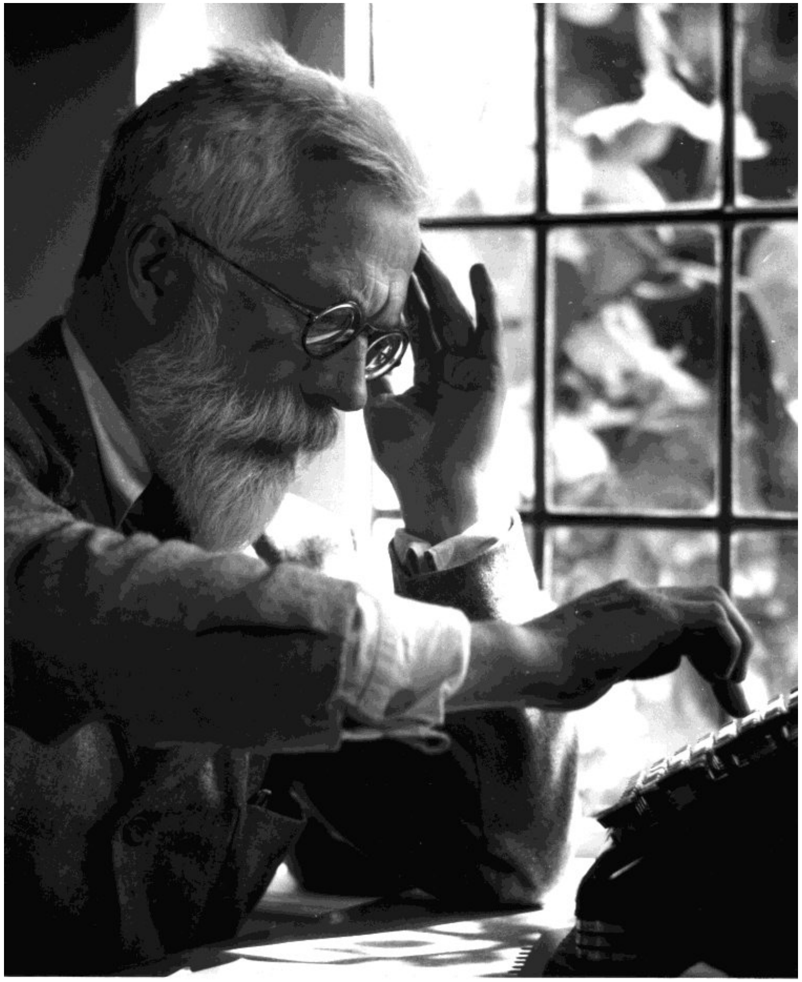
Занятие 4

Дисперсионный анализ (ANOVA): первое знакомство.

Сравнение **ДВУХ** И **БОЛЕЕ** групп

Дисперсионный анализ ANOVA (analysis of variance)



Sir Ronald Aylmer **FISHER**

ANOVA

Мы тестировали гипотезы о среднем значении для одной и двух выборок.
Как быть, если выборок **три или больше**?

В 4-х зоопарках содержатся тигры, и кормят их по-разному. Различается ли средняя масса тигра при разном питании?

ANOVA

Одна <u>зависимая</u> переменная (variable): масса;
Одна <u>независимая</u> (группирующая, factor) – рацион.

$\longrightarrow$ One-way ANOVA

Гипотеза H_0 о равенстве популяционных средних:

Тигров кормили:

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

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

Это <u>сложная гипотеза</u> (omnibus hypothesis). Она включает в себя много маленьких гипотез (для 3-х групп – 3, для 4-х – 12 …):

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

Парные (pairwise) нулевые гипотезы

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

Комплексные (complex) нулевые гипотезы

Примечание: есть другие формулировки нулевых гипотез ANOVA, и вообще это очень упрощённый взгляд

ANOVA

Как сформулировать альтернативную гипотезу? H_0 и H_1 должны быть взаимоисключающими!

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$ Может быть так?

НЕТ!

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 хотя бы одна из маленьких частных альтернативных гипотез (парных или комплексных).
Какая именно – ANOVA не говорит.

ANOVA

Почему бы не сравнить группы попарно *t*-критерием?

(Ошибка использования критерия Стьюдента)

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

Эффект <u>МНОЖЕСТВЕННЫХ СРАВНЕНИЙ</u> (multiple comparisons) возникает, если на одном массиве данных проводят несколько одинаковых статистических тестов (в т.ч., парных сравнений).

Dr. Nostat исследовал эффект 10 лекарств на уровень сахара в крови, сравнил группы попарно и в 1 группе уровень оказался достоверно ниже, чем у 9 других.

ANOVA

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.

Zar, 2010

В случае, если все H_0 на самом деле верны (т.е., различий вообще-то **НЕТ!**), суммарная вероятность ошибки 1-го рода (т.е., вероятность **получить в тесте «достоверные» различия**) в эксперименте из *C* сравнений =

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

В случае 4-х групп ≈ **26,5%**

(на самом деле, чуть меньше)

Чтобы избежать этих трудностей, надо придумать способ, как тестировать только одну гипотезу, чтобы она разом проверяла все различия.

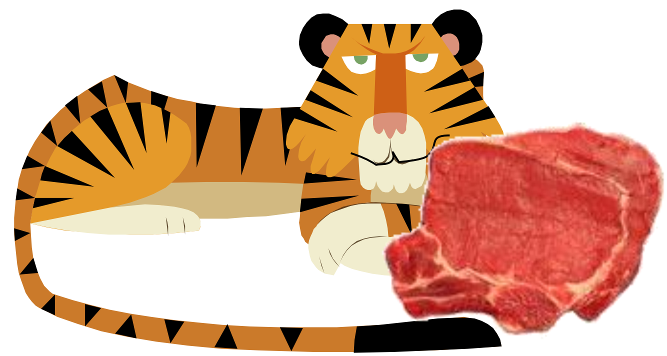
Общая логика ANOVA

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

(гипотеза о равенстве всех популяционных средних)

Пусть H_0 верна. Тогда в случайных выборках из наших популяций выборочные средние значения должны быть **примерно** одинаковы (ведь выборочное среднее – оценка популяционного среднего!).

Вопрос – **насколько** они могут быть одинаковы, если H_0 верна? Если верна H_0, получить большие различия между средними **маловероятно**.



В t-тесте сходство выборочных средних оценить легко – просто посчитать разность. Но с 3-мя (4, 5...) группами так не получится. **Как же оценить различия между средними?**

ANOVA

Для простоты пусть размер групп будет одинаков.

При верной H_0 все группы получены из популяций с одинаковыми **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 средних **между группами** (посчитаем ошибку среднего, как будто это выборочные средние, и из неё вычислим дисперсию)

$$s_{\overline{X}}^2 = \frac{\sum_{j=1}^{k} (\overline{X}_j - \overline{X}_G)^2}{k-1} \qquad SE = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

число групп

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 группах

общее среднее

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

$$\boxed{df_B = k\text{-}1}$$

число групп -1 (4-1=3)

размер группы

MS_B – **m**ean **s**quare **between** groups, оценка расстояния между средними в группах (сокращение от mean squared deviation from the mean)

Чем больше различия между средними, тем меньше вероятность получить их, если H_0 верна.

MS_B = groups MS

ANOVA

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

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

Это в случае групп, одинаковых по размеру; если они различаются, считается «взвешенное по размеру групп среднее» дисперсий.

MS_W – **m**ean **s**quare **within** groups = error MS

Статистика критерия: F

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

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

Тестирование H_0:

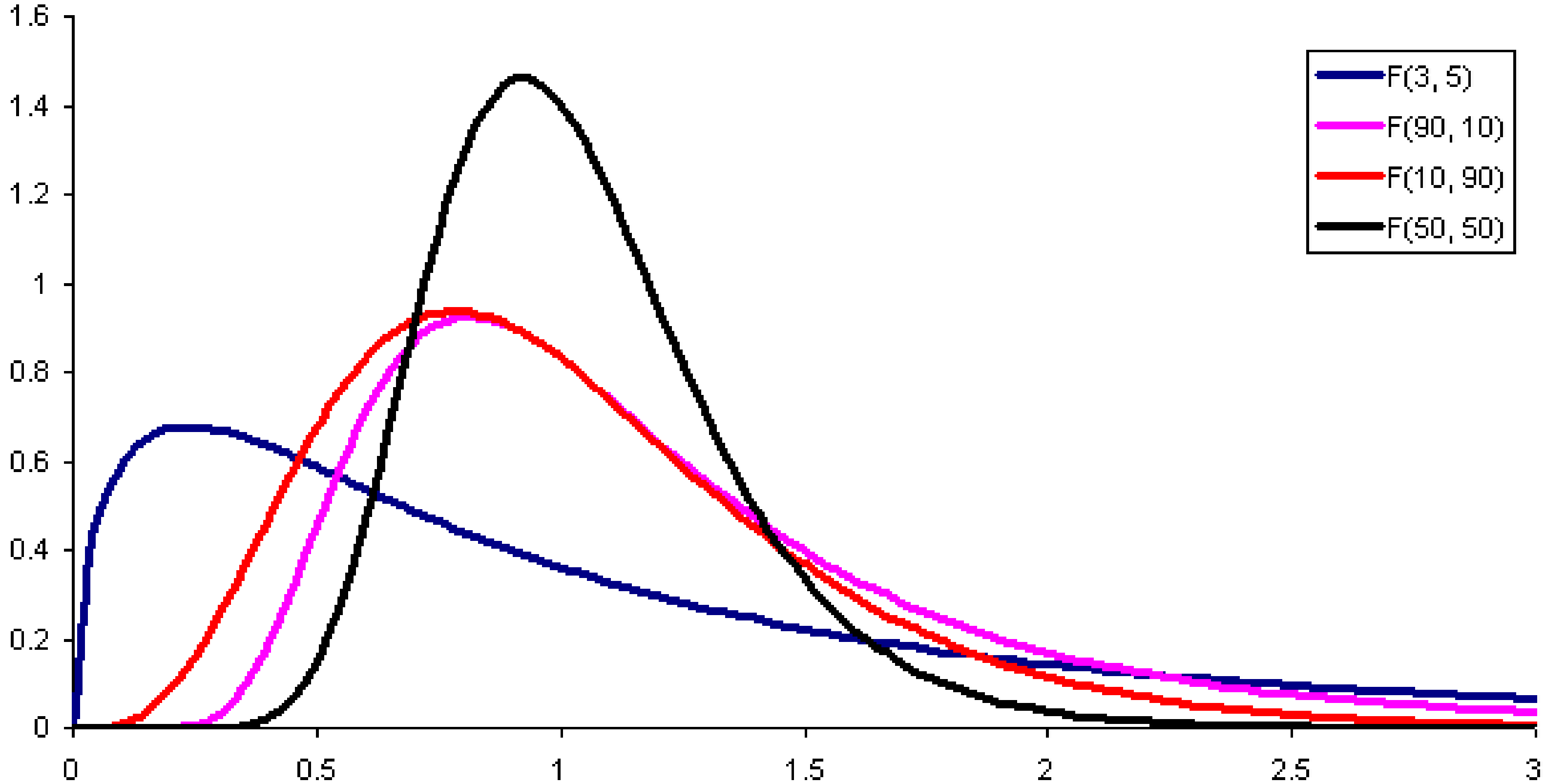
✓для заданных df рассчитывается **критическое значение F**;
✓на основе наших групп считается F и сравнивается с критическим значением;
✓если **F БОЛЬШЕ критического** – H_0 о равенстве средних в группах **отвергаем**;
✓F - это отношение дисперсий, оно имеет особое распределение, оно всегда положительно; ANOVA – принципиально односторонний тест.

ANOVA

Статистика критерия: F

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

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



ANOVA

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 - это **суммы квадратов отклонений** (sum of squared deviations):

$SS_{Between}$ - сумма квадратов отклонений каждого **среднего в группе** от <u>**общего среднего**</u> = **Effect**;

SS_{Within} – сумма квадратов отклонений **каждого измерения** от <u>среднего в соответствующей **группе**</u> = **Error**;

SS_{Total} – сумма квадратов отклонений **каждого измерения** от **общего среднего** = **Total**

ANOVA

$$\boxed{SS_T} = \sum (X_{ij} - \overline{X_G})^2 = \sum (X_{ij} - \overline{X_j})^2 + \sum (\overline{X_j} - \overline{X_G})^2 = \boxed{SS_W + SS_B}$$

$$SS_T = SS_W + SS_B$$

$$df_T = df_W + df_B$$

$df_W = n_G - k$ $df_B = k\text{-}1$

Тестирование H_0:

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

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

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

ANOVA

К чему всё это?

Таблица ANOVA показывает, что **изменчивость** (дисперсия, variance) значений зависимой переменной имеет **два источника**:

✓Различия **между группами** (group = explained variance)

✓Различия измерений **внутри групп** (residual = error).

Можно даже посчитать, какую долю в общей изменчивости составляет межгрупповая изменчивость (доля «объяснённой» изменчивости, explained).

Это используется в сложных моделях, в т.ч. для оценки размера эффекта и вклада разных факторов в изменчивость зависимой переменой.

(a) Zinc level	B	L	M	H
	0.8	0.7	1.8	2.6
	0.9	1.7	2.1	0.6
	2.4	1.0	0.6	1.2
	1.4	1.4	1.1	1.3
	1.3	1.2	2.4	2.2
	1.8	2.4	1.2	0.9
	2.1	1.1	0.9	1.9
	1.0	2.2	1.6	1.0
Means	1.4625	1.4625	1.4625	1.4625

(b) Zinc level	B	L	M	H
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
	1.2	2.3	1.8	0.7
Means	1.2	2.3	1.8	0.7

Доведённые до абсурда варианты соотношения межгрупповой и внутригрупповой дисперсий.

ANOVA

Компоненты общей дисперсии

Группа 1

Группа 2

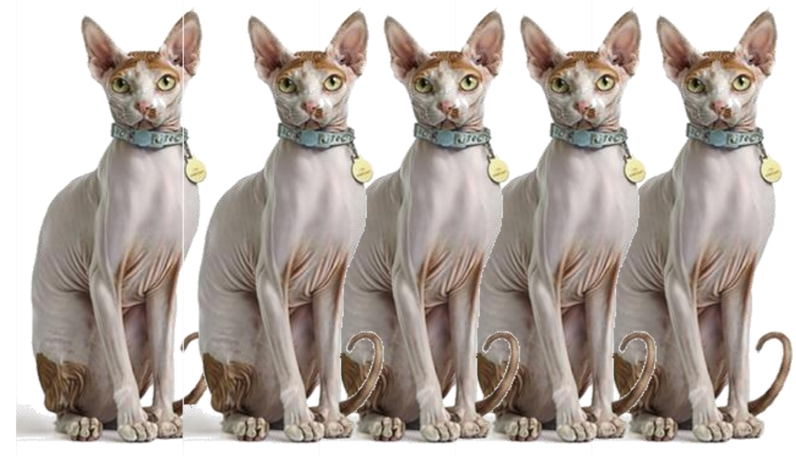
Группа 3

$\boxed{F \to 0}$

ANOVA

Компоненты общей изменчивости

Группа 1



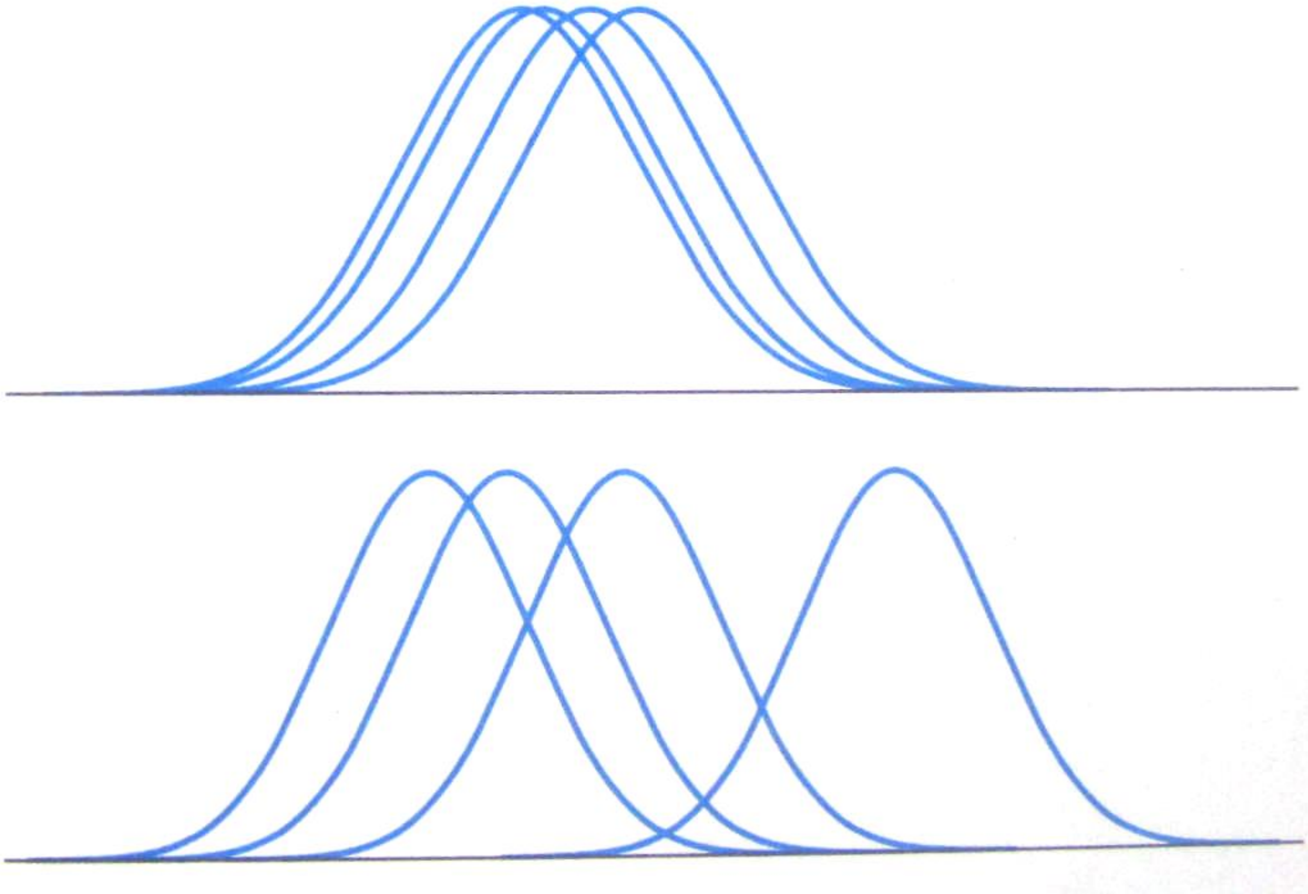
Группа 2

Группа 3

$\boxed{F \longrightarrow \infty}$

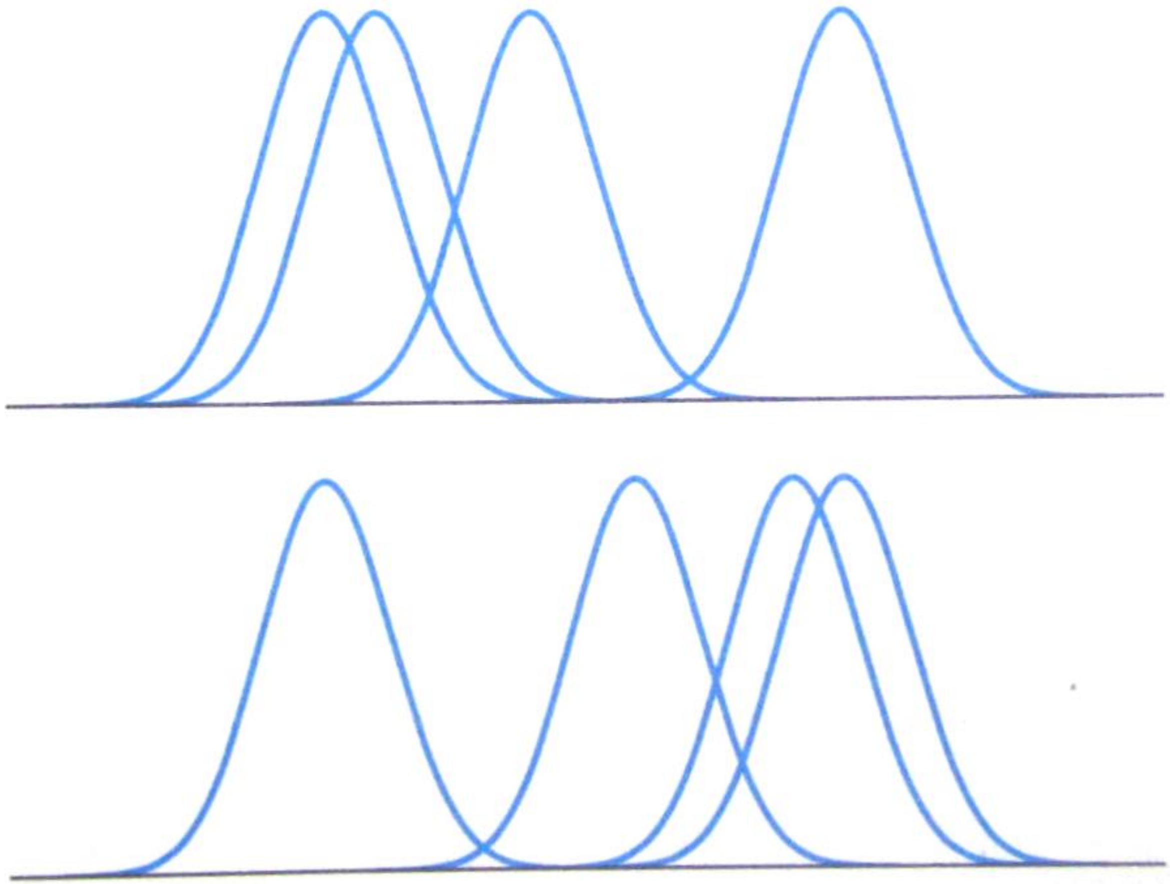
ANOVA

В каком случае значение F-статистики будет больше?



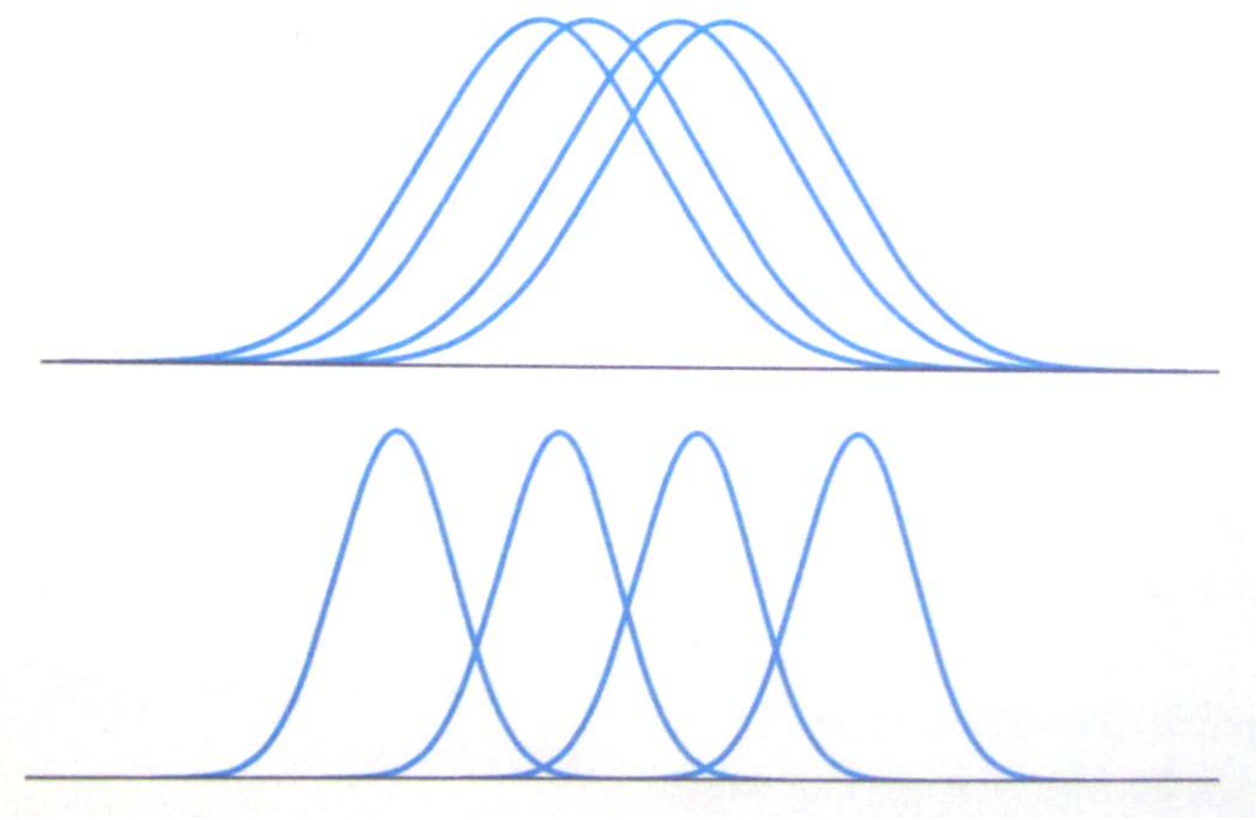
ANOVA

В каком случае значение F-статистики будет больше?



ANOVA

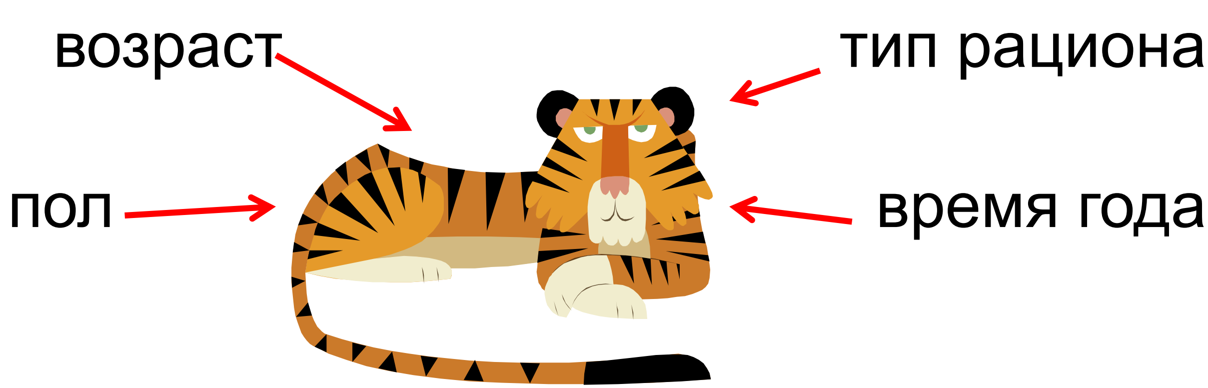
В каком случае значение F-статистики будет больше?



ANOVA: переменные

Группирующая переменная = фактор (factor, predictor).
Зависимая переменная (dependent variable, response).

Мы пока разбираем случай с **одним** фактором (one-way). В ANOVA одна зависимая переменная, а факторов может быть **несколько**, и они могут составлять довольно сложные конструкции.



Факторы в ANOVA бывают ДВУХ ТИПОВ, и это важно!

ANOVA: переменные

Факторы

FIXED

Рассматриваются ИМЕННО ЭТИ значения фактора. Другие значения не существуют/не интересуют нас.

Примеры: Лекарство1, Лекарство2, Лекарство 3, Контроль.
Самцы, Самки.
Вид1, Вид2, Вид3.
…

Модель 1

RANDOM

Рассматриваются СЛУЧАЙНО ВЫБРАННЫЕ значения фактора из многих возможных. За пределами исследования существуют другие значения фактора.

Примеры: Ручей1, Ручей2, Ручей3...
Особь1, Особь2, Особь3, Особь4...
Выводок1, Выводок2, Выводок3…
Дерево1, Дерево2, Дерево3…
Наблюдатель1, Наблюдатель2…
…

Модель 2

ANOVA: переменные

Факторы

FIXED

Исследователь: Сравню-ка я эффективность анальгина и лекарства СтопБобо, под контролем плацебо!

Как бы ни было поставлено это исследование, группы будут три, и именно эти, **других нет**.

RANDOM

Исследователь: Изучу-ка я, различается ли масса лягушек в разных прудах!

Количество прудов в исследовании может быть разным, **существуют неисследованные** пруды.

Для этих типов факторов **по-разному оценивается межгрупповая изменчивость**. Когда фактор один, это не важно, но в сложных моделях с несколькими факторами эти различия очень важны!

ANOVA effect size

1. «доля объяснённой изменчивости»

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

$\eta^2 = 0.01$ – маленький эффект,
$\eta^2 = 0.06$ – средний эффект,
$\eta^2 = 0.14$ – 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

Нет однозначных рекомендаций как считать размер эффекта для ANOVA; чаще всего используют долю объяснённой изменчивости.

Требования к выборкам для one-way ANOVA (assumptions)

1. Выборки должны быть случайными, измерения – <u>независимыми;</u>
2. <u>Размеры групп</u> должны различаться как можно меньше!!!
   Стремиться надо к **равенству размеров групп** (balanced design). Неравенство размеров (unbalanced design) сразу делает результаты анализа уязвимыми к отклонениям от требований нормальности и гомогенности.
   Желательно, чтобы размер групп был ≥ 10

Требования к выборкам для one-way ANOVA (assumptions)

3. Соответствие нормальному распределению (в каждой группе по отдельности!).

Как проверять:

- ✓построить усатые ящики (проверить, симметричны ли они; поверить аутлаеры);
- ✓построить гистограмму;
- ✓применить тест Шапиро-Уилкса;
- ✓построить картинку с residuals (подробнее – в теме про регрессии).

Если есть **АУТЛАЕР**, который нельзя обоснованно исключить, и его исключение/включение качественно меняет результат (достоверно/недостоверно), такие результаты либо не докладывать, либо докладывать оба варианта.

ANOVA

Требования к выборкам для one-way ANOVA (assumptions)

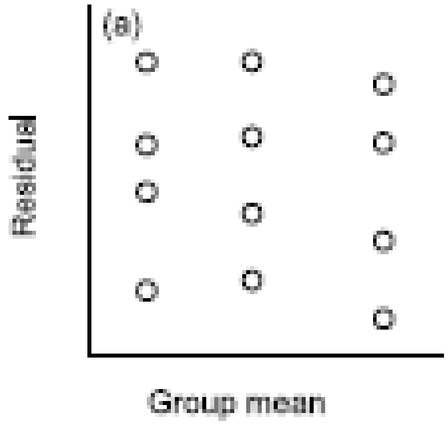
4. Равенство дисперсий в группах.

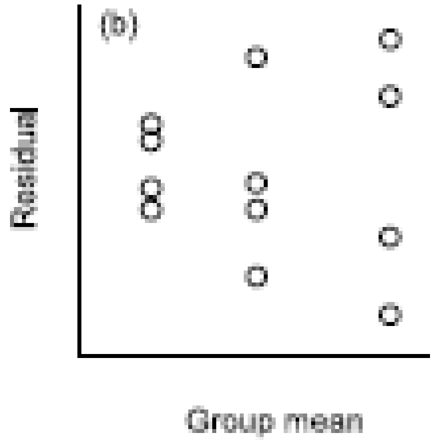
Это **более серьёзное требование, чем нормальность**!

Самое плохое, когда большая дисперсия – в меньшей по размеру группе.

Как проверять:

- ✓ Усатые ящики, гистограммы;
- ✓ Тест Левена (Levene's test);
- ✓ Самый информативный инструмент диагностики качества модели – картинки взаимосвязи residuals (отклонений каждого значения от его группового среднего) и групповых средних (predicted values).



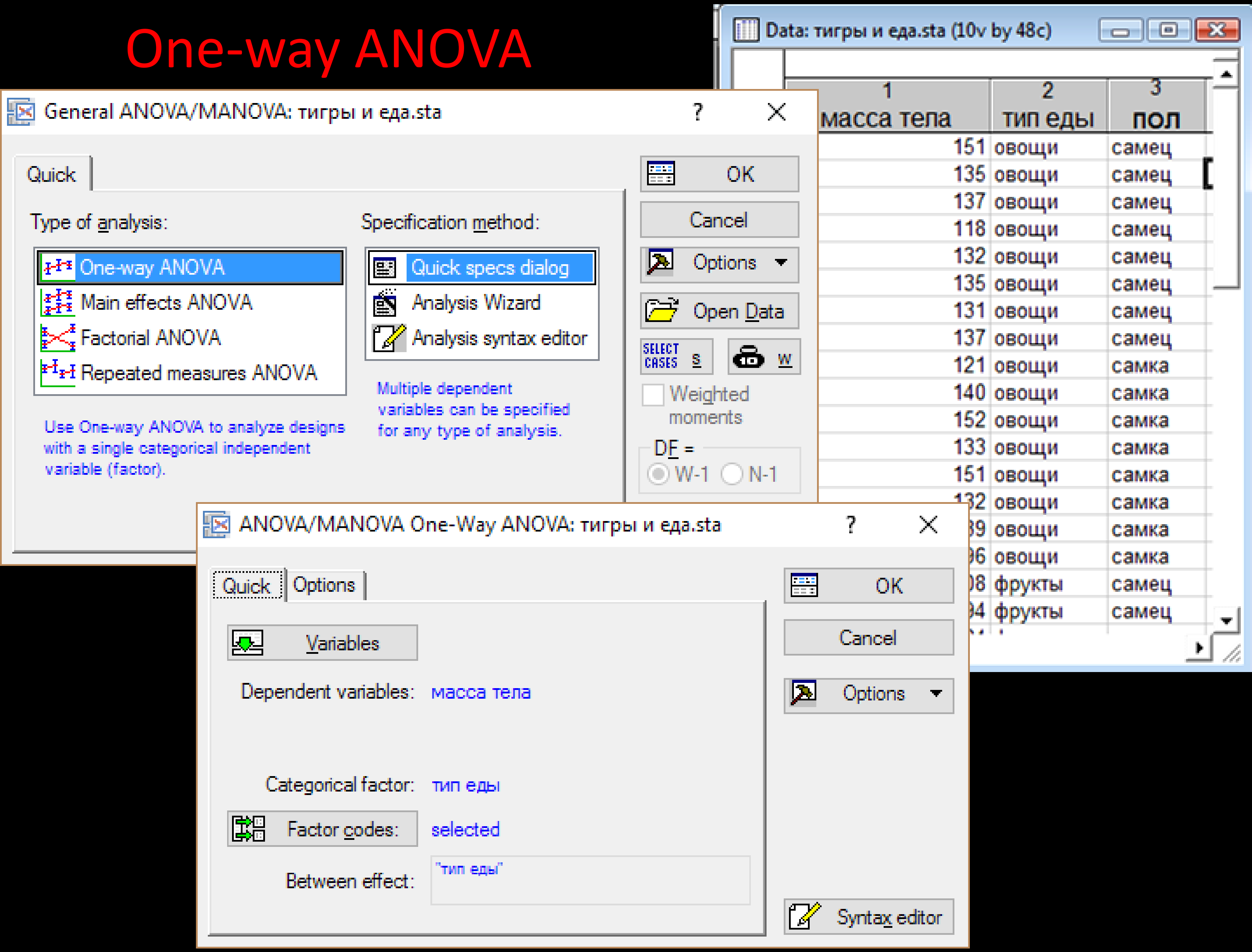


Требования к выборкам для one-way ANOVA (assumptions)

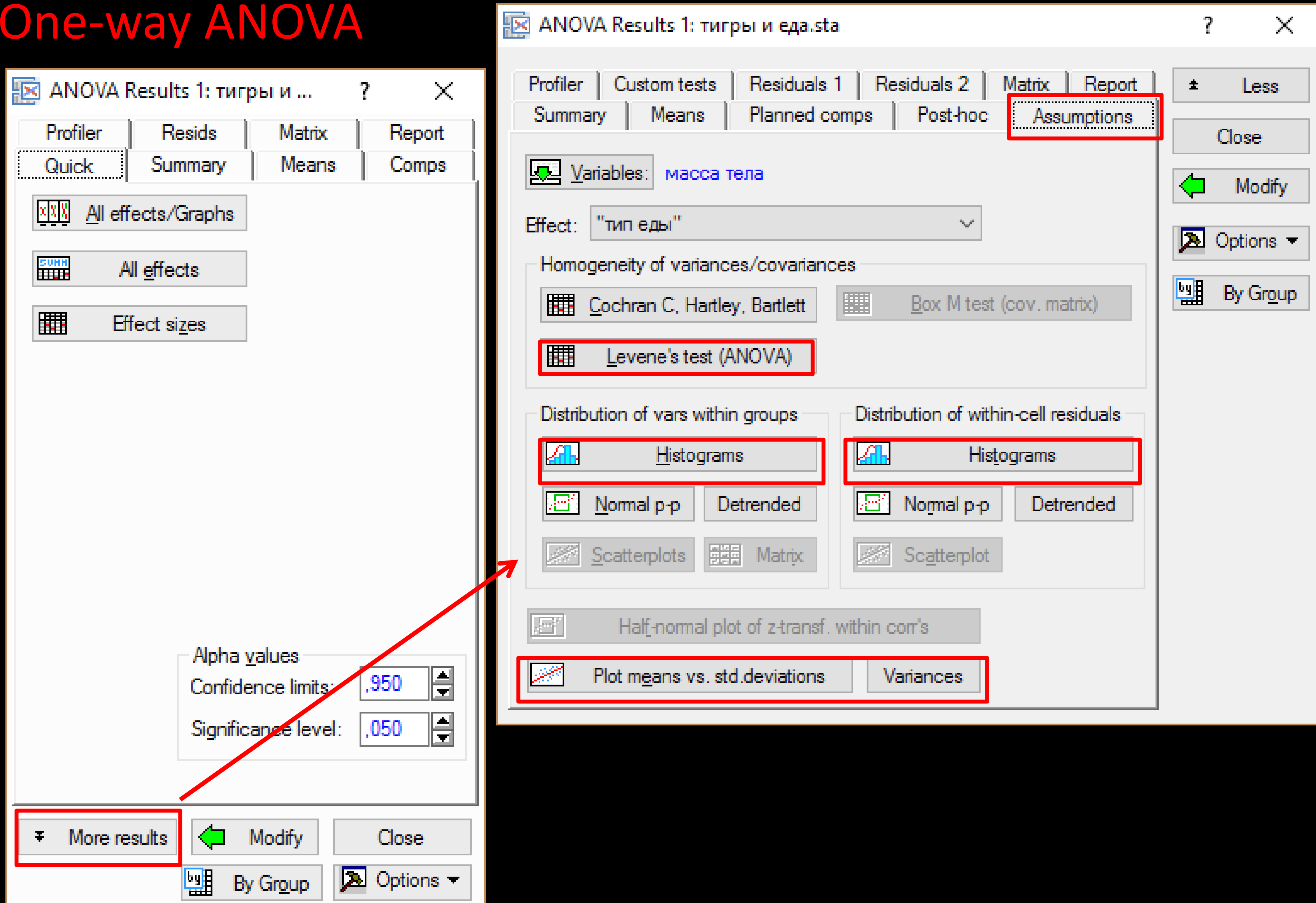
Если проблемы есть:

1. Небольшими отклонениями от гомогенности и даже серьёзными отклонениями от нормальности можно пренебречь при **равенстве размеров групп** (особенно при N>20)
2. **Трансформация данных** часто решает все проблемы разом!
3. Если ничего не помогает, допустима **ранговая** трансформация.
4. Существуют **непараметрические аналоги** АНОВы, о них разговор пойдёт позже.

One-way ANOVA



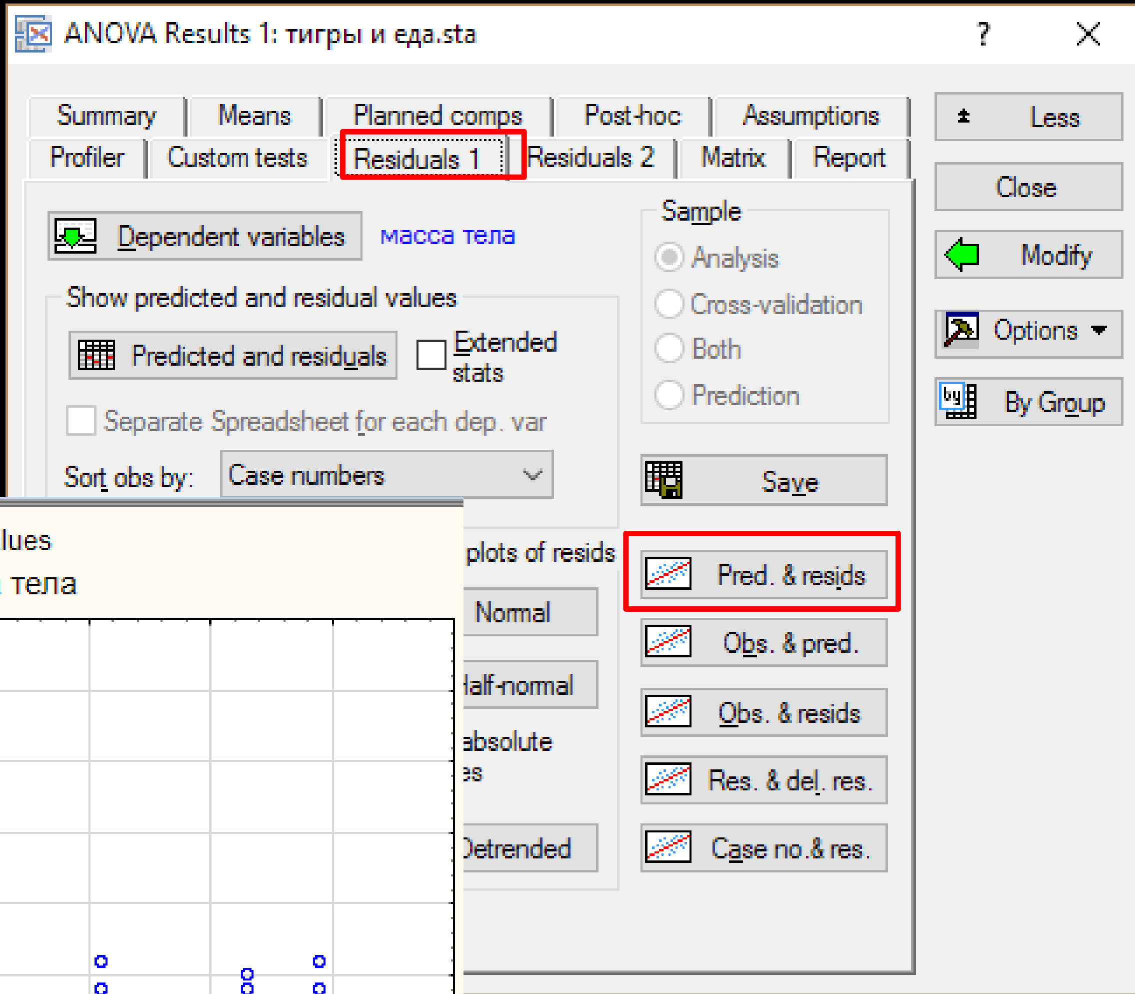
One-way ANOVA

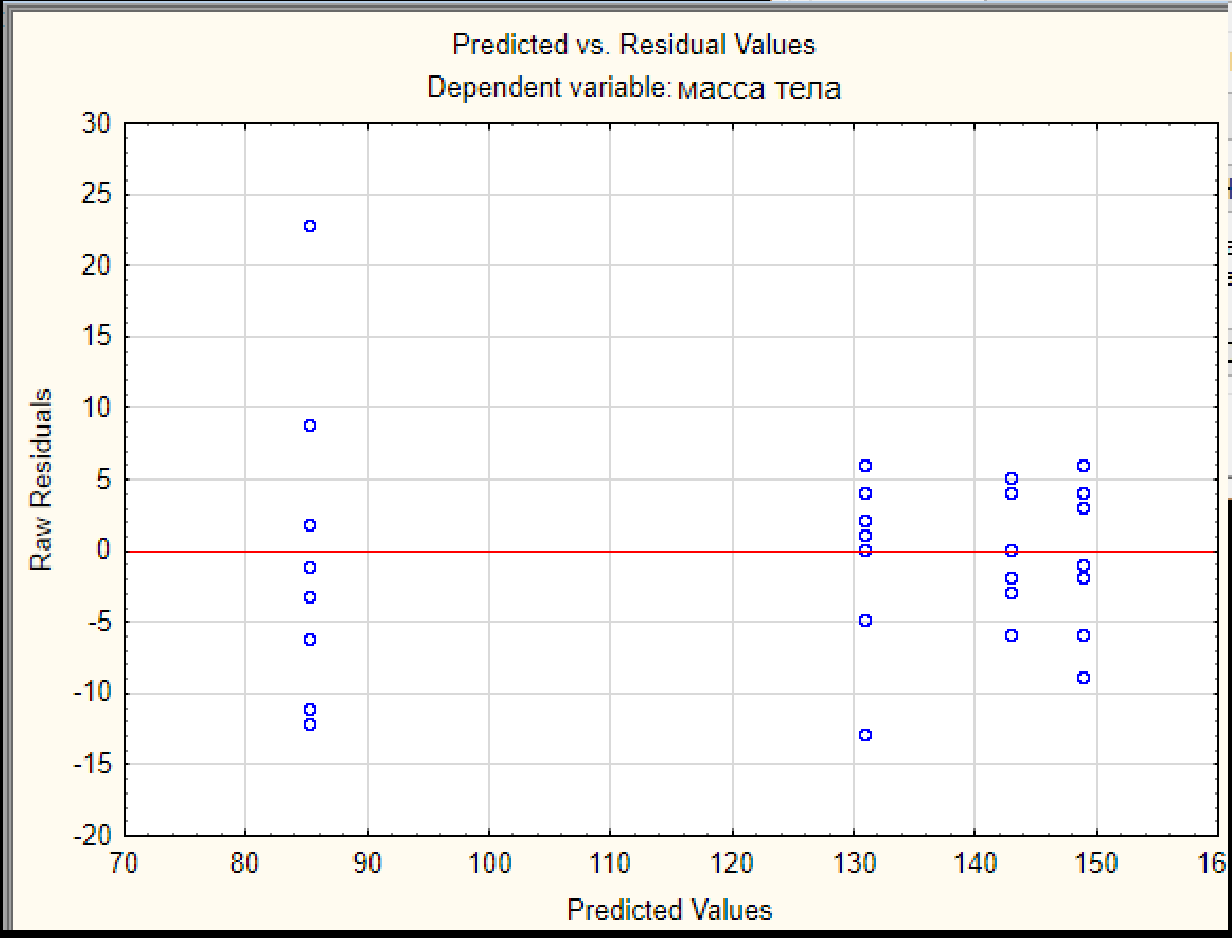


Лучше использовать тест Левена, тест Барлетта – нежелательно (Zar, 2010)

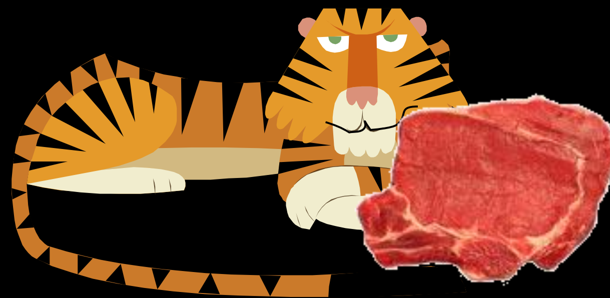
One-way ANOVA

Картинка для диагностики проблем с равенством дисперсий

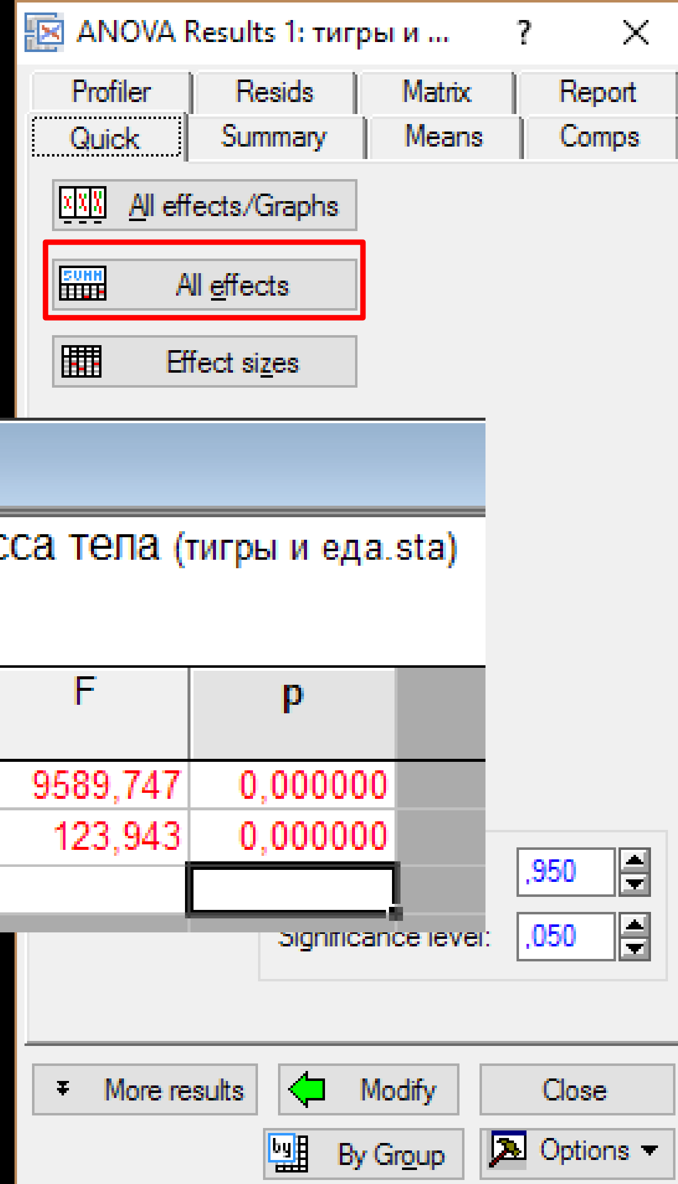




One-way ANOVA



ANOVA table



Univariate Tests of Significance for масса тела (тигры и еда.sta)

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

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	515620,1	1	515620,1	9589,747	0,000000
тип еды	19992,4	3	6664,1	123,943	0,000000
Error	1505,5	28	53,8		

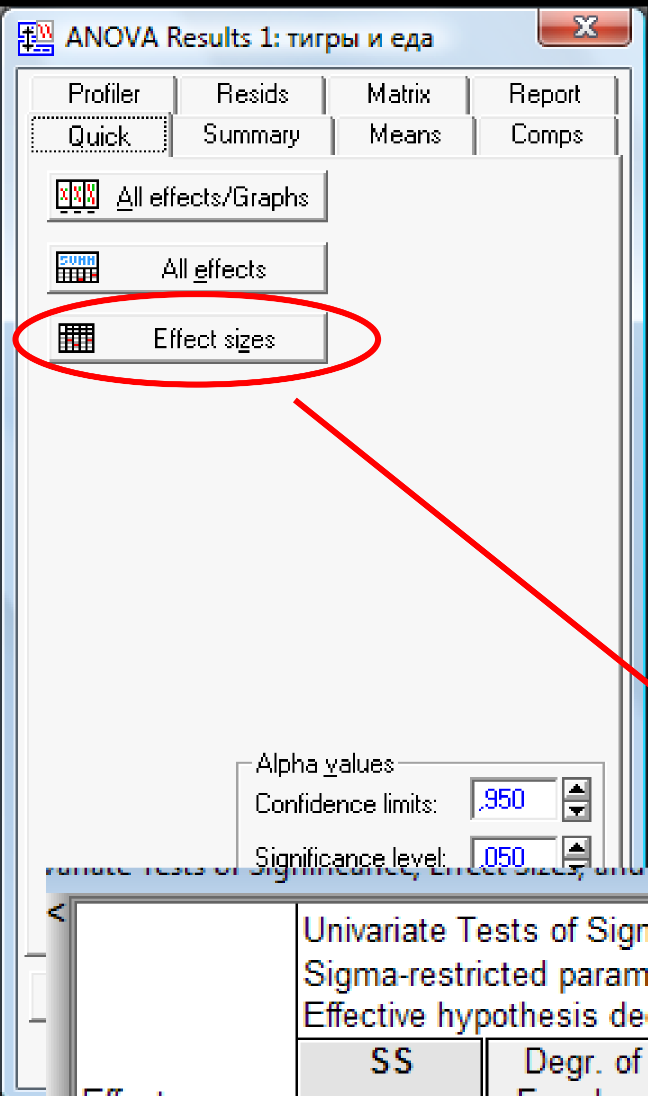
Между группами

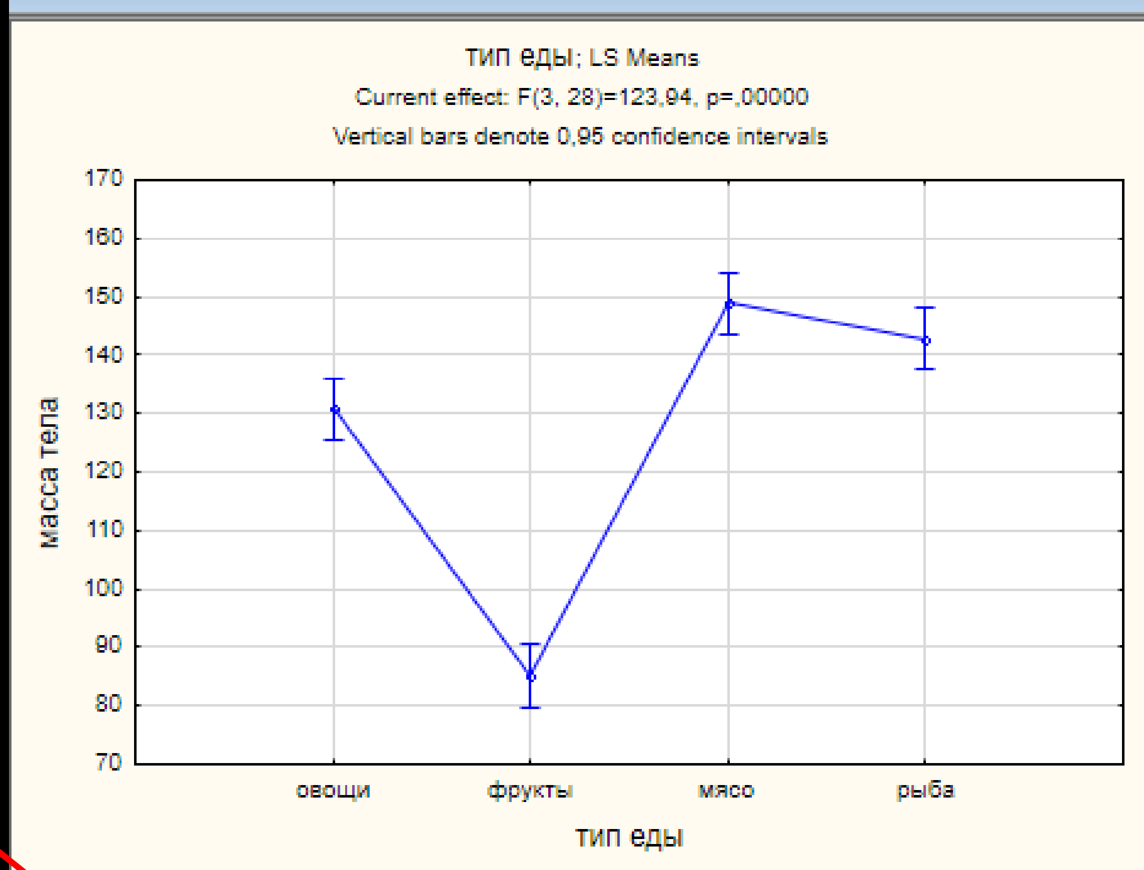
Внутри групп

мы отвергаем H_0.
Рацион влиял на массу тигров

Intercept term is computed as the grand sum of all the count data, squared, then divided by N, the number of observations.

One-way ANOVA





Univariate Tests of Significance, Effect Sizes, and Powers for масса тела (тигры и еда.sta)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p	Partial eta-squared	Non-centrality	Observed power (alpha=0,05)
Intercept	515620,1	1	515620,1	9589,747	0,000000	0,997089	9589,747	1,000000
тип еды	19992,4	3	6664,1	123,943	0,000000	0,929970	371,828	1,000000
Error	1505,5	28	53,8					

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

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

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

$df_W = n_G - k$

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

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

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

$df_B = k\text{-}1$

Для публикаций

В секции **методов** следует указать:

1. Что выборки удовлетворяли критериям нормального распределения и то, каким тестом это было установлено (например, … conform to the assumptions of normality (Shapiro-Wilk's W test, $p<0.05$));

2. Что выборки удовлетворяли условиям гомогенности дисперсий, и то, каким тестом это было установлено;

3. Что данные были проанализированы однофакторным дисперсионным анализом (one-way ANOVA).

В результатах:

F, *df*, *p*, effect size index (η^2); желательно – средние значения с показателями разброса (напр., SD, CI – conf. interval)

…For variables that conformed to a normal distribution (Shapiro–Wilk's W test, p>0.05), we used one-way ANOVA for group comparisons. The samples were homoscedastic (Levene's test, $p>0.05$).

… the body mass varied significantly among age classes ($F_{[2,80]}=10.1$, $p=0.002$, $\eta^2=0.11$).

ANOVA post hoc tests
= unplanned comparisons

Сложная «омнибусная» гипотеза АНОВЫ:

$$\boxed{H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = ... = \mu_k}$$

Похожа на стрельбу из дробовика: непонятно, куда какая дробинка попала.

Если мы отвергли H_0,
непонятно, какая же из отдельных гипотез не верна?

Ответить поможет апостериорный (post hoc) тест!

ANOVA post hoc tests

При сравнении средних значений в ≥3 группах:

1. Сначала сравнить ВСЕ группы между собой с помощью ANOVA
2. Если различия есть, использовать методы множественного сравнения (сравнивают группы попарно, сохраняя общую $\alpha=0.05$)
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 α уже будет не 0.05 (эффект множественных сравнений).

ANOVA post hoc tests

Поправка Бонферрони (*Bonferroni correction*)

если мы хотим обеспечить уровень значимости α, то в каждом из *k* сравнений (т-тестов) нужно принять уровень значимости α/*k*

Простейшая поправка, но очень грубая!

Не работает при большом числе групп – с увеличением их числа очень сильно падает мощность теста.

Сегодня почти не используется, её даже не во все учебники включают.

Behavioral Ecology Vol. 15 No. 6: 1044–1045
doi:10.1093/beheco/arh107
Advance Access publication on June 30, 2004

A farewell to Bonferroni: the problems of low statistical power and publication bias

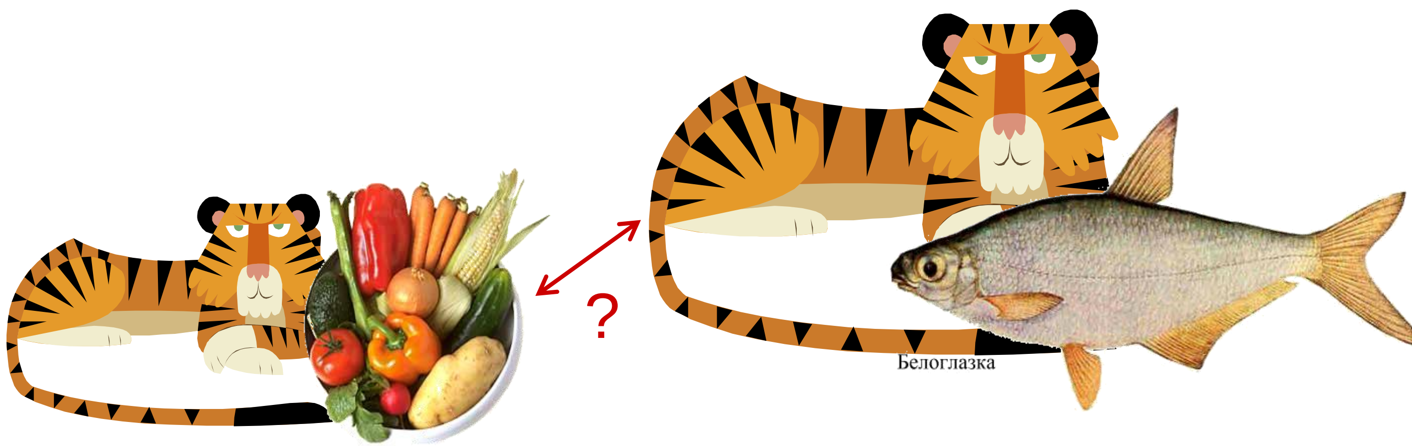
Shinichi Nakagawa
Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom

associated with the standard Bonferroni procedure is a substantial reduction in the statistical power of rejecting an incorrect H_o in each test (e.g., Holm, 1979; Perneger, 1998; Rice, 1989). The sequential Bonferroni procedure also incurs reduction in power, but to a lesser extent (which is the reason that the sequential procedure is used in preference by some researchers; Moran, 2003). Thus, both procedures exacerbate the existing problem of low power, identified by Jennions and

Тест Тьюки (Tukey HSD test)

- ✓Наиболее распространённый и рекомендуемый в литературе тест (Quinn, Keough, 2002; Hurlburt, 2006; Zar, 2010…).
- ✓ строго контролирует α (0.05)
- ✓ разом проверяет все парные гипотезы;
- ✓ плохо работает, если размер групп сильно различается;
- ✓ чувствителен к неравенству дисперсий;
- ✓ считает статистику (Q) на основе MS_{within} и df

$$H_{01}: \mu_1 = \mu_2$$
$$H_{02}: \mu_1 = \mu_4$$
$$H_{03}: \mu_1 = \mu_3$$
$$\ldots$$



ANOVA post hoc tests

Другие апостериорные тесты

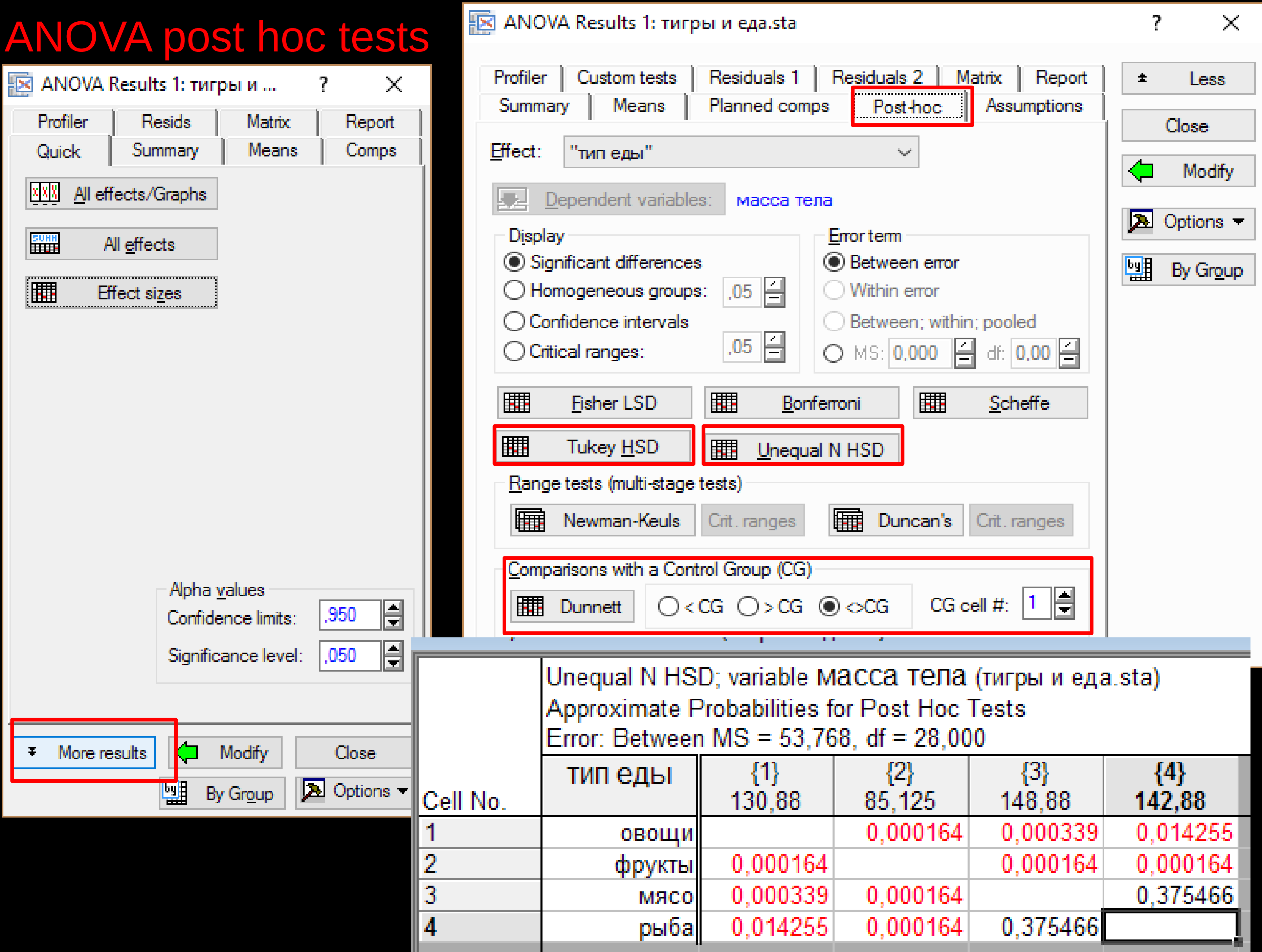
1. Критерий **Ньюмена-Кейлса** (*Newman-Keuls test*). Все средние упорядочивают по возрастанию и пошагово вычисляют статистики; начинают от сравнения наибольшего с наименьшим. Мощнее теста Тьюки, но плохо контрлирует ошибку 1-го рода – не рекомендуется.

2. Критерий **Шеффе** (Scheffe test) – очень консервативный, мощность меньше, чем у т. Тьюки.

3. Критерий **Даннетта** (*Dunnett test*) – используется для сравнения нескольких групп с контрольной группой, мощнее, чем т. Тьюки. Размер контрольной группы рекомендуется делать больше, чем размеры остальных групп в $\sqrt{k-1}$ раз.

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.

ANOVA post hoc tests



Unequal N HSD; variable масса тела (тигры и еда.sta)
Approximate Probabilities for Post Hoc Tests
Error: Between MS = 53,768, df = 28,000

Cell No.	тип еды	{1} 130,88	{2} 85,125	{3} 148,88	{4} 142,88
1	овощи		0,000164	0,000339	0,014255
2	фрукты	0,000164		0,000164	0,000164
3	мясо	0,000339	0,000164		0,375466
4	рыба	0,014255	0,000164	0,375466	

Для публикаций

В **результатах** указывают:

Обязательно – сначала достоверные результаты ANOVA, т.е., *F, df*, *p*;

Для теста Тьюки – *p* значение (например, … Tukey post hoc test, *p*=0.0001 …). Некоторые программы выдают ещё и статистику Q

Анализ контрастов
(planned comparisons)

Сложная «омнибусная» гипотеза АНОВЫ:

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

На самом деле, обычно исследователя не все возможные сравнения интересуют одинаково. Наверняка есть сравнение, важное для нас в первую очередь.

Если нас просто интересует сравнение 2-х групп, можно предпринять т-тест и с АНОВой не связываться. А вот как проверить комплексную гипотезу?

Анализ контрастов (planned comparisons)

a priori Tests =
Planned comparisons = анализ контрастов
(вместо ANOVA)

✓Важно: то, какие группы сравнивать, выбирают **ЗАРАНЕЕ**, до проведения какого-либо анализа! В идеале – ещё при постановке исследования.

✓В тесте проверяется только **одна гипотеза**;

✓можно провести **2-3 таких теста** в пределах одного «набора» групп, только надо следить, чтобы сравнения не сильно перекрывались, не были избыточными.

✓Процедура тестирования у *a priori* тестов – почти как у t-критерия Стьюдента.

✓**Мощность** теста принципиально **выше**, чем у пост-хок тестов.

✓Принципиально лучше, чем объединять группы по одну сторону знака =, т.к. учитывает межгрупповую изменчивость.

Анализ контрастов

Особенно удобно использовать для тестирования комплексных (а не парных) гипотез.

Dr. Diet разработал новую диету и собирается протестировать её эффективность. Из 20 добровольцев
группа 1 (n=5) соблюдает новую диету;
группа 2 (n=5) занимается на тренажёре;
группа 3 (n=5) занимается аэробикой;
группа 4 (n=5) бегает по утрам.

Сравнение лекарств разных групп;

анализ контрастов

Зависимая переменная – число грамм, на которое изменилась масса тела добровольцев за 3 месяца.

Можно было бы провести ANOVA затем апостериорный тест, но нас интересует лишь сравнение диеты Dr. Diet с разными видами физических упражнений.

анализ контрастов

$$H_0: \mu_1 = \frac{\mu_2 + \mu_3 + \mu_4}{3}$$

$$\boxed{H_0: \mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 = 0}$$

$$C_1\mu_1 + C_2\mu_2 + C_3\mu_3 + C_4\mu_4$$

$$\sum C_j = 0$$

«**Контраст**» = «сравнение» (contrast, comparison) – линейная комбинация средних значений.

<u>Коэффициенты сравнения</u> – константы, на которые умножены средние. В сумме = нулю.

Если на одном «наборе» групп хотят провести несколько контрастов, очень желательно, чтобы их коэффициенты при соответствующих групповых средних удовлетворяли условию

$$\sum_{i=1}^{k} c_{iA}\, c_{iB} = 0$$ (для пары контрастов A и B)

Примеры коэффициентов для анализа контрастов при разных нулевых гипотезах

Null hypothesis	Coefficients of the comparison
Some pairwise null hypotheses:	
H_0: $\mu_1 = \mu_2$	$1, -1, 0, 0$
H_0: $\mu_1 = \mu_3$	$1, 0, -1, 0$
H_0: $\mu_2 = \mu_4$	$0, 1, 0, -1$
Some complex null hypotheses:	
H_0: $\mu_1 = \dfrac{\mu_2 + \mu_3 + \mu_4}{3}$	$1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}$
H_0: $\dfrac{\mu_1 + \mu_2}{2} = \dfrac{\mu_3 + \mu_4}{2}$	$\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}$
H_0: $\mu_3 = \dfrac{\mu_1 + \mu_2 + \mu_4}{3}$	$-\frac{1}{3}, -\frac{1}{3}, 1, -\frac{1}{3}$

анализ контрастов

$$H_0 : \mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 = 0$$

$$H_1 : \mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 \neq 0$$

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

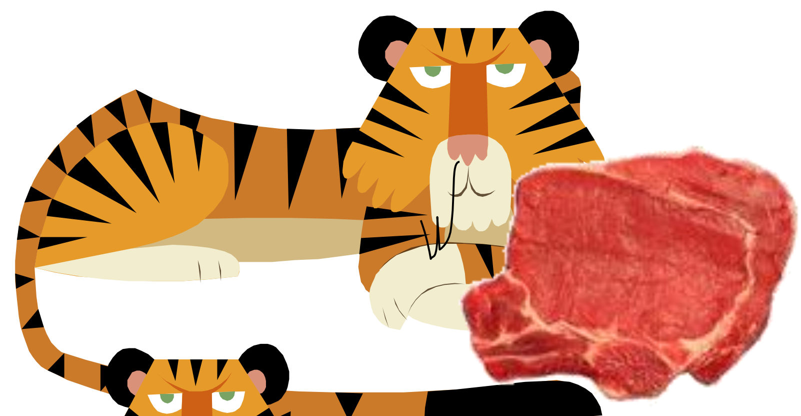
$$\text{Статистика} = \frac{\text{выборочное сравнение - 0}}{\text{стандартная } \textbf{ошибка} \text{ выборочного сравнения}}$$

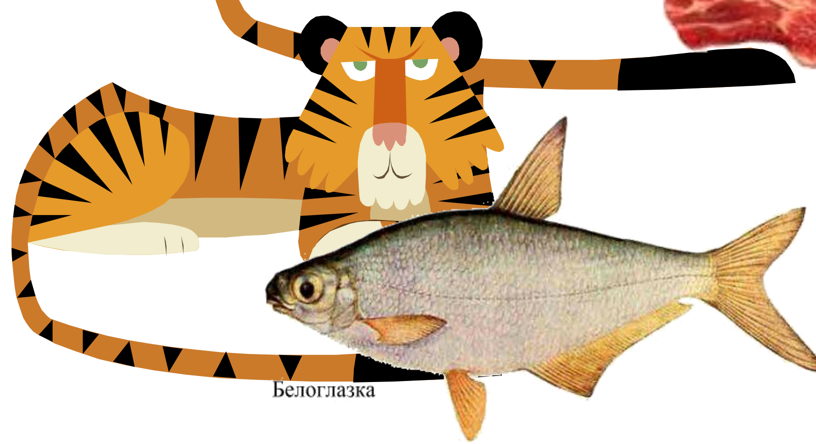
$$\boxed{t = \frac{C_1\bar{X}_1 + C_2\bar{X}_2 + C_3\bar{X}_3 + C_4\bar{X}_4}{s_{contrast}}}$$

Она имеет t-распределение

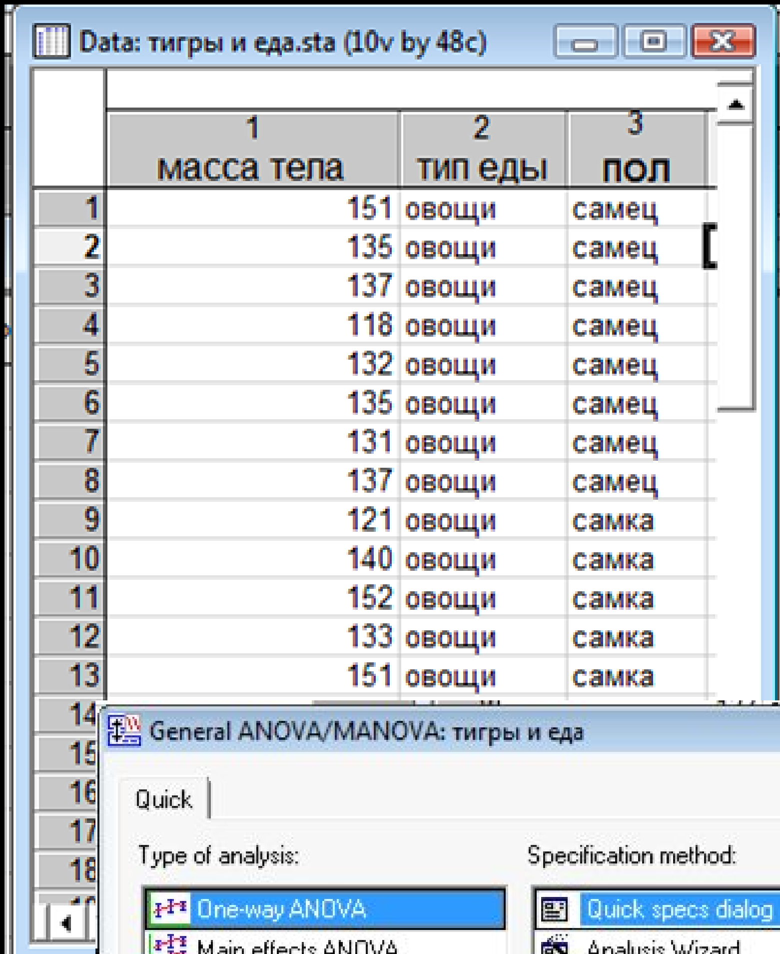
Ещё один пример:
У нас 4 группы тигров, их кормят: овощами; фруктами; рыбой; мясом.
Вопрос: отличается ли масса тигров, питающихся животной и растительной едой?

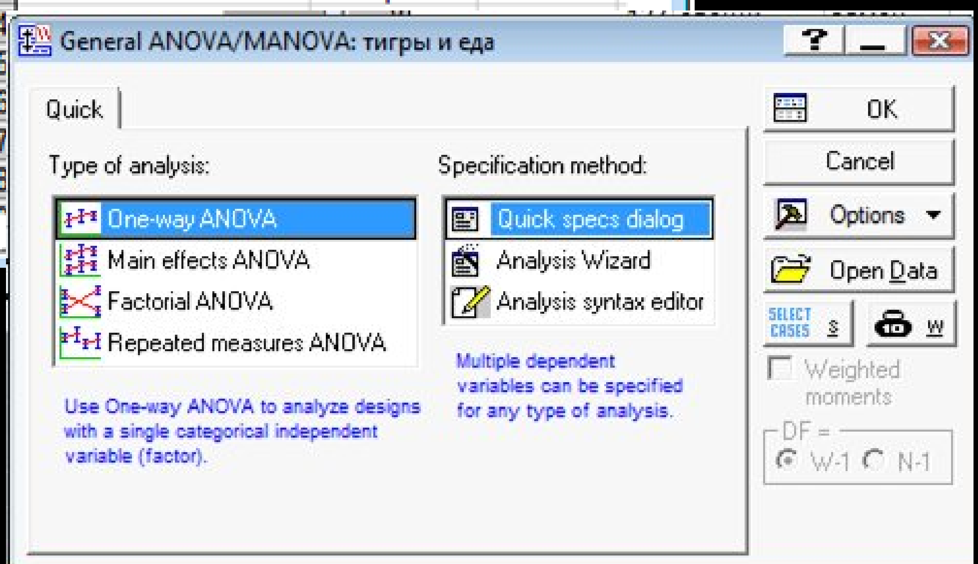
$$H_0 : \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$$

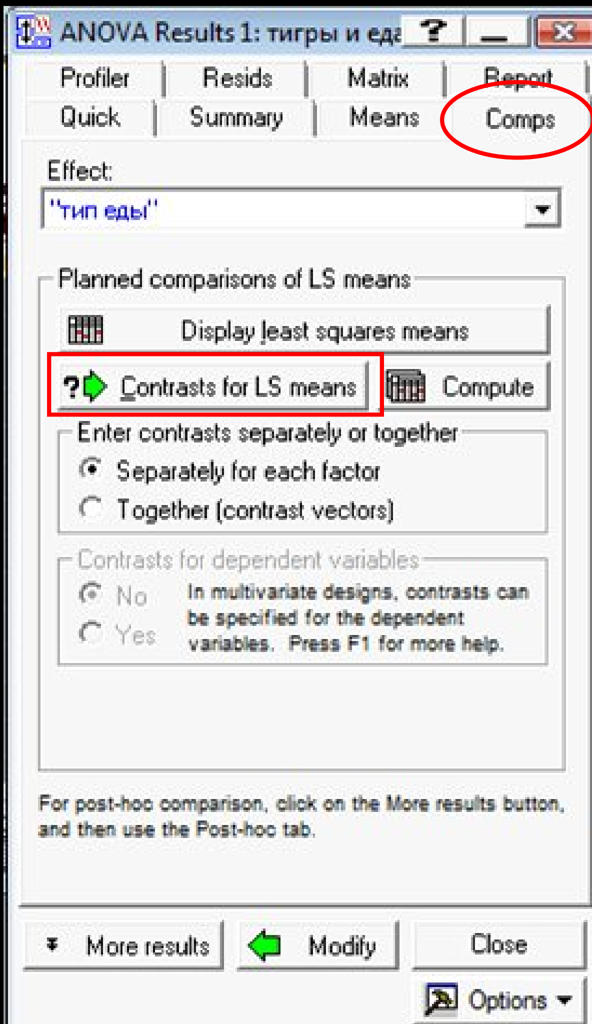




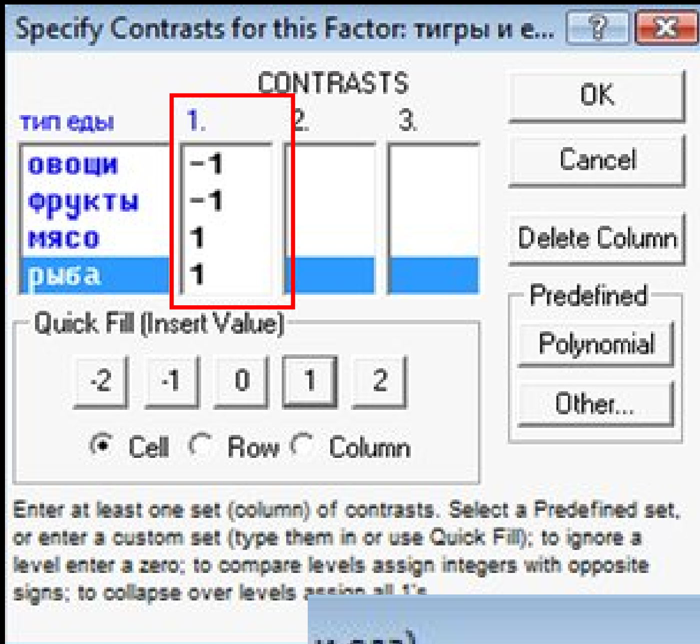
анализ контрастов

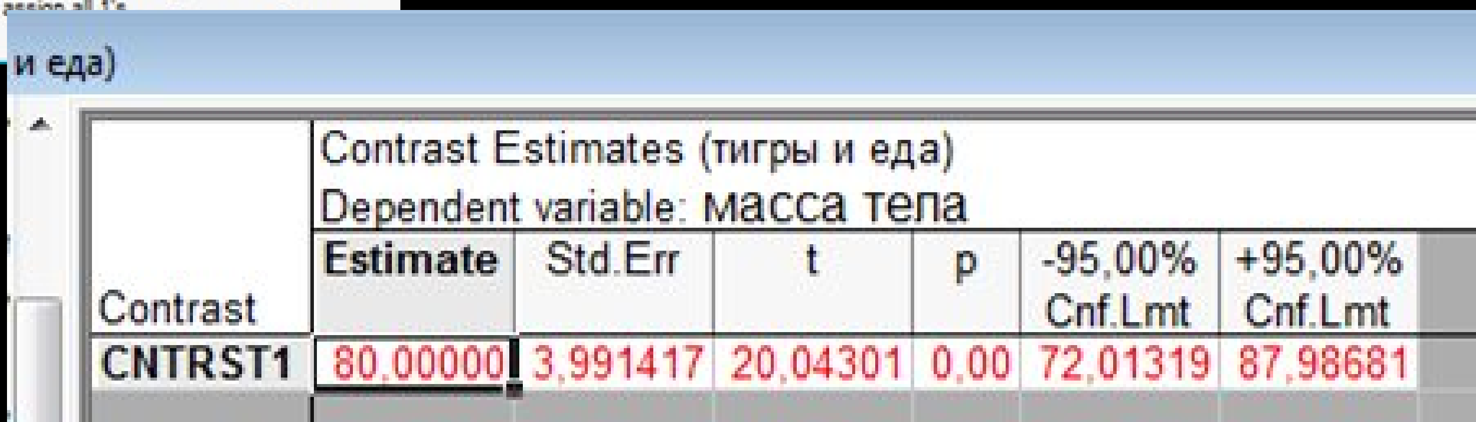






анализ контрастов





Contrast	Estimate	Std.Err	t	p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
CNTRST1	80,00000	3,991417	20,04301	0,00	72,01319	87,98681

мы отвергаем H_0.
Масса тигров, питавшихся растительной и животной едой, достоверно различалась.

К практической части

1. Файл Gator: сначала проверить на нормальное распределение, потом посчитать АНОВу в Basic stats ANOVA/MANOVA, размер эффекта не забыть. Картинки построить какие только можно, в том числе assumptions.
2. Power analysis для АНОВы вставить значения из пред. Задания
3. post-hoc test – Rats. Assumptions tab, Plot means vs. std. deviations
4. planned comparisons – Gator

1. Про то, что можно представить модель взаимодействия переменных как уравнение;
2. правда, в этом уравнении число «неизвестных» (популяционное среднее, оценки дисперсий) больше, чем количество средних значений в группах, и поэтому, чтобы его «решить» (оценить эти «неизвестные») придумали разные уловки-ограничения.
3. Про оценку силы действия фактора путём сравнения остаточной изменчивости моделей с этим фактором (полной) и модели без него

$$F_{df_B,df_W}$$