4	3.3	3.2	3.2	3.1	3.0
9	5.1	4.3	3.9	3.6	3.5	3.4	3.3	3.2	3.2	3.1	3.0	2.9	2.9	2.7
10	5.0	4.1	3.7	3.5	3.3	3.2	3.1	3.1	3.0	3.0	2.9	2.8	2.7	2.5
11	4.8	4.0	3.6	3.4	3.2	3.1	3.0	3.0	2.9	2.9	2.7	2.7	2.6	2.4
12	4.7	3.9	3.5	3.3	3.1	3.0	2.9	2.9	2.8	2.8	2.6	2.5	2.5	2.3
13	4.7	3.8	3.4	3.2	3.0	2.9	2.8	2.8	2.7	2.7	2.5	2.5	2.4	2.2
14	4.6	3.7	3.3	3.1	3.0	2.9	2.8	2.7	2.7	2.6	2.5	2.4	2.3	2.1
15	4.5	3.7	3.3	3.1	2.9	2.8	2.7	2.6	2.6	2.5	2.4	2.3	2.2	2.1
16	4.5	3.6	3.2	3.0	2.8	2.7	2.7	2.6	2.5	2.5	2.3	2.3	2.2	2.0
17	4.4	3.6	3.2	2.9	2.8	2.7	2.6	2.5	2.5	2.4	2.3	2.2	2.1	2.0
18	4.4	3.5	3.2	2.9	2.8	2.7	2.6	2.5	2.5	2.4	2.3	2.2	2.1	1.9
19	4.4	3.5	3.1	2.9	2.7	2.6	2.5	2.5	2.4	2.4	2.2	2.2	2.1	1.9
20	4.3	3.5	3.1	2.9	2.7	2.6	2.5	2.4	2.4	2.3	2.2	2.1	2.0	1.8
21	4.3	3.5	3.1	2.8	2.7	2.6	2.5	2.4	2.4	2.3	2.2	2.1	2.0	1.8
22	4.3	3.4	3.0	2.8	2.7	2.5	2.5	2.4	2.3	2.3	2.1	2.1	2.0	1.8
23	4.3	3.4	3.0	2.8	2.6	2.5	2.4	2.4	2.3	2.3	2.1	2.0	1.9	1.8
24	4.3	3.4	3.0	2.8	2.6	2.5	2.4	2.4	2.3	2.2	2.1	2.0	1.9	1.7
26	4.2	3.4	3.0	2.7	2.6	2.5	2.4	2.3	2.3	2.2	2.1	2.0	1.9	1.7
28	4.2	3.3	2.9	2.7	2.6	2.4	2.4	2.3	2.2	2.2	2.0	2.0	1.9	1.6
30	4.2	3.3	2.9	2.7	2.5	2.4	2.3	2.3	2.2	2.2	2.0	1.9	1.8	1.6
40	4.1	3.2	2.8	2.6	2.4	2.3	2.2	2.2	2.1	2.1	1.9	1.8	1.7	1.5
60	4.0	3.1	2.8	2.5	2.4	2.2	2.2	2.1	2.0	2.0	1.8	1.7	1.6	1.4
120	3.9	3.1	2.7	2.4	2.3	2.2	2.1	2.0	2.0	1.9	1.7	1.7	1.6	1.2
∞	3.8	3.0	2.6	2.4	2.2	2.1	2.0	1.9	1.9	1.8	1.7	1.6	1.5	1.0

однофакторный дисперсионный анализ

One-way ANOVA: assumptions (рекомендации и требования к выборкам)

1. Выборки должны быть случайными
2. Размеры выборок должны различаться как можно меньше (с увеличением разницы в размерах между выборками мощность теста резко падает);
3. Соответствие нормальному распределению
4. Равенство дисперсий в выборках
5. В выборках не должно быть очевидных аутлаеров (они сильно влияют на дисперсии)
6. Желательно, чтобы размер выборок был ≥ 10

N.B. Небольшое отклонение от какого-нибудь из требований компенсируется соблюдением остальных.

Как повысить мощность ANOVA:

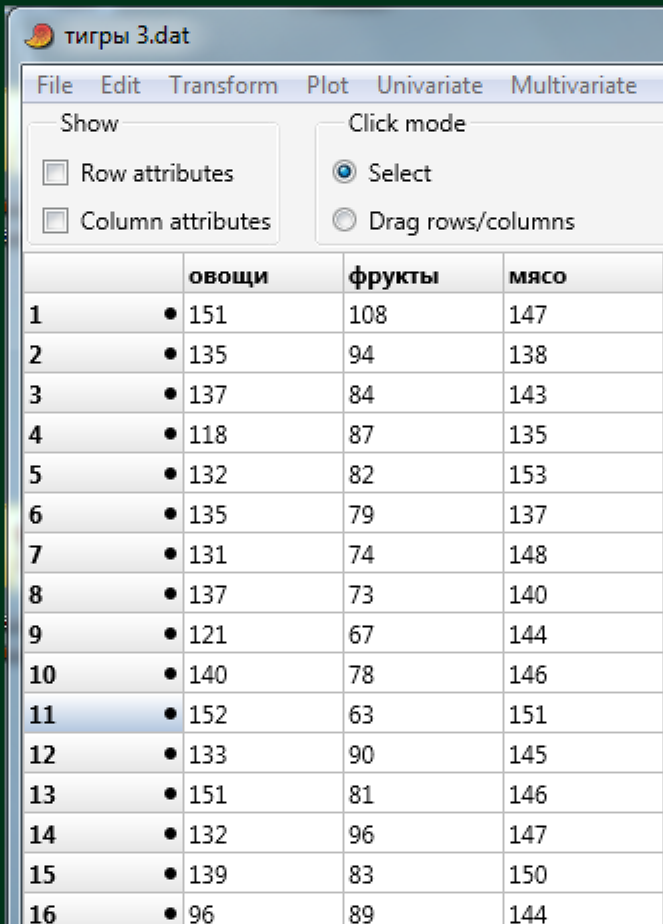
1. Увеличить размер выборок;
2. Уменьшить число групп;
3. Уменьшить внутригрупповую изменчивость.

Как и для двухвыборочного t-критерия, для ANOVA можно перед проведением эксперимента рассчитать:

✓размеры выборок для заданных мощности и размера эффекта;
✓мощность для выборок данного размера с конкретным размером эффекта.

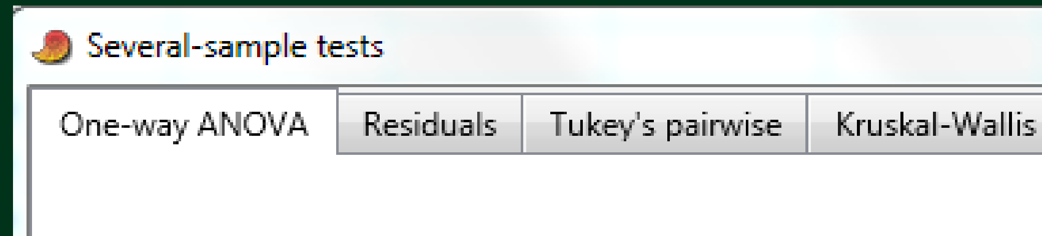
однофакторный дисперсионный анализ

One-way ANOVA

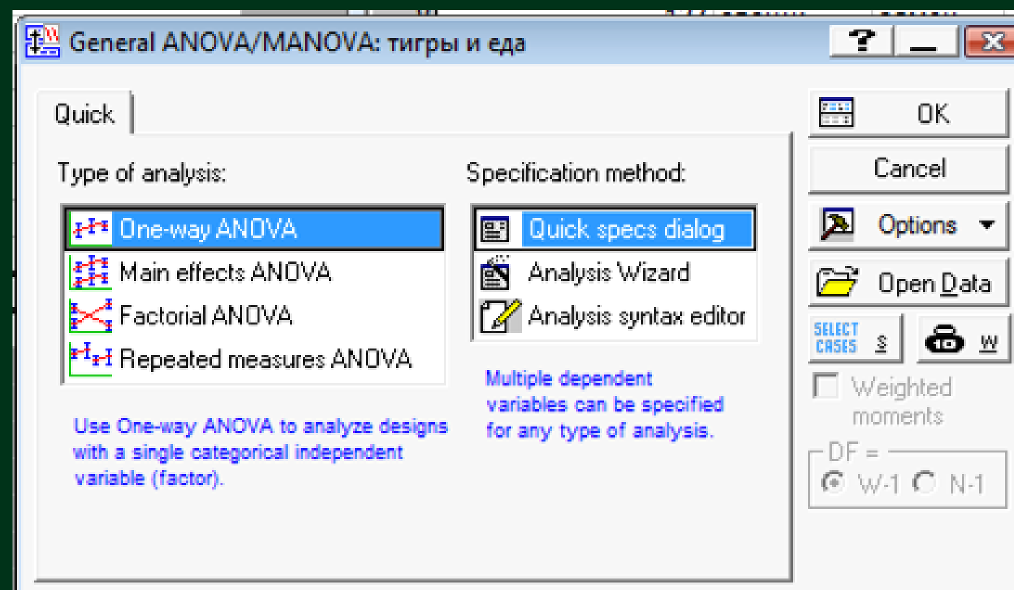


	овощи	фрукты	мясо
1 •	151	108	147
2 •	135	94	138
3 •	137	84	143
4 •	118	87	135
5 •	132	82	153
6 •	135	79	137
7 •	131	74	148
8 •	137	73	140
9 •	121	67	144
10 •	140	78	146
11 •	152	63	151
12 •	133	90	145
13 •	151	81	146
14 •	132	96	147
15 •	139	83	150
16 •	96	89	144

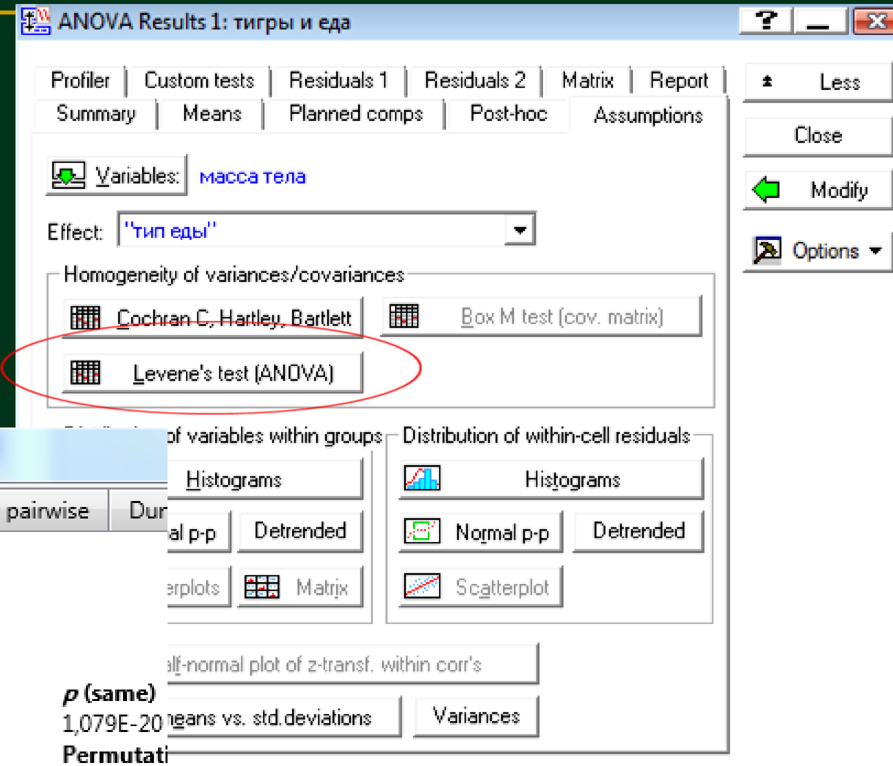
В «Пасте»

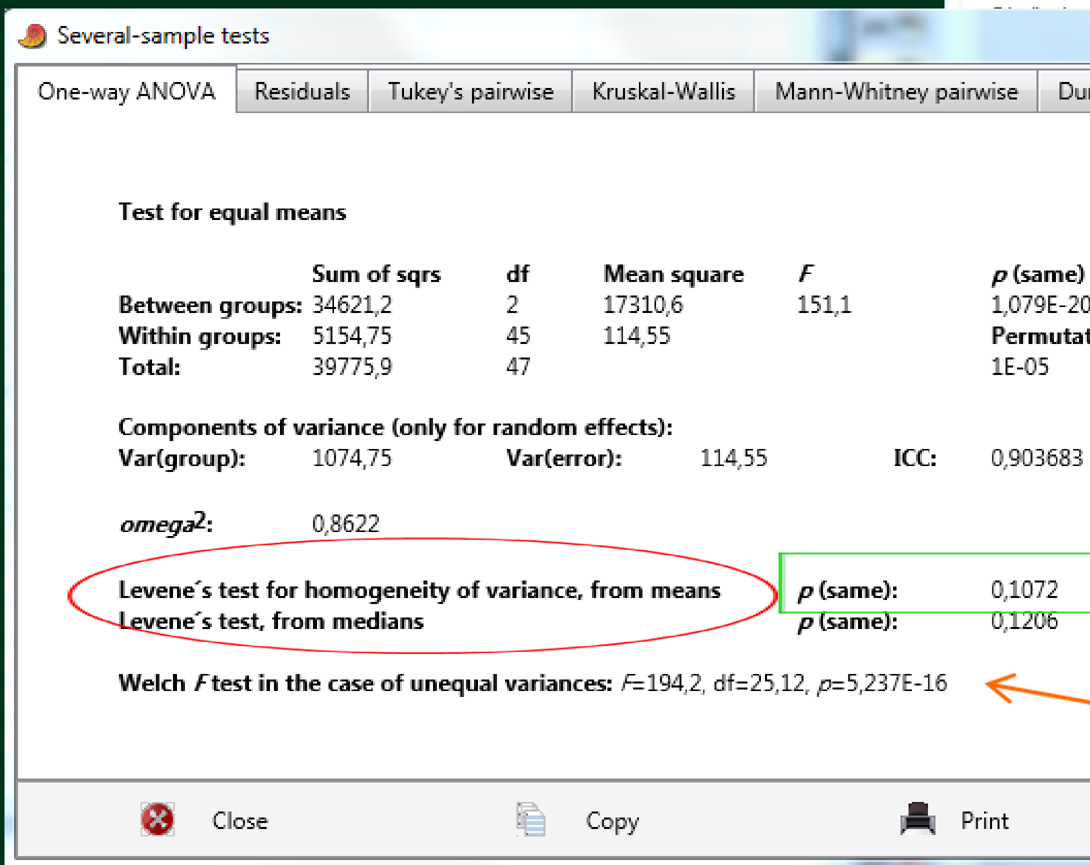


В «Статистике»



Проверяем на нормальность распределения данных и гомогенность дисперсий





Если дисперсии имеют значимые различия, используем результаты подхода Уэлча

One-way ANOVA

между группами

внутри групп

Several-sample tests

One-way ANOVA | Residuals | Tukey's pairwise | Kruskal-Wallis | Mann-Whitney pairwise | Dunn's post hoc

Test for equal means

	Sum of sqrs	df	Mean square	*F*	*p* (same)
Between groups:	34621,2	2	17310,6	151,1	1,079E-20
Within groups:	5154,75	45	114,55		**Permutation *p* (n=99999)**
Total:	39775,9	47			1E-05

Components of variance (only for random effects):

Var(group): 1074,75 **Var(error):** 114,55 **ICC:** 0,903683

***omega*2:** 0,8622

Levene´s test for homogeneity of variance, from means ***p*** **(same):** 0,1072

Levene´s test, from medians ***p*** **(same):** 0,1206

Welch *F* test in the case of unequal variances: *F*=194,2, df=25,12, *p*=5,237E-16

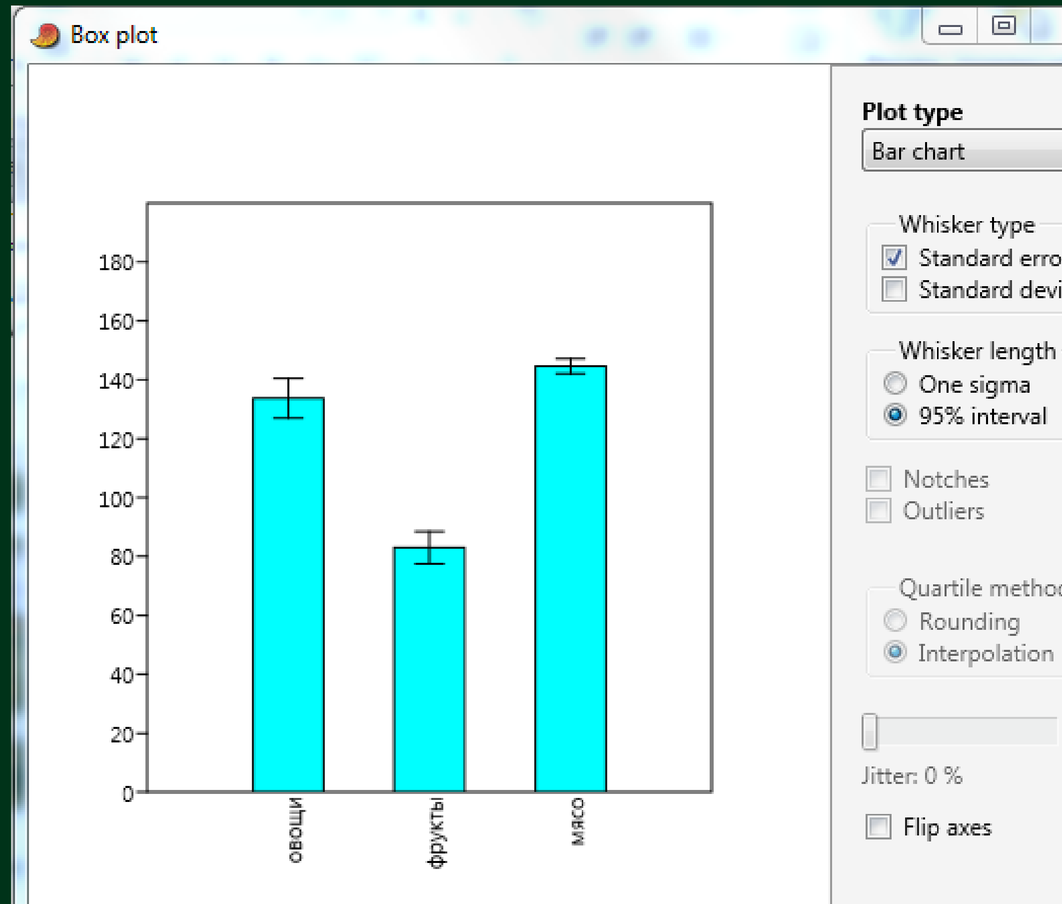
ignificance for масса тела (тигры и еда)

Univariate Tests of Significance for масса тела (тигры и еда)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	696490,1	1	696490,1	6080,228	0,00
тип еды	34621,2	2	17310,6	151,118	0,00
Error	5154,8	45	114,6		

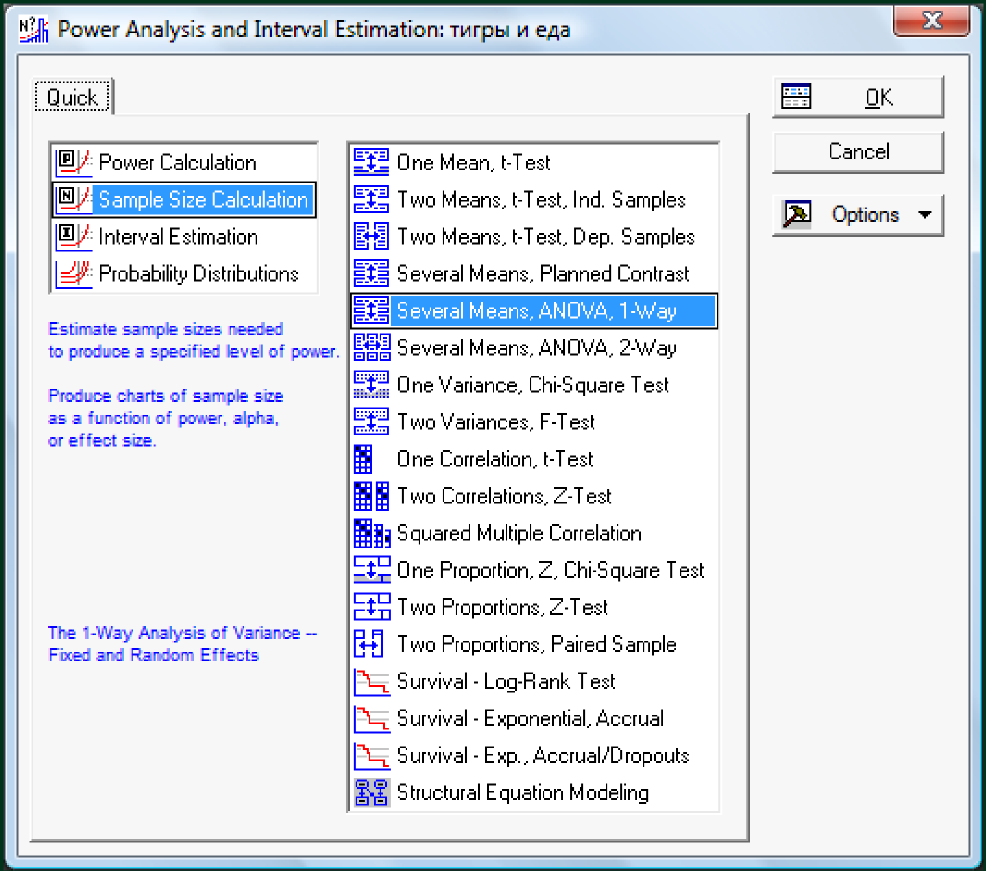
F	p (same)
151,1	1,079E-20

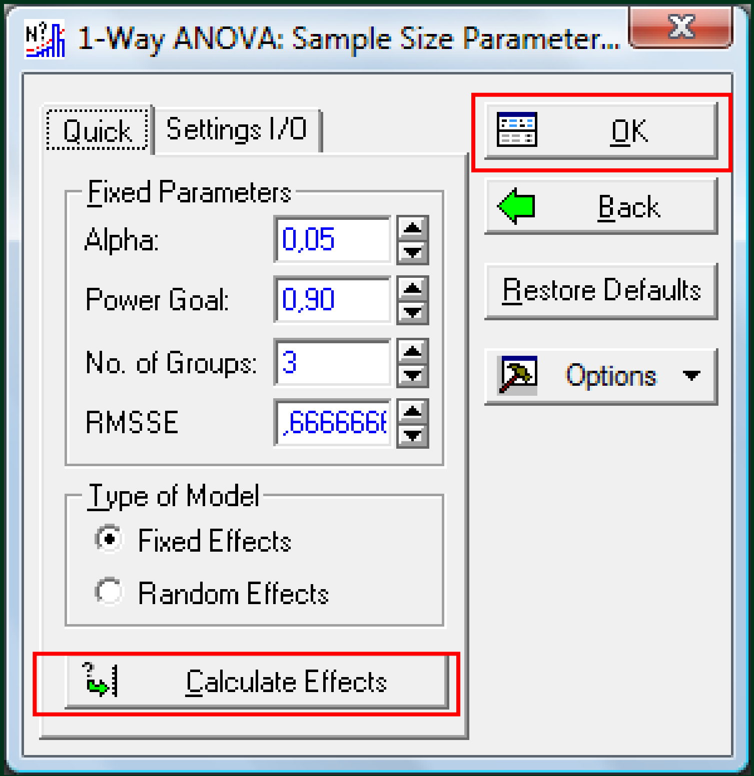
мы отвергаем H_0.
тип еды влиял на массу тигров

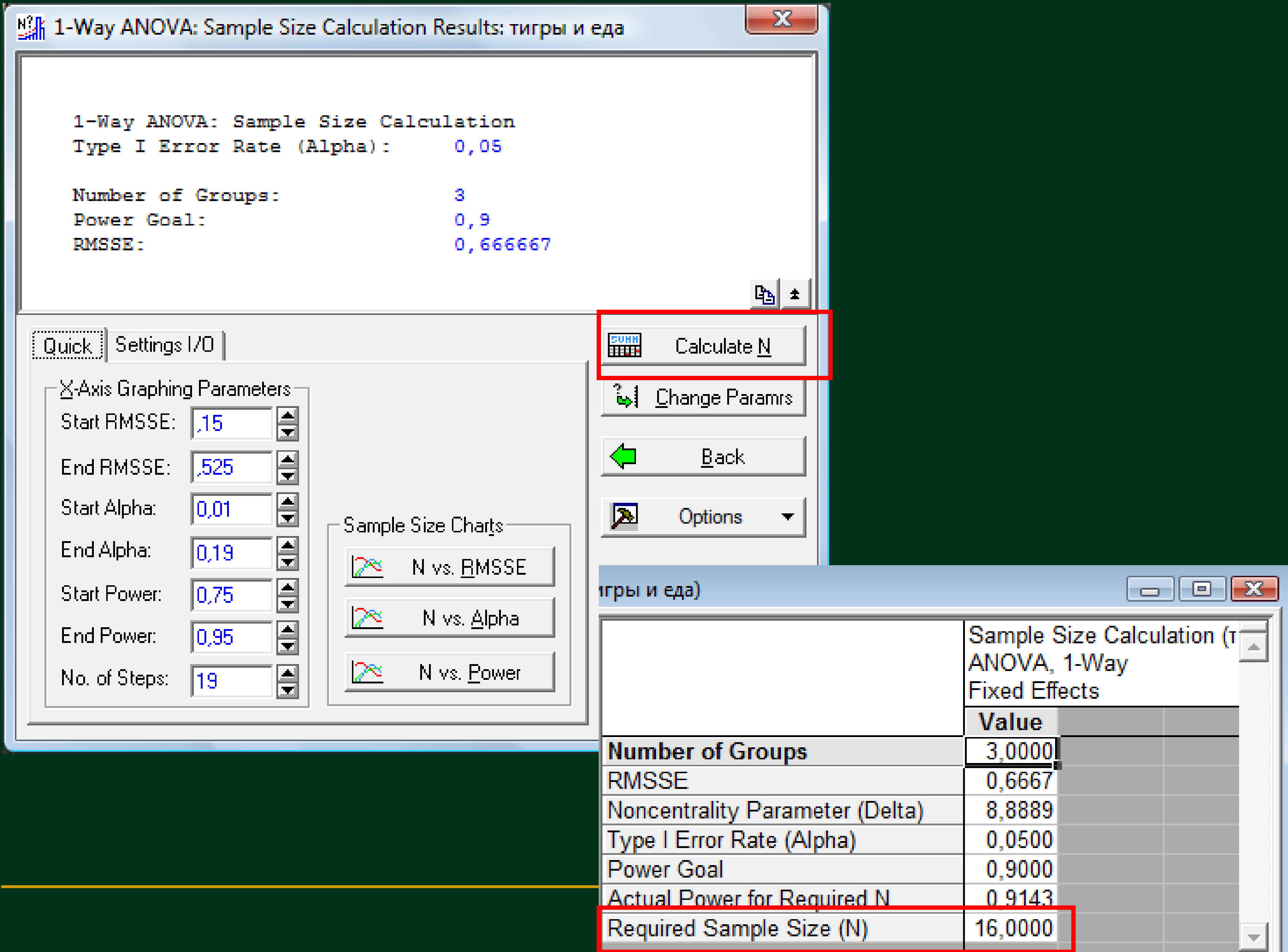


В «Статистике»

возможен Расчёт размера выборки (для каждой группы) при планировании исследования для 1-Way ANOVA






1-Way ANOVA: Sample Size Calculation Results: тигры и еда
1-Way ANOVA: Sample Size Calculation
Type I Error Rate (Alpha): 0,05
Number of Groups: 3
Power Goal: 0,9
RMSSE: 0,666667
Quick
Settings I/O
X-Axis Graphing Parameters
Start RMSSE: ,15
End RMSSE: ,525
Start Alpha: 0,01
End Alpha: 0,19
Start Power: 0,75
End Power: 0,95
No. of Steps: 19
Sample Size Charts
N vs. RMSSE
N vs. Alpha
N vs. Power
Calculate N
Change Params
Back
Options
игры и еда)
Sample Size Calculation (т
ANOVA, 1-Way
Fixed Effects
Value
Number of Groups 3,0000
RMSSE 0,6667
Noncentrality Parameter (Delta) 8,8889
Type I Error Rate (Alpha) 0,0500
Power Goal 0,9000
Actual Power for Required N 0,9143
Required Sample Size (N) 16,0000

На всякий случай:

Возможно провести one-way ANOVA в случае, если у нас в руках есть только средние значения, показатели разброса (SD, SE, s^2) и размер выборок (например, из какой-нибудь статьи)

Поскольку для каждой группы $s^2 = SD^2 = n(SE)^2$, для k групп

$$SS_w = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

$$df_W = n_G - k$$

$$MS_W = \frac{SS_W}{df_W}$$

$$SS_B = \sum_{i=1}^{k} n_i \overline{X}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \overline{X}_i\right)^2}{\sum_{i=1}^{k} n_i}$$

$$df_B = k-1$$

$$MS_B = \frac{SS_B}{df_B}$$

$$F = \frac{MS_B}{MS_W}$$

ANOVA post hoc tests

Сложная «омнибусная» гипотеза АНОВЫ:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = ... = \mu_k$$

Похожа на стрельбу из дробовика: не нужно особенно точно целиться, **НО** непонятно, какая дробинка попала в какую мишень!

<u>Какая же из отдельных гипотез не верна?</u>

Ответить поможет апостериорный (post hoc) тест!

Если у нас 3 и более групп:

1. Сначала сравнить ВСЕ группы между собой с помощью ANOVA
2. Если различия есть, использовать методы множественного сравнения (группы сравнивают попарно)
3. Если различий нет, мы НЕ ИМЕЕМ ПРАВА ПРЕДПРИНИМАТЬ ДАЛЬНЕЙШИЙ АНАЛИЗ!

Двухвыборочный t-критерий для сравнения групп попарно после проведения ANOVA тоже не годится!

Например, если мы сравним две крайние группы, это уже будут не случайные выборки из генеральной совокупности, и уровень значимости α уже будет не 0.05!

Тест Тьюки (Tukey HSD test)

Наиболее распространённый и рекомендуемый в литературе тест (Hurlburt, 2006; Zar, 2010).

Рекомендуется для близких по размеру групп.

Проверяет только ПАРНЫЕ (но не комплексные) гипотезы.

$$
\begin{aligned}
H_{01} &: \mu_1 = \mu_2 \\
H_{02} &: \mu_1 = \mu_4 \\
H_{03} &: \mu_1 = \mu_3 \\
&\ldots
\end{aligned}
$$

?

Другие апостериорные тесты

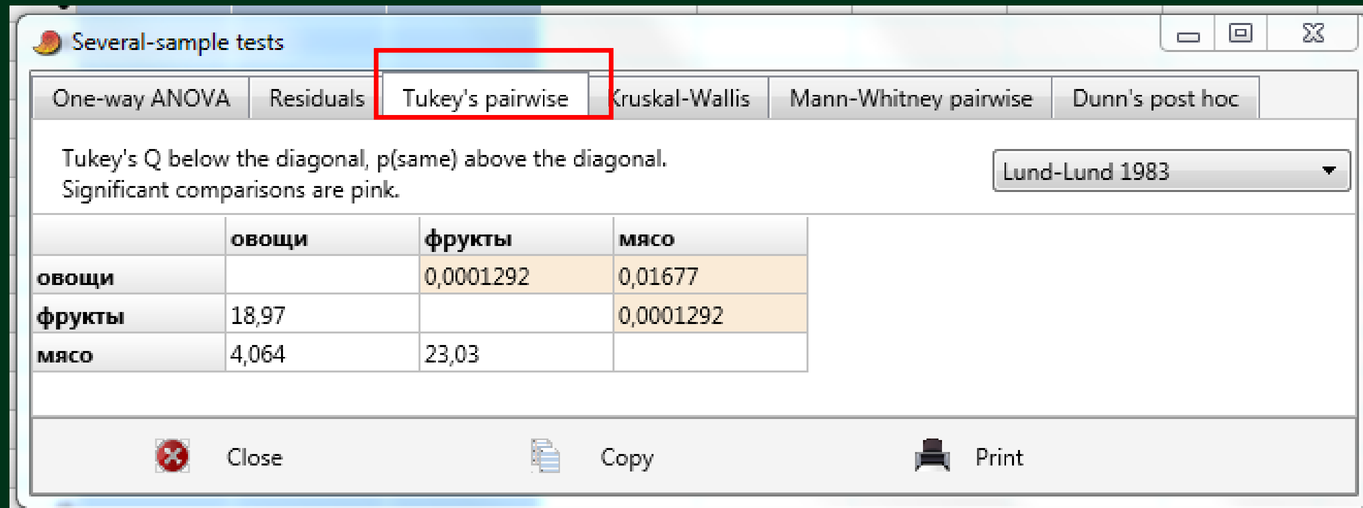
1. Критерий Ньюмена-Кейлса (*Newman-Keuls test*) - наименее строгий. Все средние упорядочивают по возрастанию и вычисляют критерий; начинают от сравнения наибольшего с наименьшим.
2. Критерий Шеффе (Scheffe test) – проверяет не только парные гипотезы, но и комплексные.
3. Критерий Даннетта (*Dunnett test*) – используется для сравнения нескольких групп с контрольной группой.

…

Бывает так, что в ANOVA нулевая гипотеза отвергается, а пост-хок тесты не обнаруживают различий, так как их мощность ниже. В этом случае необходимо увеличивать размер выборки.

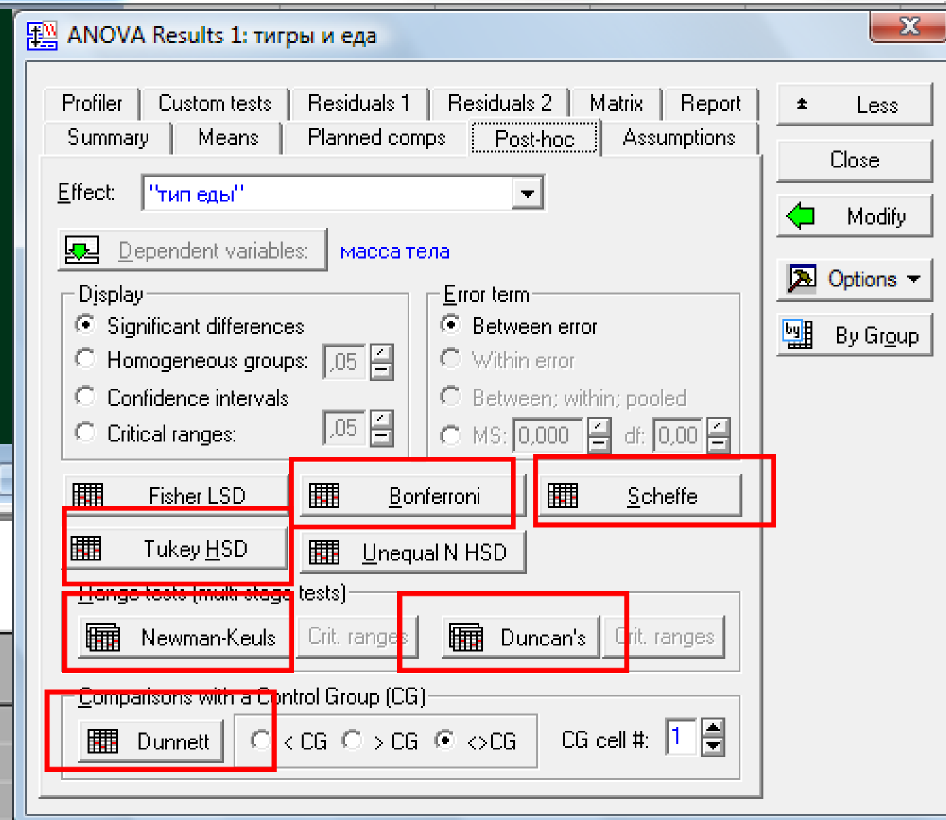
апостериорные тесты

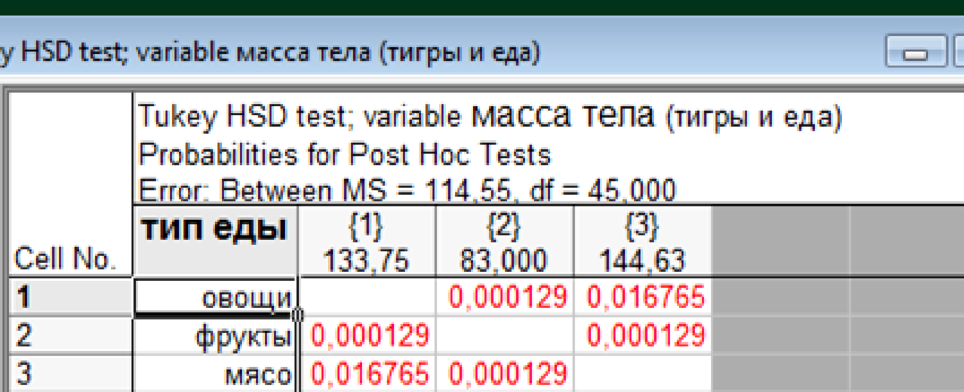
В «Пасте»



Апостериорные сравнения методом Тьюки показали, что статистически значимые различия в дисперсионном анализе связаны с различиями средних масс тигров во всех трех группах

В «Статистике»

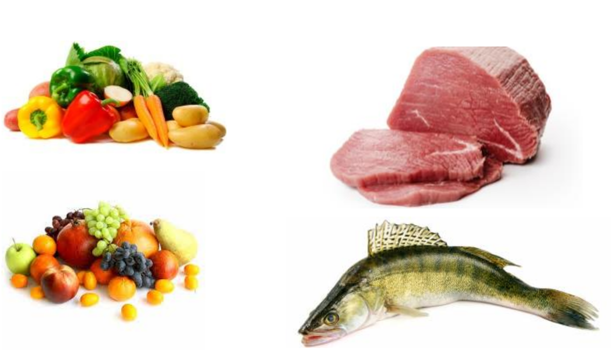




Factorial ANOVA

Многофакторный дисперсионный анализ

При однофакторном дисперсионном анализе **One-way ANOVA:**



Одна зависимая переменная, variable (масса тела);

Одна группирующая = фактор (тип пищи).

Одна нулевая гипотеза $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Что делать, если нужно проанализировать влияние двух (трёх и т.д.) факторов на зависимую переменную?

Например,
Мы изучаем влияние размножения на массу тела у самок африканских земляных белок разного возраста.
Зависимая переменная – масса тела.
Фактор А – наличие выводка (1. есть; 2. нет)
Фактор В – возраст (1 год, 2 года, 3 года и старше).

Фактора ДВА, наш выбор - two-way ANOVA

Двухфакторный дисперсионный анализ

	1 год	2 года	≥3 года
без выводка	440 438 429 502 602	892 868 855 866 932	1575 849 759 1602 1327
с выводком	308 328 326 326 325	737 798,5 876 810 861	1000,5 901 958 1032 883

Получилось $a \times b = 2 \times 3 = 6$ групп белок – 6 ячеек (cells) в таблице.

Заметим, что во ВСЕХ ячейках должны выполняться условия соответствия нормальному распределению и равенства дисперсий.

Пусть в каждой ячейке по n наблюдений.

	1 год	2 года	≥3 года
без выводка	440 438 429 502 602	892 868 855 866 932	1575 849 759 1602 1327
с выводком	308 328 326 326 325	737 798,5 876 810 861	1000,5 901 958 1032 883

Формулируем 3 нулевые гипотезы (и 3 альтернативные):

H_0: наличие выводка не влияет на массу самки ($\mu_{\text{б/выводка}} = \mu_{\text{с выводком}}$)
H_0: возраст самки не влияет на массу самки ($\mu_1 = \mu_2 = \mu_{\geq 3}$)
H_0: нет взаимодействия между факторами.

$$MS_{factorA} = \frac{SS_{factorA}}{df_{factorA}} = \frac{bn\sum_{i=1}^{a}(\bar{X}_i - \bar{X}_G)^2}{a-1}$$

$$F_{factorA} = \frac{MS_{factorA}}{MS_{error}}$$

$$MS_{factorB} = \frac{SS_{factorB}}{df_{factorB}} = \frac{an\sum_{j=1}^{b}(\bar{X}_j - \bar{X}_G)^2}{b-1}$$

$$F_{factorB} = \frac{MS_{factorB}}{MS_{error}}$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\left[\sum_{l=1}^{n}(X_{ijl} - \bar{X}_{ij})^2\right]}{N-ab}$$

Дисперсия между ячейками не равна сумме изменчивостей между уровнями фактора А и уровнями фактора В. Эта разница определяется взаимодействием факторов:

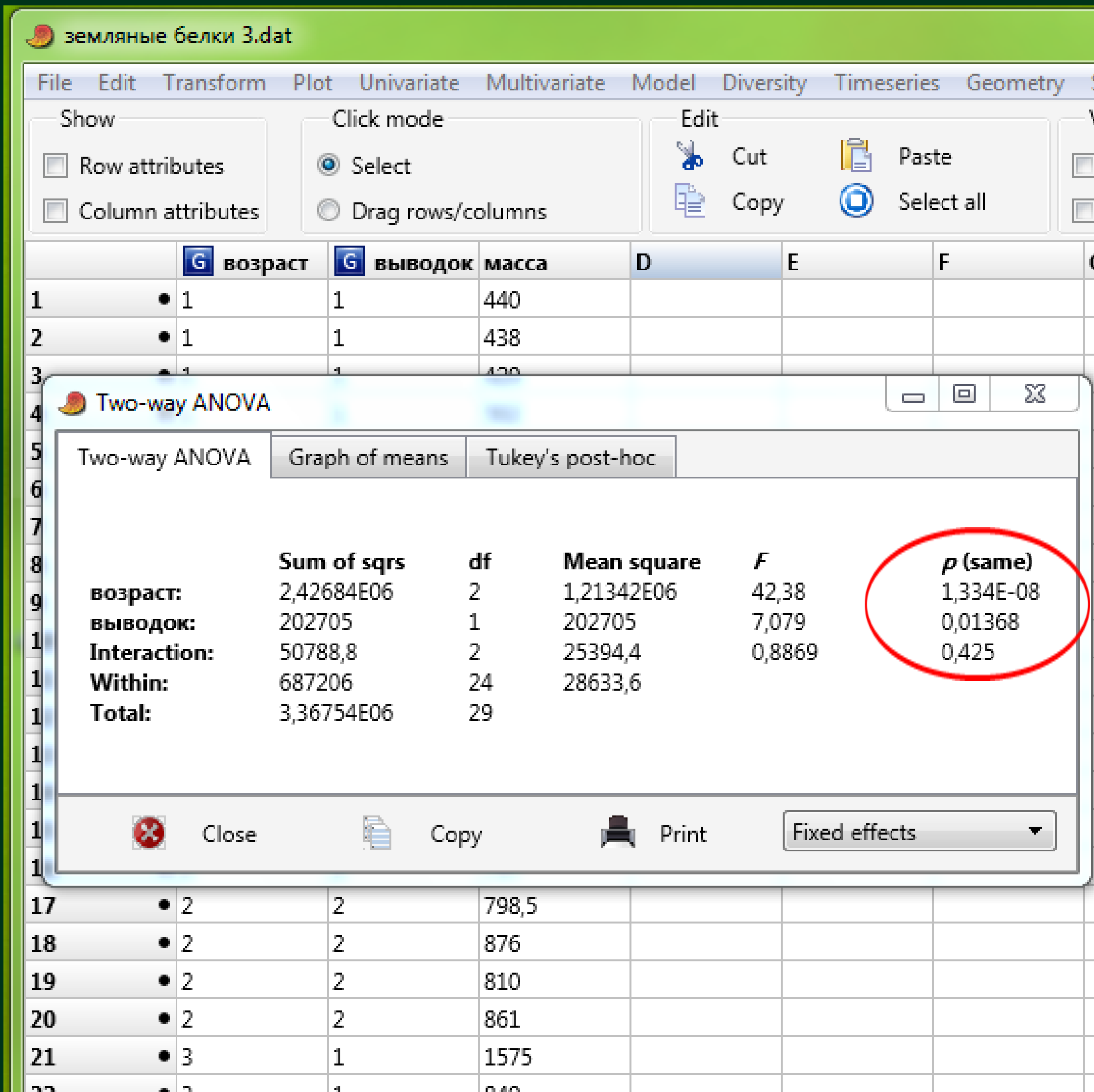
$$MS_{AB\,\mathrm{interaction}} = \frac{SS_{AB\,\mathrm{interaction}}}{df_{AB\,\mathrm{interaction}}} = \frac{SS_{cells} - SS_{factorA} - SS_{factorB}}{df_{cells} - df_{factorA} - df_{factorB}}$$

$df_{interaction} = df_{factorA} \times df_{factorB}$

$$\boxed{F_{AB\,\mathrm{interaction}} = \frac{MS_{AB\,\mathrm{interaction}}}{MS_{error}}}$$

$$SS_{cells} = \sum_{i=1}^{a}\sum_{j=1}^{b} n(\bar{X}_{ij} - \bar{X}_G)^2 \qquad df_{cells} = ab - 1$$

Factorial ANOVA
африканские земляные белки



первые две нулевых гипотезы мы отвергаем: и возраст, и наличие выводка влияют на массу белок,
третью не отвергаем: взаимодействия факторов НЕТ

То есть, для каждой гипотезы мы рассчитываем своё F-значение и сравниваем его со своим критическим уровнем.

Изменчивость между строками

$$F_{R_{obs}} = \frac{MS_R}{MS_W}$$

MS_{error}, средняя по ячейкам внутригрупповая изменчивость

$$F_{C_{obs}} = \frac{MS_C}{MS_W}$$

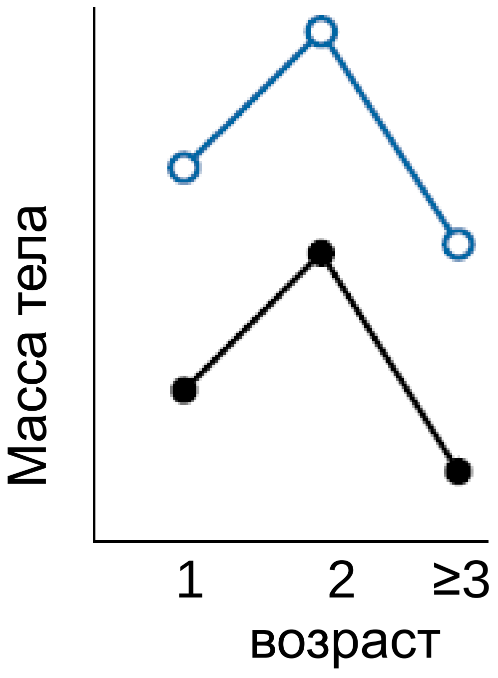
Изменчивость между столбцами

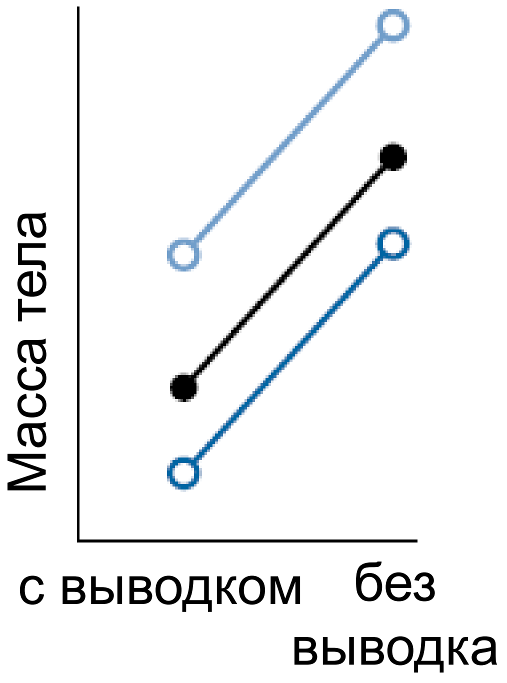
$$F_{RC_{obs}} = \frac{MS_{RC}}{MS_W}$$

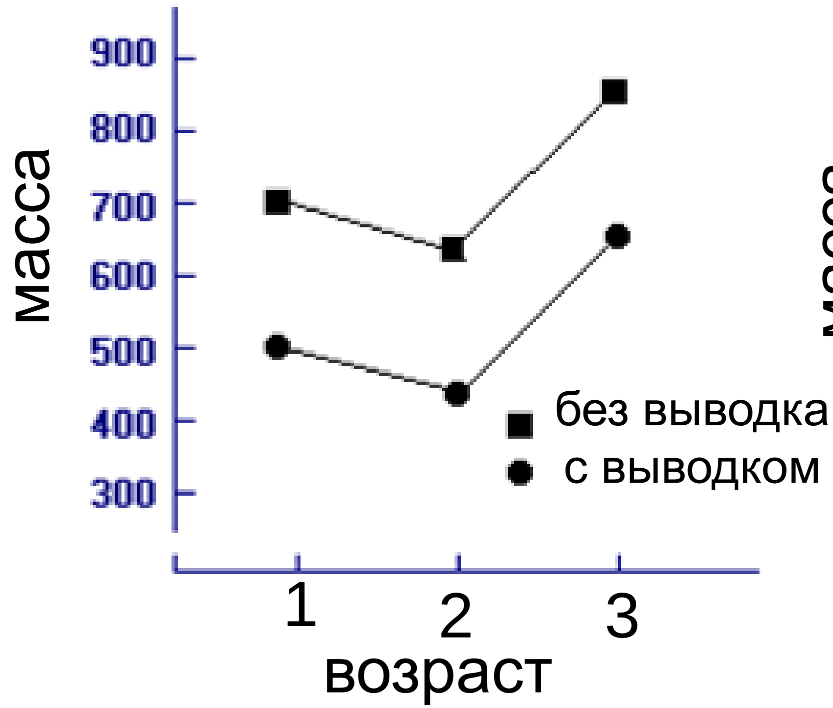
«взаимодействие» факторов

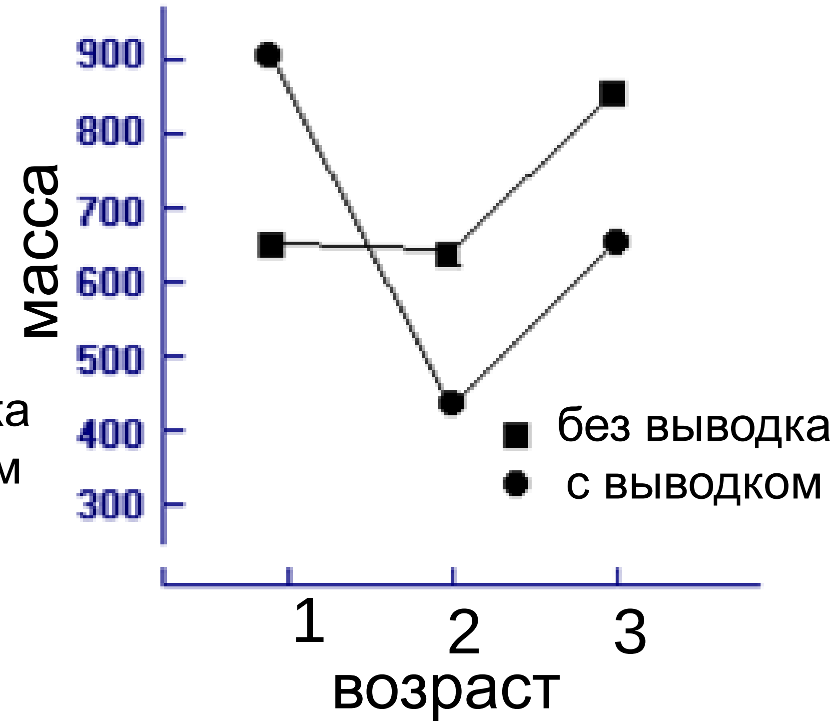
Достоверное взаимодействие факторов говорит о том, что различия между уровнями одного из факторов неодинаковы для всех уровней другого фактора.

Примерный вид графического представления:









и размножение, и возраст влияют на массу;
взаимодействия факторов НЕТ

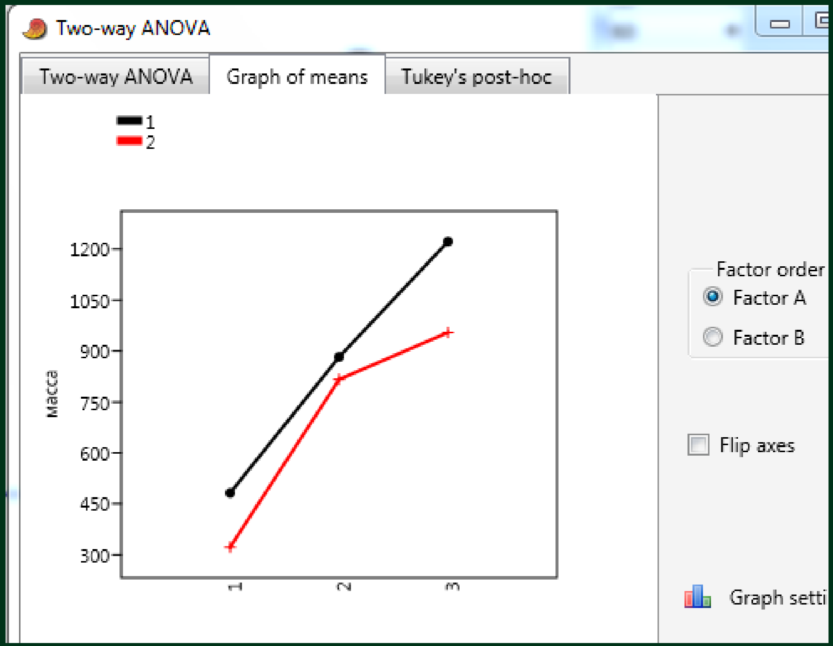
возраст влияет на массу, размножение – нет;
взаимодействие ЕСТЬ

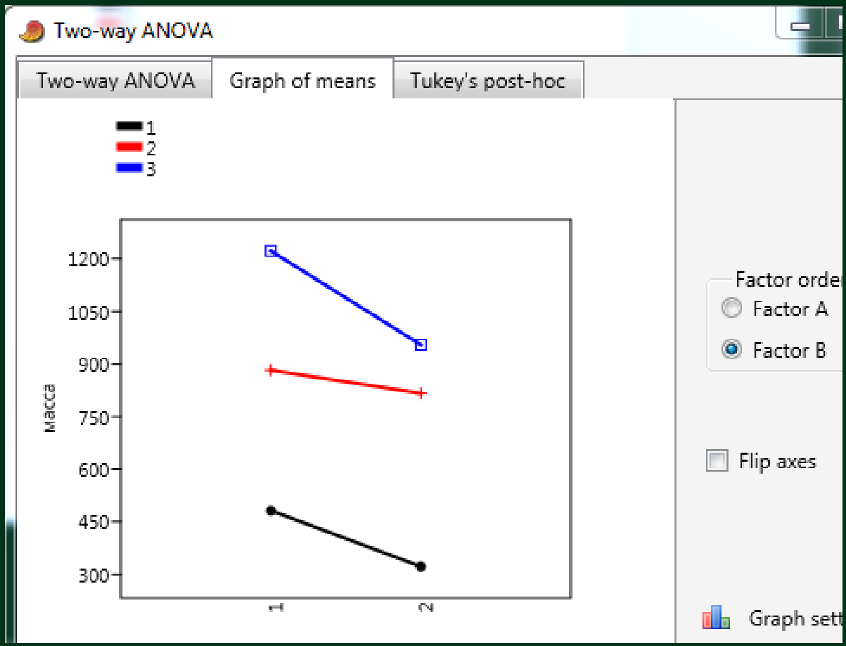
если линии на рисунке ПАРАЛЛЕЛЬНЫ, взаимодействия факторов НЕТ.
если НЕ ПАРАЛЛЕЛЬНЫ, взаимодействие ЕСТЬ.

(насколько они параллельны, решает ANOVA)

В нашем случае:

Factorial ANOVA
африканские земляные белки

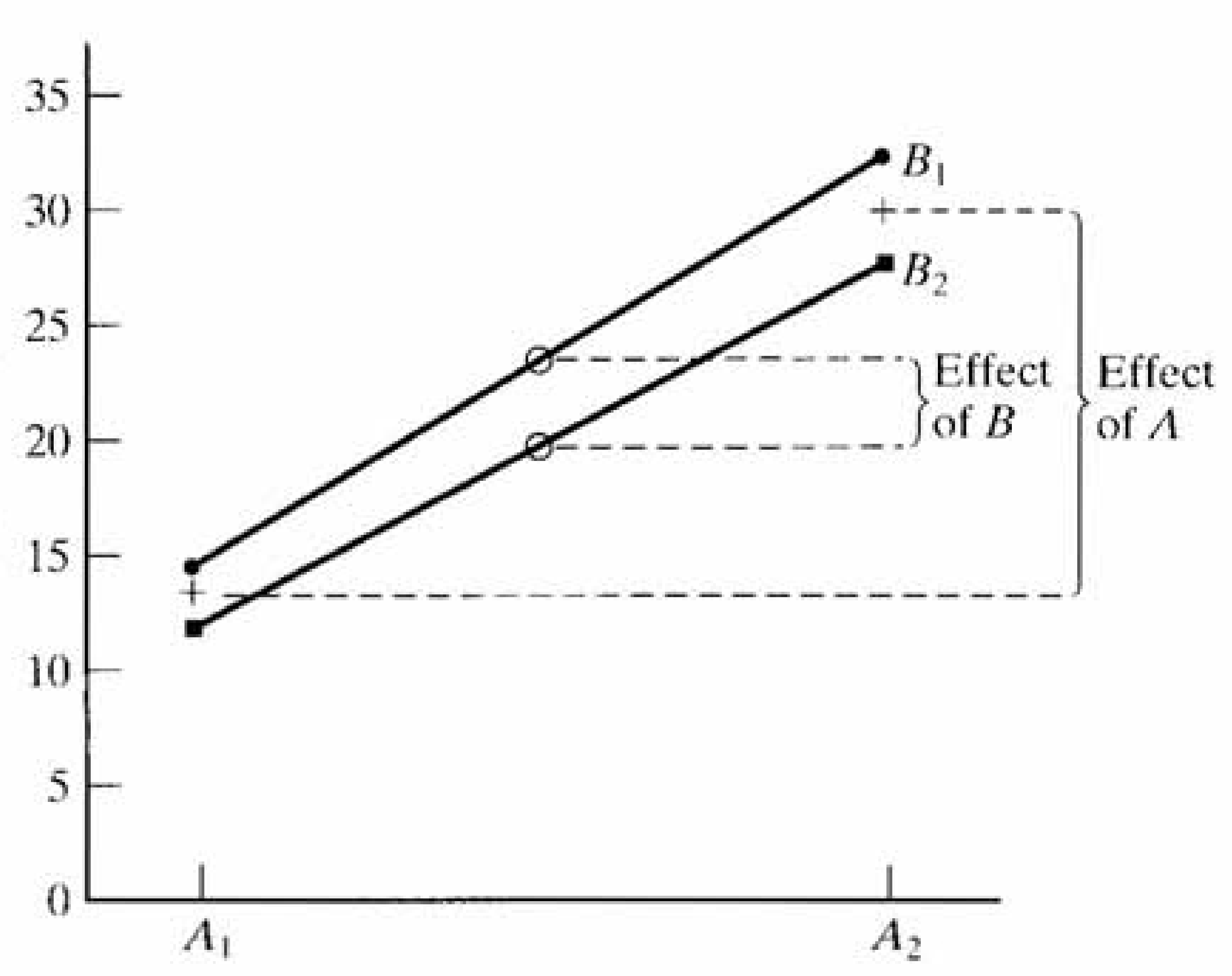




Черная линия – отсутствие выводка
Красная линия – наличие выводка

Черная линия – возраст 1 год
Красная линия – возраст 2 года
Синяя линия – возраст 3 и больше года

и размножение, и возраст влияют на массу;
взаимодействия факторов НЕТ



Как определить на глаз влияние каждого из факторов

Взаимодействие между факторами ≠ корреляция между факторами!!!

	1 год	2 года	≥3 года
без выводка	900	1200	1500
с 1-м выводком	600	900	1200
с 2-мя выводками	300	600	900

Взаимодействия факторов НЕТ, при этом между ними есть корреляция

	1 год	2 года	≥3 года
без выводка	900	1200	600
с 1-м выводком	600	900	1200
с 2-мя выводками	300	600	1200

Взаимодействие факторов ЕСТЬ

У старых самок участие в размножении по-другому сказывается на физическом состоянии, чем у молодых

Если факторов не 2 а много, а зависимая переменная ОДНА, анализ называется

Multiway ANOVA Многофакторный дисперсионный анализ

В этом случае становится много гипотез о взаимодействии факторов (для 3-х факторов 4 гипотезы об их взаимодействии).
Не рекомендуется исследовать действие более 4-х факторов, так как затрудняется интерпретация результатов.

Расчёт статистик в таких сложных случаях производится с использованием принципов регрессионного анализа.

Repeated measures ANOVA

Сравнение связанных групп

Преподаватель решил узнать, как у его студентов продолжительность занятий зависит от дня недели (он поделил время на 15-минутные блоки).

	Time blocks (15-minute periods) spent studying				
Person	*Monday*	*Tuesday*	*Wednesday*	*Thursday*	*Person mean*
Pat	15	10	8	7	10
Bobby	10	11	4	7	8
Riki	4	9	7	0	5
Jean	10	10	7	1	7
Lynn	4	2	4	2	3
Jo	12	17	9	14	13
Means	9.167	9.833	6.500	5.167	

Представим, что эти группы независимы и проведём ANOVA. Различия между ними недостоверны. Почему?
Из-за большой внутригрупповой изменчивости?

Студенты по усердию **сильно различаются между собой**!\

Как элиминировать межиндивидуальные различия (between-subjects effect)?

Вычесть из каждого измерения среднее значение для каждого студента!

Person	Monday	Tuesday	Wednesday	Thursday	Mean
	Deviations from person's own mean				
Pat	5	0	−2	−3	0
Bobby	2	3	−4	−1	0
Riki	−1	4	2	−5	0
Jean	3	3	0	−6	0
Lynn	1	−1	1	−1	0
Jo	−1	4	−4	1	0
Mean	1.500	2.167	−1.167	−2.500	

Вот теперь измерения стали независимы («исправленные»), и дальше можно сравнить их ANOVA

(от обычной ANOVA отличается число степеней свободы внутри измерений – $df_w = (k-1)(n-1)$)

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1: нулевая гипотеза не верна

Обычная ANOVA:

$$F = \frac{\text{оценка дисперсии } \textbf{между группами}}{\text{оценка дисперсии } \textbf{внутри} \text{ групп}} \qquad F = \frac{MS_B}{MS_W}$$

Repeated measures ANOVA:

$$F = \frac{\text{оценка дисперсии } \textbf{между измерениями}}{\textbf{«ошибка»} \text{ внутри } \textbf{исправленных} \text{ измерений}} \qquad F = \frac{MS_B}{MS_{err}}$$

Изменчивость:

1. Между измерениями;
2. Между особями (получается из средних значений для особей);
3. «ошибка» (внутри «исправленных» измерений) – error, residual

Теперь H_0 будет отвергнута, т.е., преподаватель сможет утверждать, что усердие его учеников зависит от дня недели.

Мощность дисперсионного анализа для повторных измерений выше, чем обыкновенного дисперсионного анализа (в случае связанных выборок).

Другой пример: к тиграм-самцам пришёл новый служитель, а потом – новая уборщица. И возможно, они стали по-другому питаться. Мы хотим узнать, менялась ли их масса.

Мы анализируем влияние служителя и уборщицы на массу тигров-самцов. *Зависимая переменная* – масса.
Для каждой особи по 3 измерения (3 столбика в таблице).

Каждый тигр ТРИ раза участвует в наблюдениях.

	ДО	СЛУЖ	УБОР
1 тигр	567	589	579
2 тигр	653	645	630
3 тигр	465	498	486
4 тигр	547	567	555
5 тигр	564	598	588
6 тигр	457	438	440

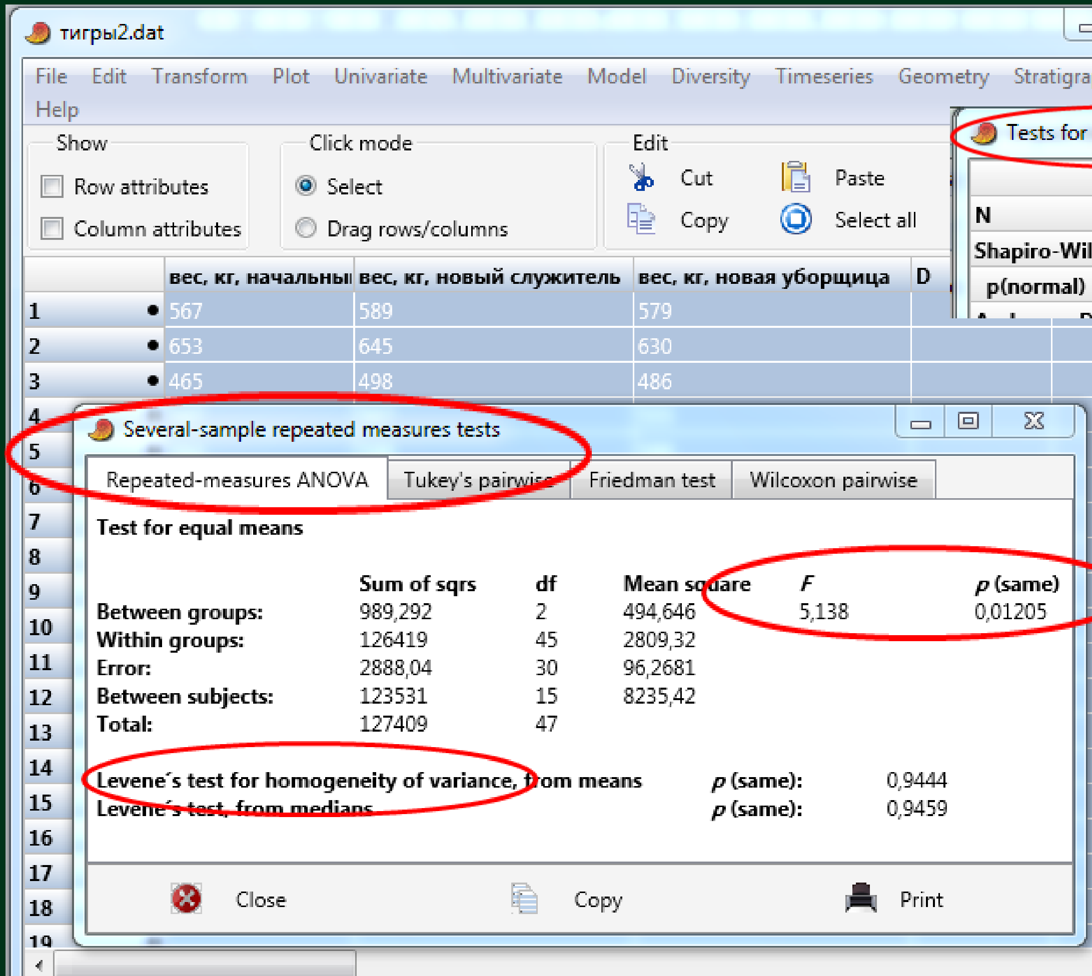
…

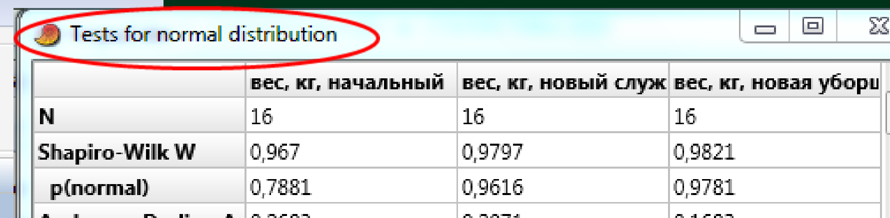
$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$F = \frac{\text{оценка дисперсии между измерениями}}{\text{«ошибка»}}$$

Сравнение связанных групп

В «Пасте»



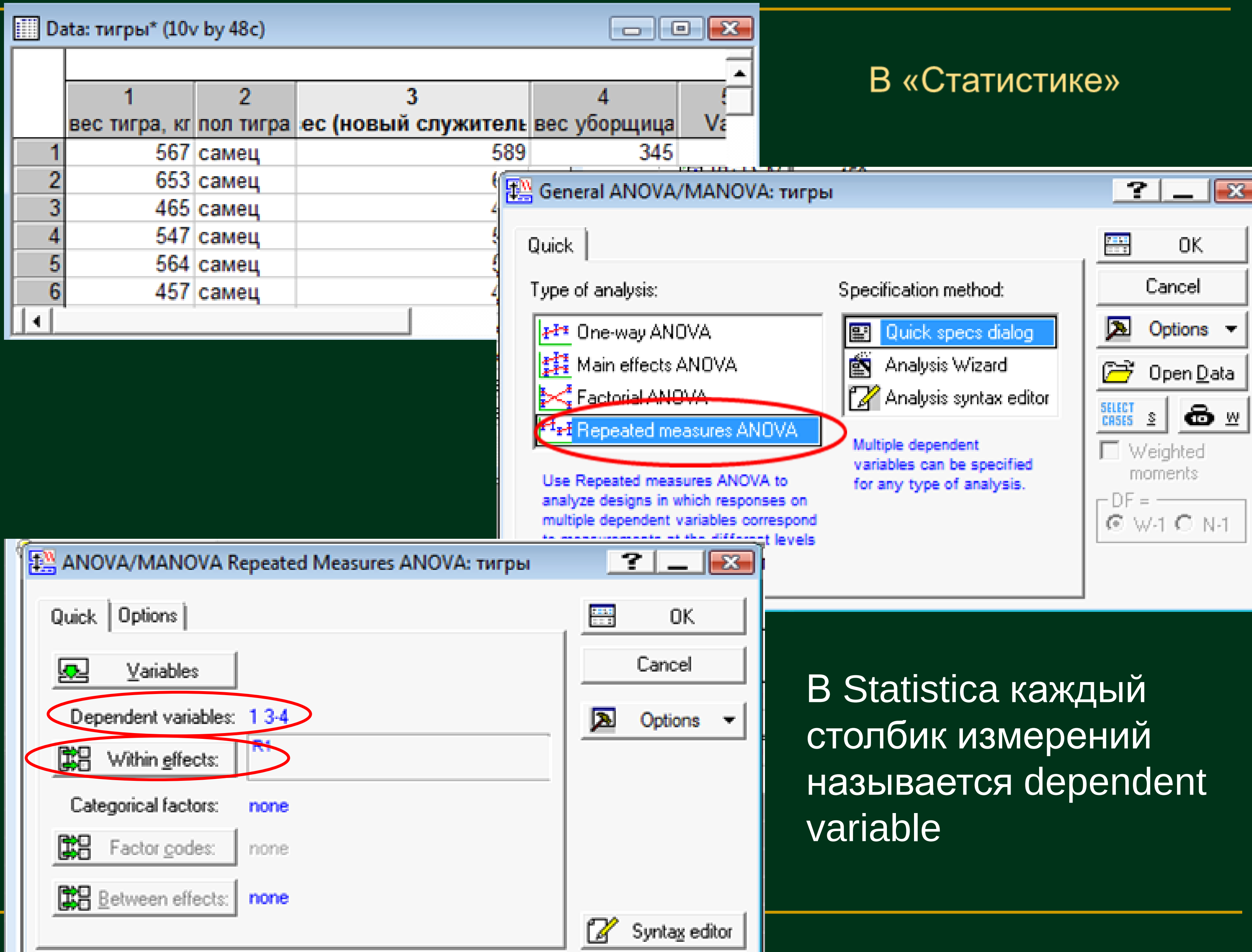


Проверяем:
на нормальность распределения,
гомогенность дисперсий.

Проводим дисперсионный анализ
для связанных групп
(= для повторных измерений)

При P ≤ 0,05 отвергаем
нулевую гипотезу H_0

В «Статистике»



В Statistica каждый столбик измерений называется dependent variable

Отвергаем H_0:
Масса тигров в среднем достоверно изменилась после прихода нового служителя и новой уборщицы.

А теперь можно провести апостериорный (post-hoc) тест.
И выяснить, кто и как повлиял на массу тигров.

Спасибо за внимание!