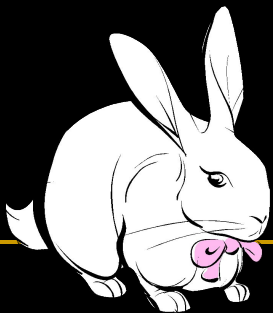


Занятие 3

Мощность статистического
теста.

Дисперсионный анализ
ANOVA



Мощность

Мощность - вероятность отвергнуть H_0 в эксперименте, когда H_0 действительно неверна.

	Истинное (но неизвестное нам) положение дел	
	Верна H_0	Верна H_1
Мы «приняли» H_0	ПРАВИЛЬНО!	ОШИБКА 2-го рода = β
Мы отвергли H_0	ОШИБКА 1-го рода = α	ПРАВИЛЬНО! МОЩНОСТЬ критерия = $1-\beta$

Мощность

Т.е., масса землероек в Заповеднике на самом деле больше, чем 90 г.

Мощность – вероятность того, что проведённое нами исследование установит этот факт.

$$H_0: \mu \leq 90 \text{ г};$$

$$H_1: \mu > 90 \text{ г}$$



Ошибка 2-го рода + мощность = 1

$$\beta + (1-\beta) = 1$$

(это 2 возможных результата теста, если H_0 не верна)

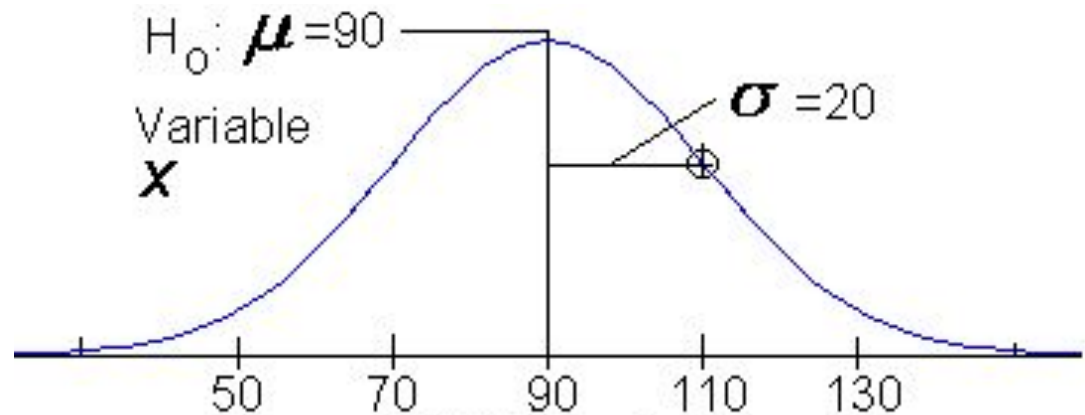
Мощность

Мощность предполагаемого статистического теста - ключевой элемент планирования исследования

«Реальное значение» параметра:

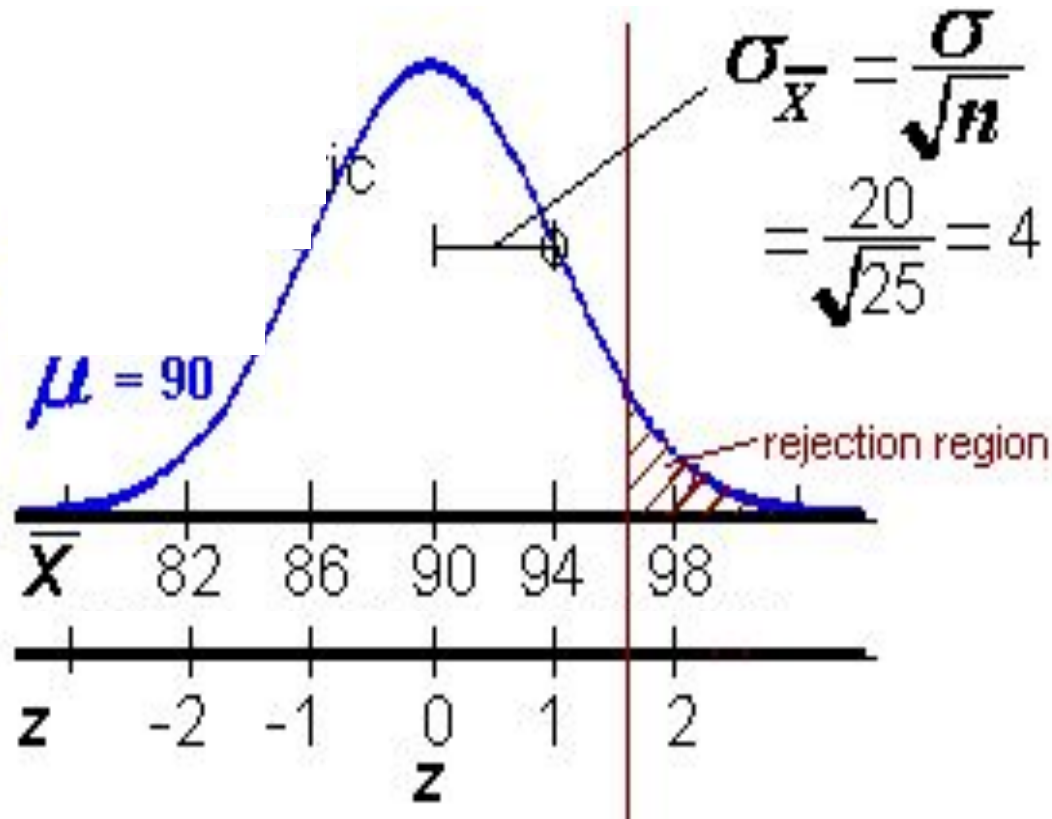
Во всей мировой популяции землероек $\mu = 90$ г.

Пусть «реальное значение» средней массы в заповеднике = 94 г.



Мощность

Нарисуем распределения выборочных средних для $\mu = 90$ и $\mu = 94$ (стандартное отклонение $\sigma = 20$).

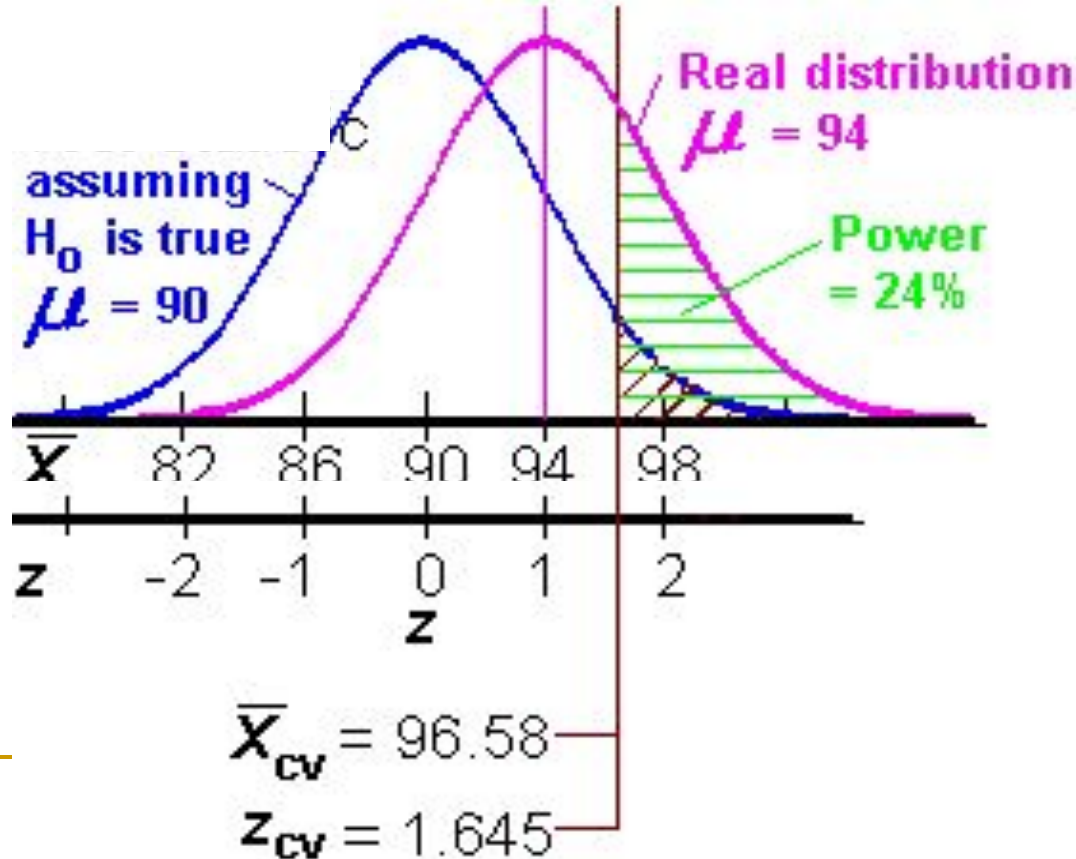


Размер
выборки $n =$
25 зверей



Мощность

Если мы поймаем 25 землероек в заповеднике, у нас есть вероятность лишь 24%, что мы найдём различия! Т.к. лишь в 24% случаев среднее из нашей будущей выборки попадёт в критическую область.



Как увеличить мощность?

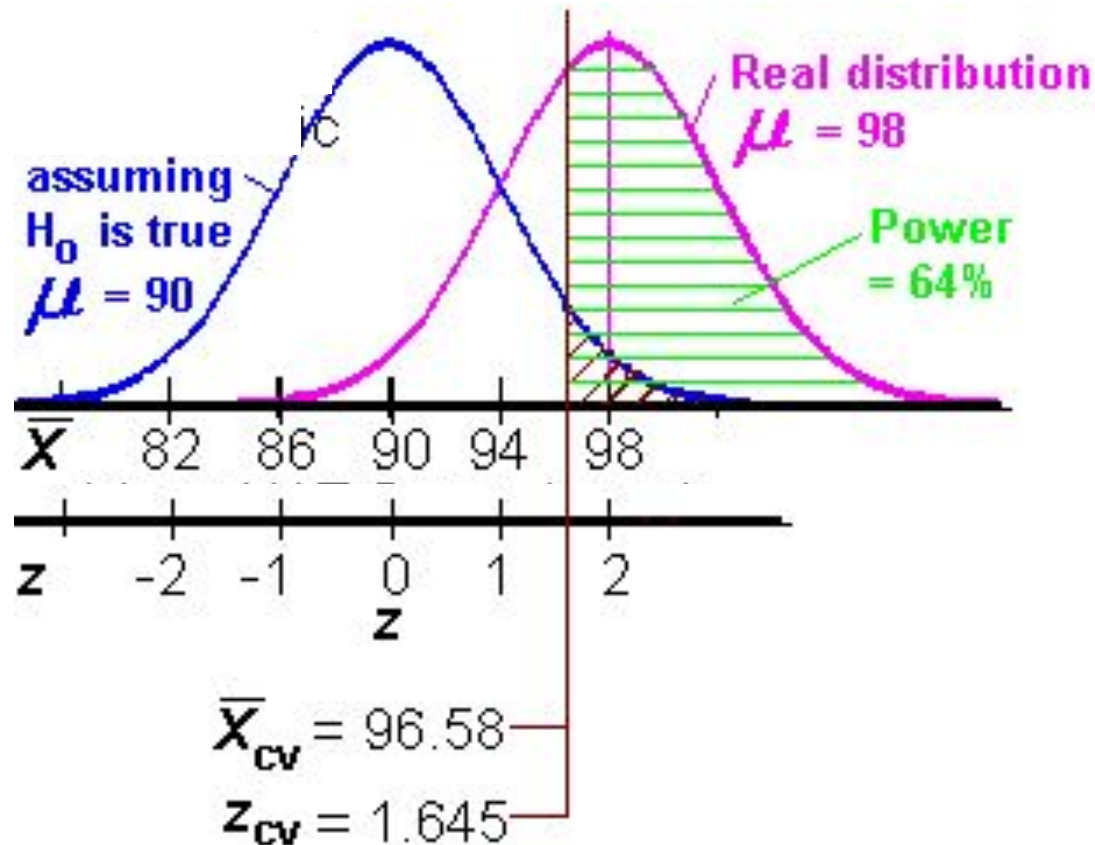
Большей **МОЩНОСТИ** критерия способствуют:

1. Большой размер выборки;
2. Большие различия между популяциями (effect size);
3. Маленькое стандартное отклонение;
4. Большой уровень значимости ($\alpha=0.05$ а не $\alpha=0.01$);
5. Выбор одностороннего теста вместо двустороннего



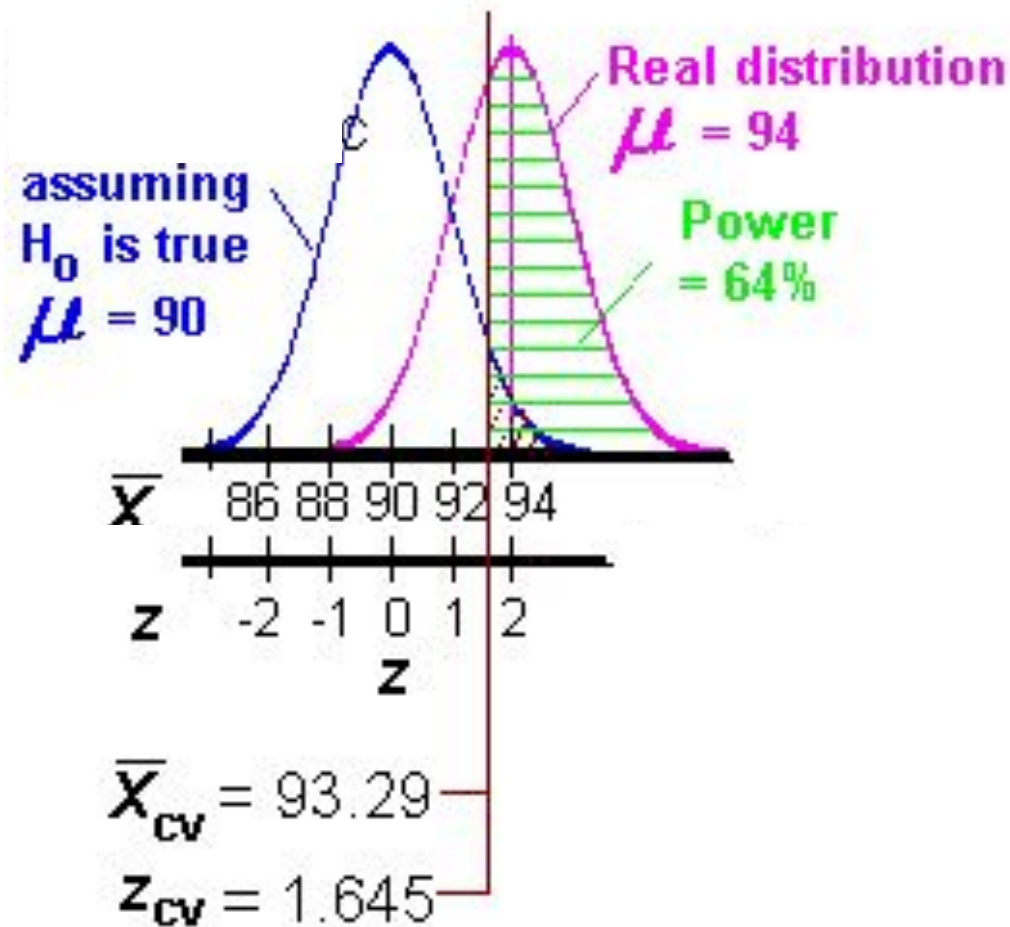
Мощность

Если в действительности средняя масса землероек в заповеднике равна 98 г, мощность теста будет уже 64%.



Мощность

Здесь стандартное отклонение уменьшили вдвое, и мощность теста тоже стала 64%.

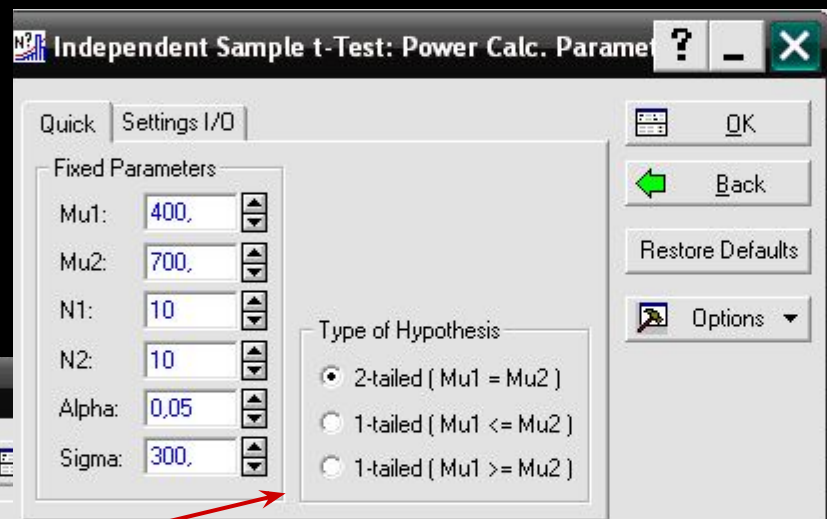
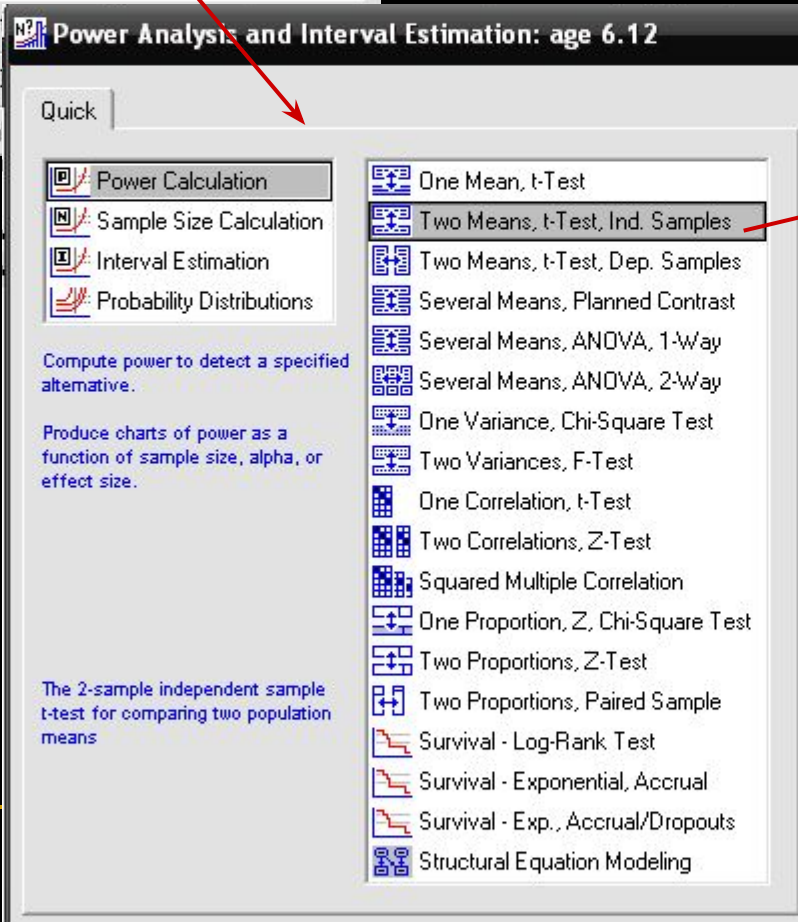
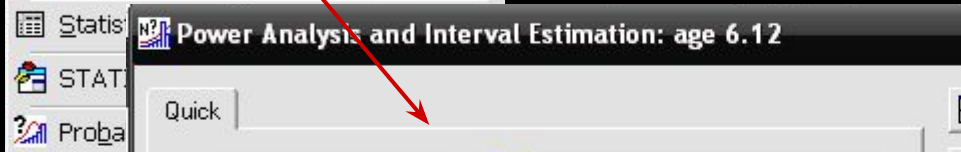
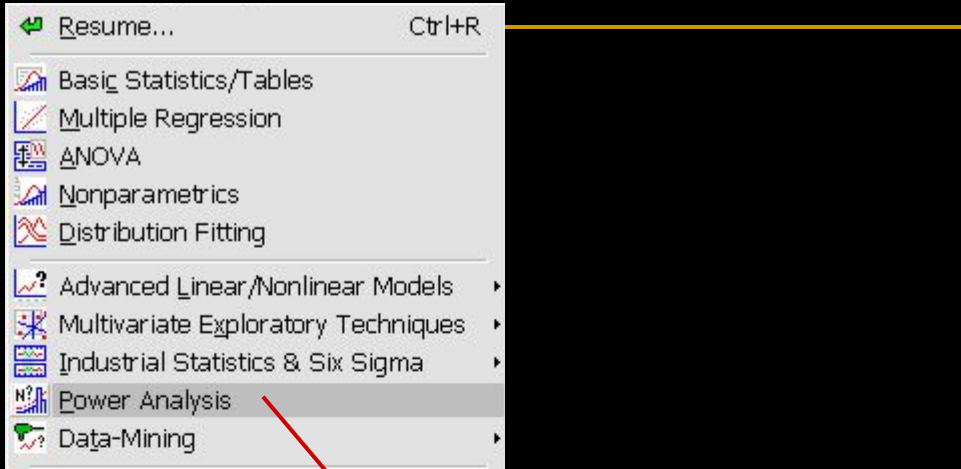


Как использовать понятие мощности критерия:

При **планировании** исследования мы можем рассчитать размер выборки, необходимый для того, чтобы «разглядеть» предполагаемые различия между выборками.

(Реальные различия нам, очевидно, неизвестны, но можно задать минимальные, имеющие биологическое значение).

Ещё мы можем **после** проведения теста (в котором мы не отвергли H_0) оценить вероятность ошибки (2-го рода).



Расчёт мощности
для
двухвыборочного t-
критерия для
независимых
выборок.

Independent Sample t-Test: Power Calc. Results: age 6

Independent Sample t-Test: Power Calculation

H0: $\mu_1 = \mu_2$

Type I Error Rate (Alpha): 0,05

Population Mean μ_1 : 400

Population Mean μ_2 : 700

Sample Size N1: 10

Sample Size N2: 10

Population S.D. (Sigma): 300

Standardized Effect (Es): -1



Quick Settings I/O

X-Axis Graphing Parameters

Start N: 10

End N: 100

Start Es: 0,30

End Es: 0,90

Start Alpha: 0,01

End Alpha: 0,25

No. of Steps: 10

Power Charts

Power vs. N

Power vs. N1

Power vs. N2

Power vs. Es

Power vs. Alpha

Calculate Power

Change Params

Back

Options

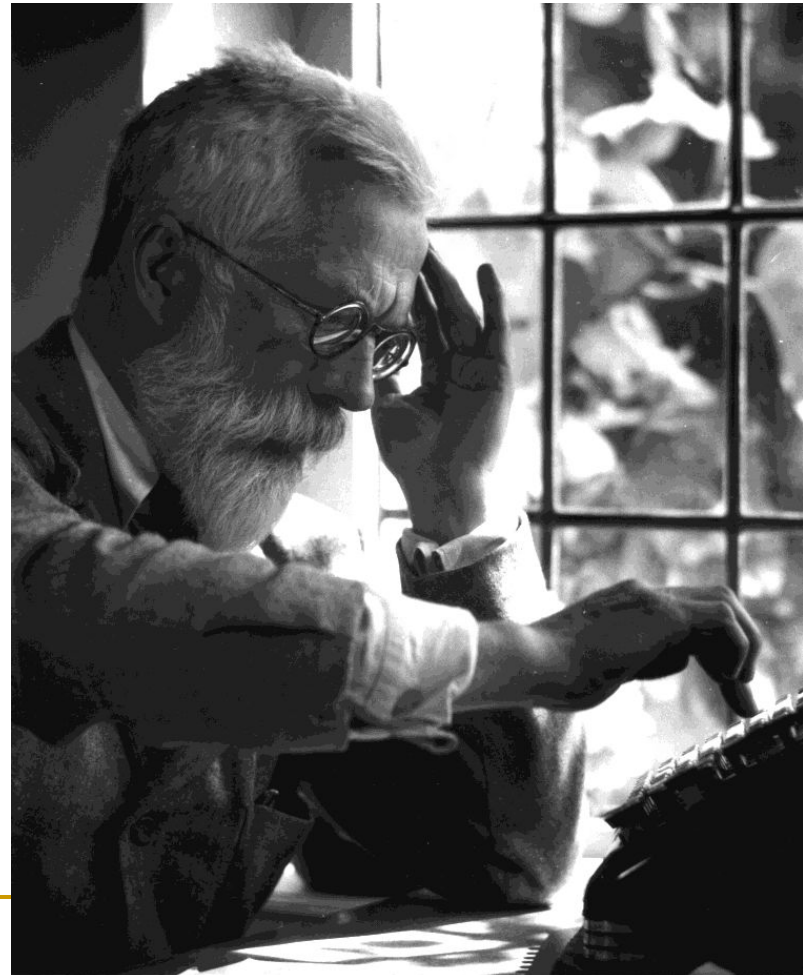
Power Calculation (age 6.12)	
Two Means, t-Test, Ind. Samples	
H0: $\mu_1 = \mu_2$	
	Value
Population Mean μ_1	400,0000
Population Mean μ_2	700,0000
Population S.D. (Sigma)	300,0000
Standardized Effect (Es)	-1,0000
Sample Size N1	10,0000
Sample Size N2	10,0000
Type I Error Rate (Alpha)	0,0500
Critical Value of t	2,1009
Power	0,5620

ANOVA

Сравнение **ДВУХ И БОЛЕЕ** групп

Дисперсионный анализ
ANOVA (analysis of variance)

Sir Ronald Aylmer
FISHER

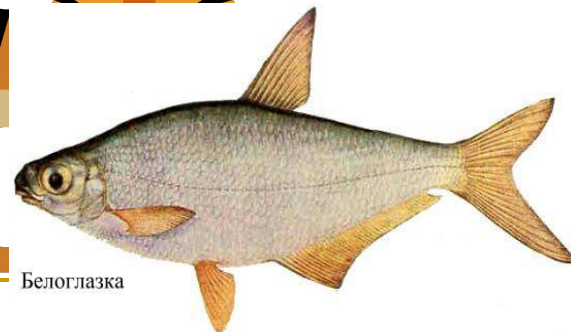
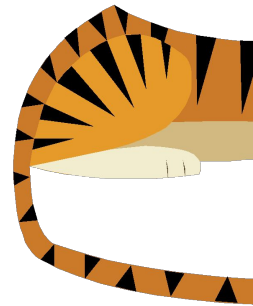


ANOVA

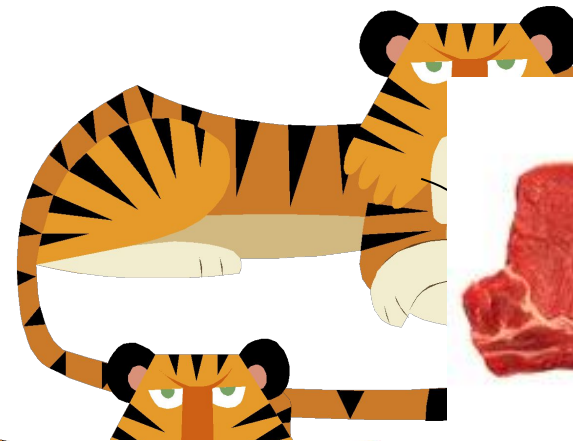
Мы тестировали гипотезы о среднем значении для одной и двух выборок.

Как быть, если выборка **три или больше**?

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



Белоглазка



ANOVA

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 (для 3-х групп – 3, для 4-х – 12 ...):

$$H_{01} : \mu_1 = \mu_2$$

$$H_{02} : \mu_1 = \mu_4$$

$$H_{03} : \mu_1 = \mu_3$$

$$H_{04} : \mu_2 = \mu_3$$

$$H_{05} : \mu_2 = \mu_4$$

$$H_{06} : \mu_3 = \mu_4$$

Парные
(pairwise)
нулевые
гипотезы

$$H_{07} : \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$$

$$H_{08} : \mu_1 = \frac{\mu_2 + \mu_3 + \mu_4}{3}$$

...

Комплексные
(complex)
нулевые
гипотезы

Зависимая переменная: масса;

Независимая (группирующая) – тип еды.

Формулируем альтернативную гипотезу:

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

НЕВЕРНО!

$$H_1: \mu_1 \neq \mu_2 \quad \text{или} \quad \mu_1 \neq \mu_3 \quad \text{или} \quad \mu_1 \neq \mu_4 \quad \dots$$

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. 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 попарном сравнении нескольких групп).

При уровне значимости $\alpha=0,05$ вероятность ошибиться в хотя бы в одном из k сравнений примерно равна:

$$P_{\text{ошибки}} = 1 - (1 - 0,05)^k$$

Например, для попарного сравнения 4-х групп $k=6$, т.е., $P_{\text{ошибки}} = 1 - (1 - 0,05)^6 = 0,265$

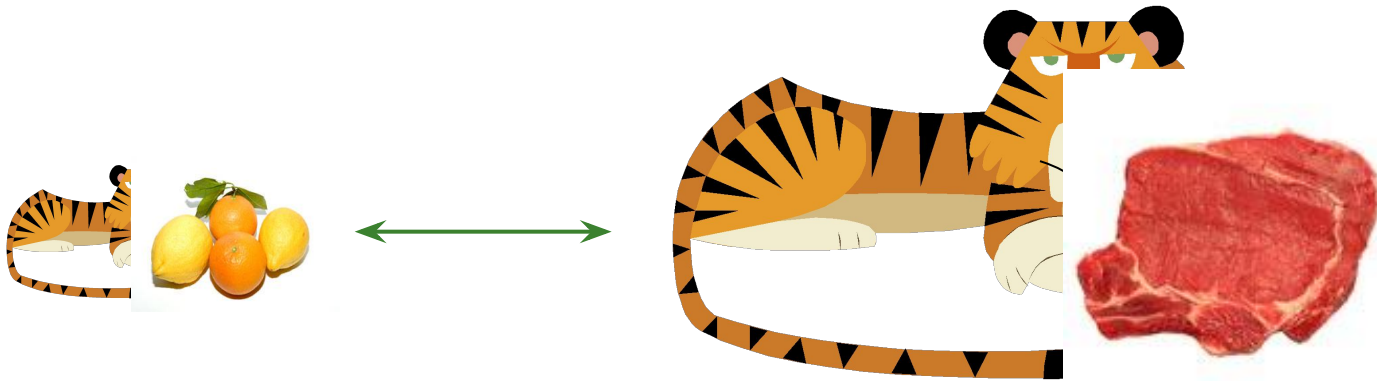
$$(P_{\text{ошибки}} \sim 0,05k)$$

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

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 (насколько примерно?)** одинаковы.

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



В t-тесте сходство выборочных средних оценить легко – просто посчитать разность. Но с 3-мя (4, 5...) группами так не получится!

ANOVA

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 оценки σ^2 и сравним их!

На этой идее основана ANOVA.

овоци	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

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

$$df_B = k-1$$

число групп -1 (3 - 1 = 2)

размер группы

MS_B – mean square between groups, оценка расстояния между средними в группах.

различия большие - H_0 не верна

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

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

число групп k

$$df_W = n_G - k$$

статистика:

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

Статистика критерия: F

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

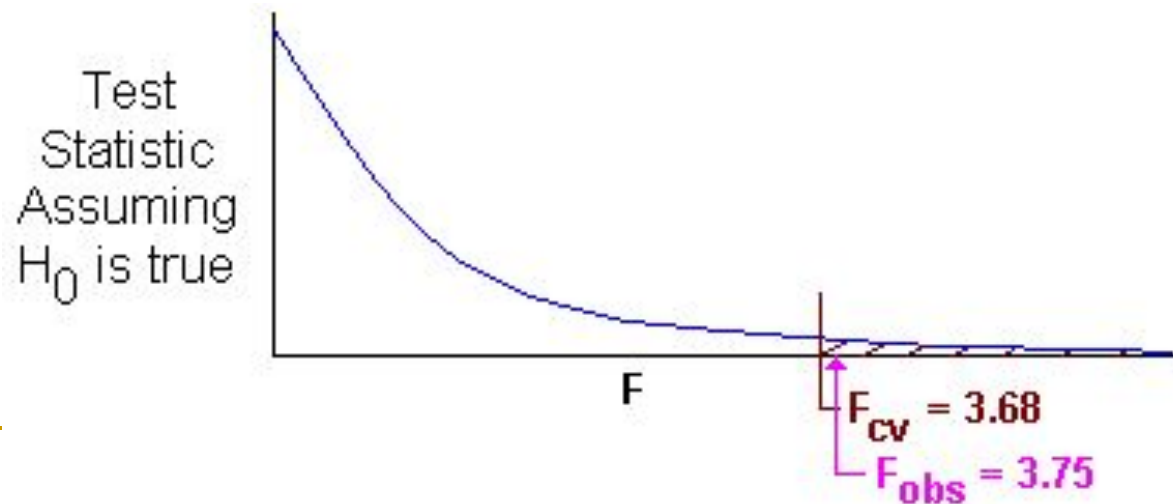
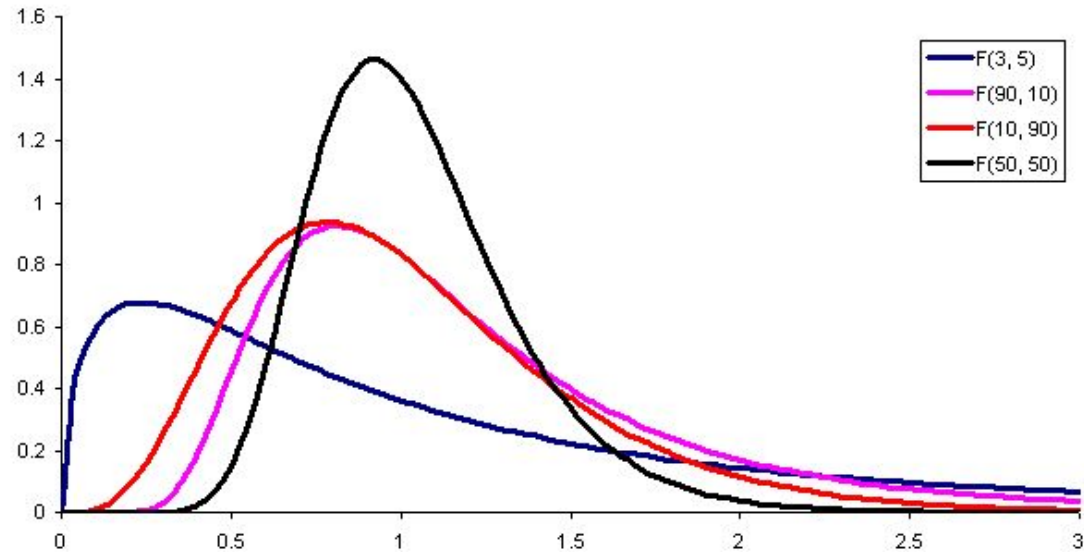
$$F = \frac{MS_B}{MS_W}$$

✓ Не соответствует общей формуле

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

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

Принципиально ненаправленный (двусторонний) тест



ANOVA

ANOVA table

<i>источник изменчивости</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
между	SS_B	df_B	MS_B	F
внутри	SS_W	df_W	MS_W	
общее	SS_T	df_T		

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

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

$$SS_T = \sum (X_{ij} - \bar{X}_G)^2 = \sum (X_{ij} - \bar{X}_j)^2 + \sum (\bar{X}_j - \bar{X}_G)^2 = SS_W + SS_B$$

$$df_T = df_W + df_B$$

$$MS_B = \frac{SS_B}{df_B} \quad MS_W = \frac{SS_W}{df_W}$$

ANOVA effect size

«Практическая значимость» результата:

$$1. \quad f = \frac{s_{\bar{X}}}{\sqrt{MS_W}}$$

$f = 0.1$ – маленький эффект

$f = 0.25$ – средний эффект

$f = 0.4$ – большой эффект

2. «доля объяснённой изменчивости»

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

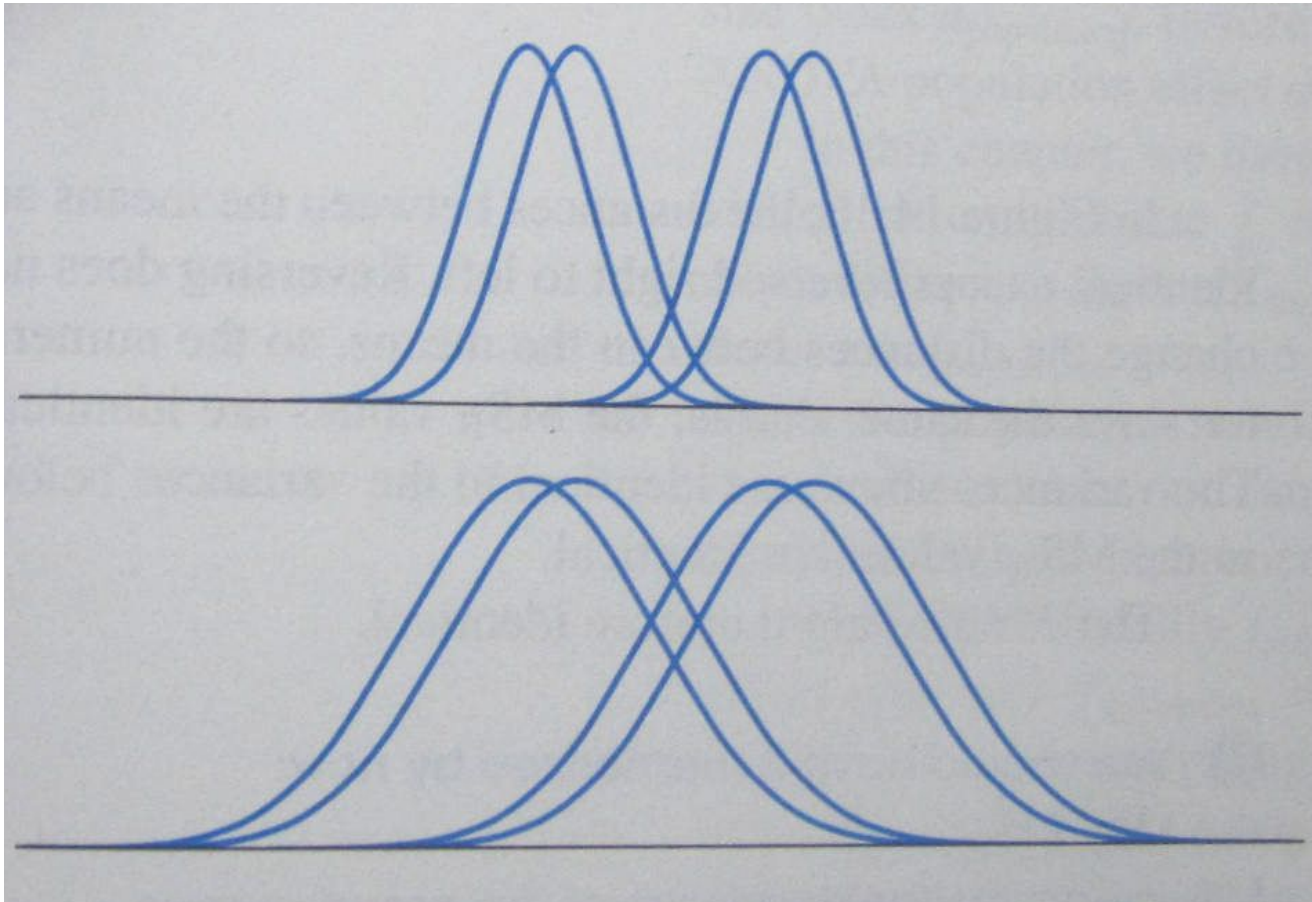
$R^2 = 0.01$ – маленький эффект

$R^2 = 0.06$ – средний эффект

$R^2 = 0.14$ – большой эффект

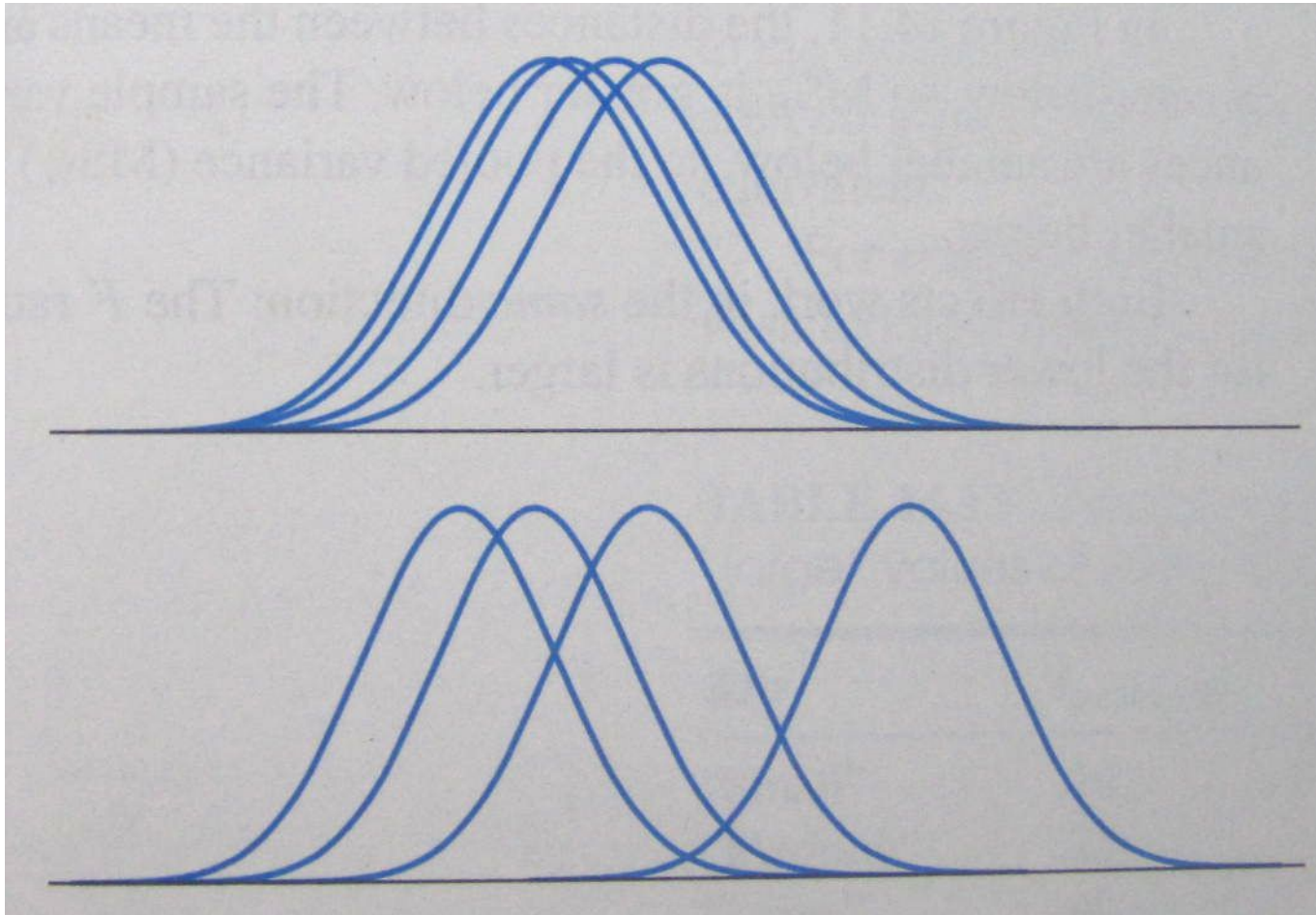
ANOVA

В каком случае значение F-статистики будет больше?



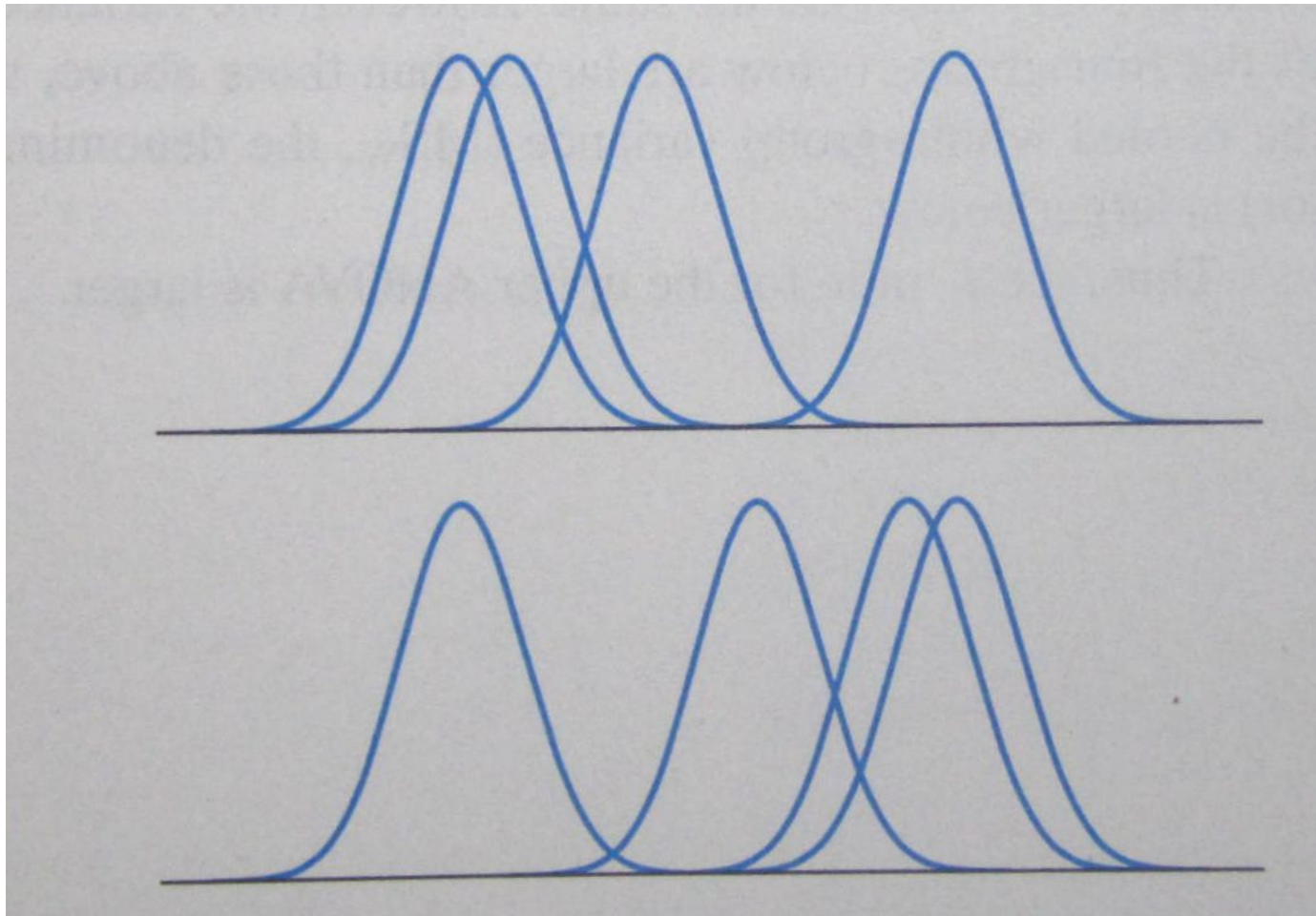
ANOVA

В каком случае значение F-статистики будет больше?



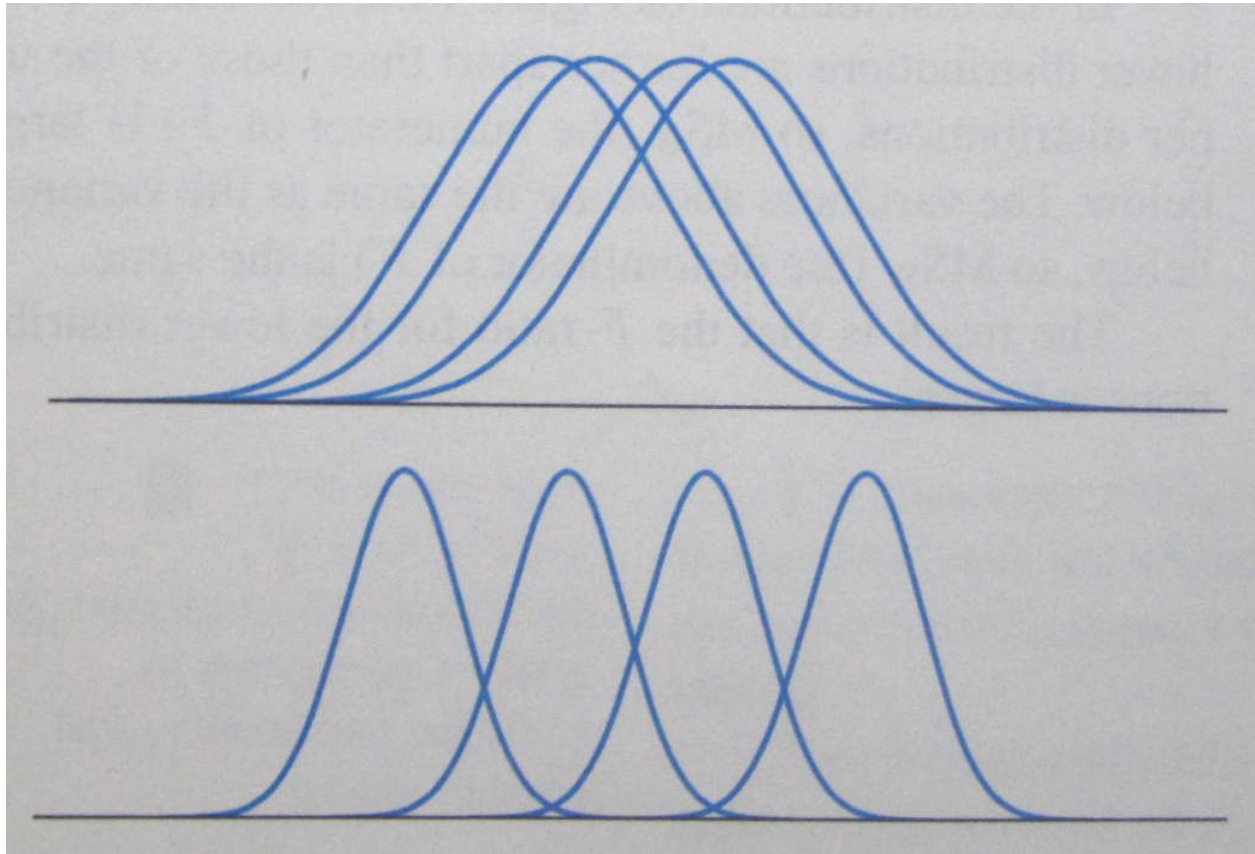
ANOVA

В каком случае значение F-статистики будет больше?



ANOVA

В каком случае значение F-статистики будет больше?



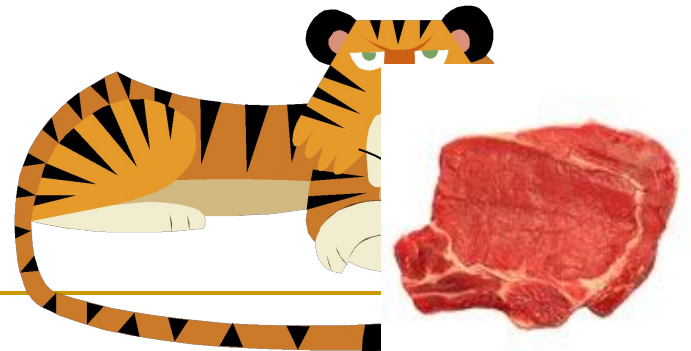
ANOVA

У нас только одна независимая (группирующая) переменная.

Такой анализ называется

One-way ANOVA

требования и ограничения – как в критерии Стьюдента



One-way ANOVA

Data: тигры и еда.sta (10v by 48c)

	1	2	3
	масса тела	тип еды	ПОЛ
1	151	овощи	самец
2	135	овощи	самец
3	137	овощи	самец
4	118	овощи	самец
5	132	овощи	самец
6	135	овощи	самец
7	131	овощи	самец
8	137	овощи	самец
9	121	овощи	самка
	140	овощи	самка
	152	овощи	самка
	133	овощи	самка
	151	овощи	самка
	132	овощи	самка
	139	овощи	самка
	96	овощи	самка
	108	фрукты	самец
	94	фрукты	самец

General ANOVA/MANOVA: тигры и еда

Quick

Type of analysis:

- One-way ANOVA
- Main effects ANOVA
- Factorial ANOVA
- Repeated measures ANOVA

Specification method:

- Quick specs dialog
- Analysis Wizard
- Analysis syntax editor

Multiple dependent variables can be specified for any type of analysis.

Use One-way ANOVA to analyze designs with a single categorical independent variable (factor).

OK

Cancel

Options

Open Data

SELECT CASES

Weighted moments

DF =

W-1 N-1

assumptions: нормальность, гомогенность

ANOVA Results 1: тигры и еда

Profiler | Custom tests | Residuals 1 | Residuals 2 | Matrix | Report
Summary | Means | Planned comps | Post-hoc | Assumptions

Variables: масса тела

Effect: "тип еды"

Homogeneity of variances/covariances

Cochran C, Hartley, Bartlett | Box M test (cov. matrix)
Levene's test (ANOVA)

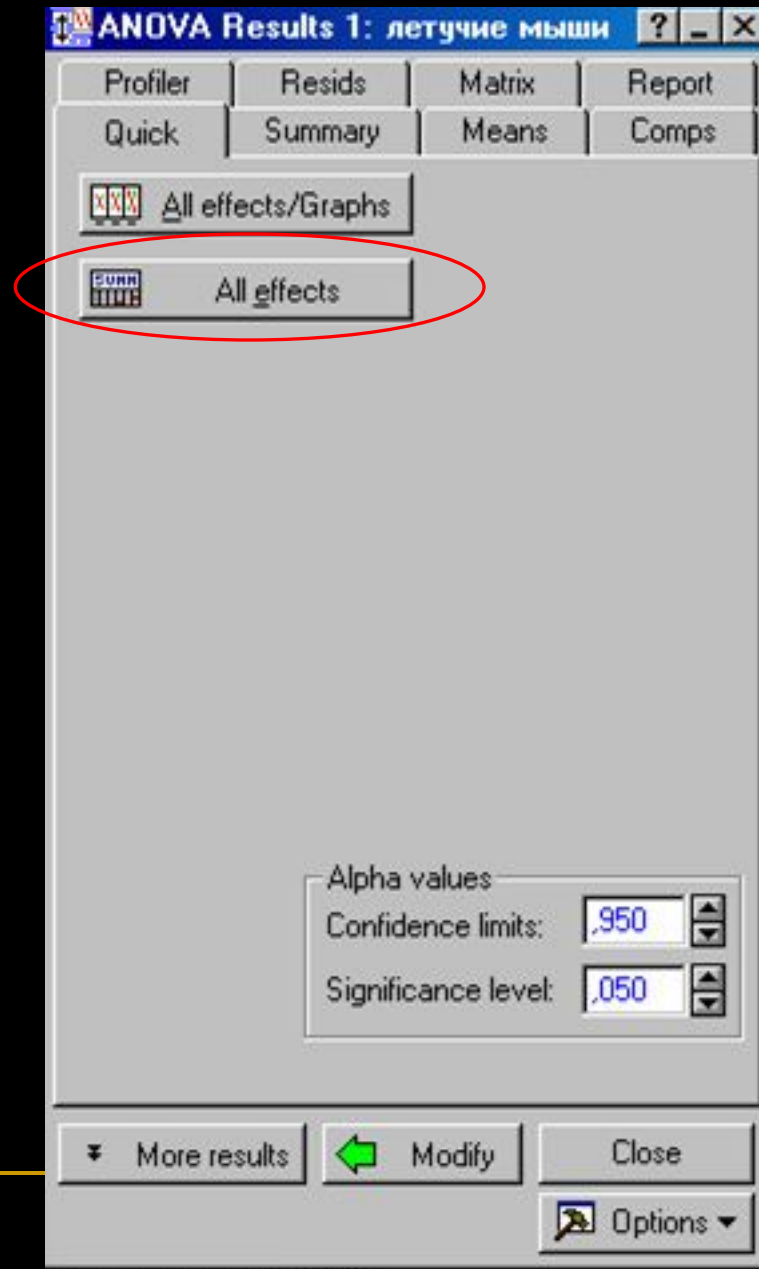
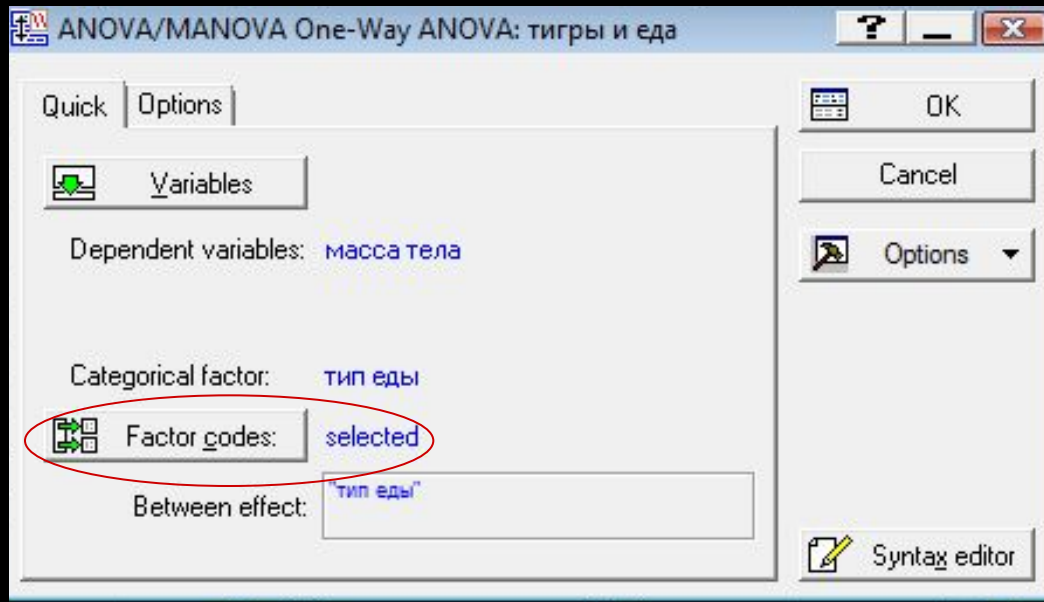
Distribution of variables within groups | Distribution of within-cell residuals

Histograms | Histograms
Normal p-p | Detrended | Normal p-p | Detrended
Scatterplots | Matrix | Scatterplot

Half-normal plot of z-transf. within corr's
Plot means vs. std. deviations | Variances

Less
Close
Modify
Options

One-way ANOVA



Significance for масса тела (тигры и еда)

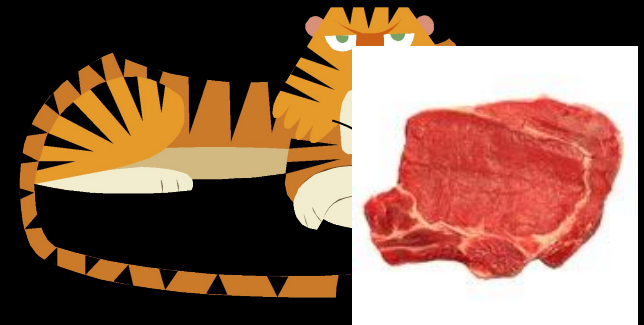
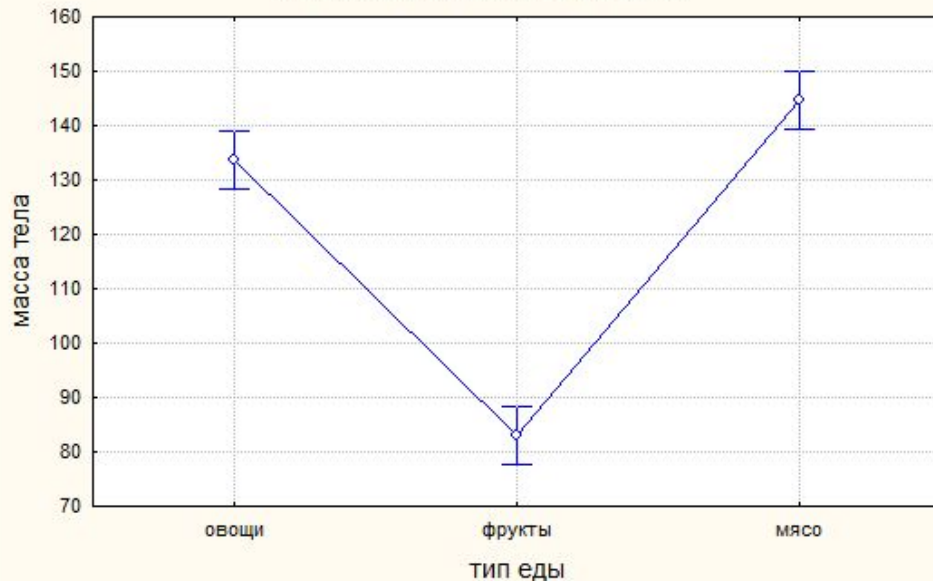
Univariate Tests of Significance for масса тела (тигры и еда)
Sigma-restricted parameterization
Effective hypothesis decomposition

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		

между группами

внутри групп

ТИП ЕДЫ; LS Means
Current effect: F(2, 45)=151,12, p=0,0000
Effective hypothesis decomposition
Vertical bars denote 0,95 confidence intervals



мы отвергаем H_0 .
тип еды влиял на
массу тигров

ANOVA post hoc tests

Сложная «омнибусная» гипотеза АНОВЫ:

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

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



Какая же из отдельных гипотез не верна?

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, но вводят поправки)
 3. Если различий нет, мы НЕ ИМЕЕМ ПРАВА ПРЕДПРИНИМАТЬ ДАЛЬНЕЙШИЙ АНАЛИЗ!
-

ANOVA post hoc tests

Поправка Бонферрони (*Bonferroni correction* для небольших k)

если мы хотим обеспечить уровень значимости α , то в каждом из k сравнений 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.

ANOVA post hoc tests

Тест Тьюки (Tukey HSD test)

Наиболее распространённый и рекомендуемый в литературе тест.

Рекомендуется для близких по размеру групп.

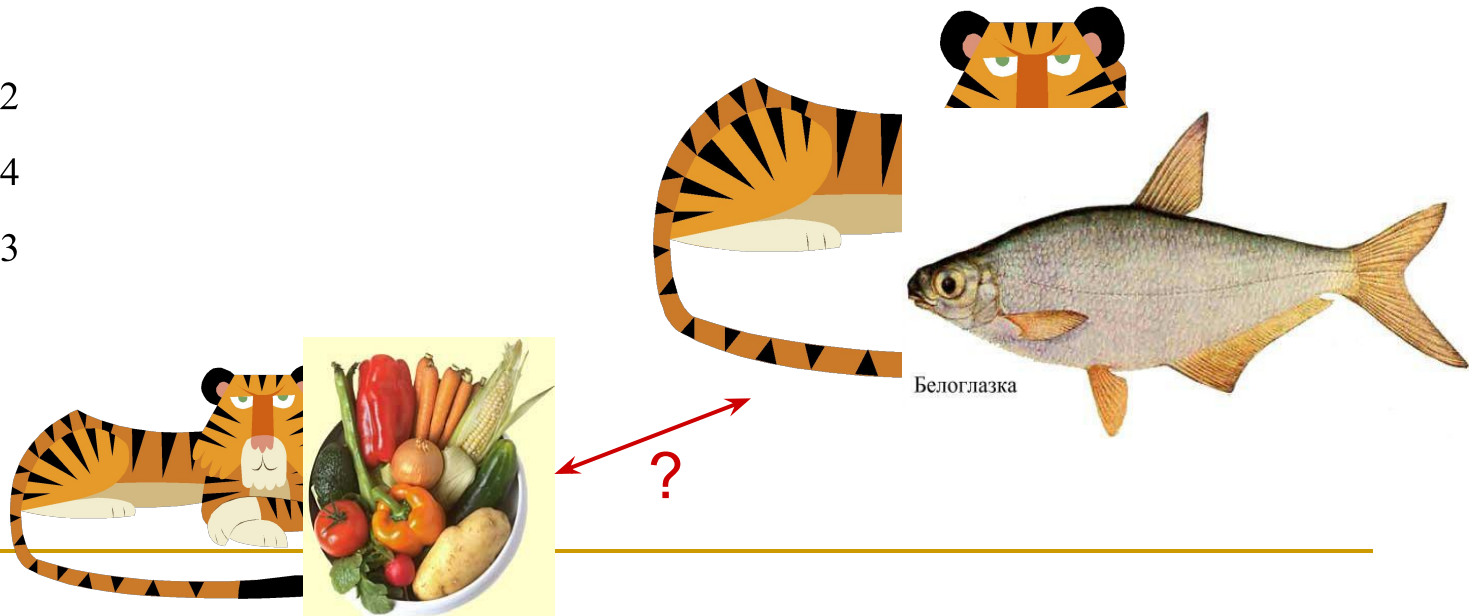
Проверяет только ПАРНЫЕ (но не комплексные) гипотезы.

$$H_{01} : \mu_1 = \mu_2$$

$$H_{02} : \mu_1 = \mu_4$$

$$H_{03} : \mu_1 = \mu_3$$

...



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.
-

Поправки для
множественных
сравнений и
сравнений с
контрольной
группой

Profiler | Custom tests | Residuals 1 | Residuals 2 | Matrix | Report
Summary | Means | Planned comps | Post-hoc | Assumptions

Effect: "пещера"

Dependent variables: длина крыла

Display

- Significant differences
- Homogeneous groups: .05
- Confidence intervals
- Critical ranges: .05

Error term

- Between error
- Within error
- Between; within; pooled
- MS: 0,00 df: 0

Fisher LSD | Bonferroni | Scheffe
Tukey HSD | Unequal N HSD

Range tests (multi-stage tests)

Newman-Keuls | Crit. ranges | Duncan's | Crit. ranges

Comparisons with a Control Group (CG)

Dunnett | < CG | > CG | <>CG | CG cell #: 1

Tukey HSD test; variable масса тела (тигры и еда)

Probabilities for Post Hoc Tests

Error: Between MS = 114,55, df = 45,000

Cell No.	тип еды	{1}	{2}	{3}
1	овощи	133,75	0,000129	0,016765
2	фрукты	0,000129		0,000129
3	мясо	0,016765	0,000129	



