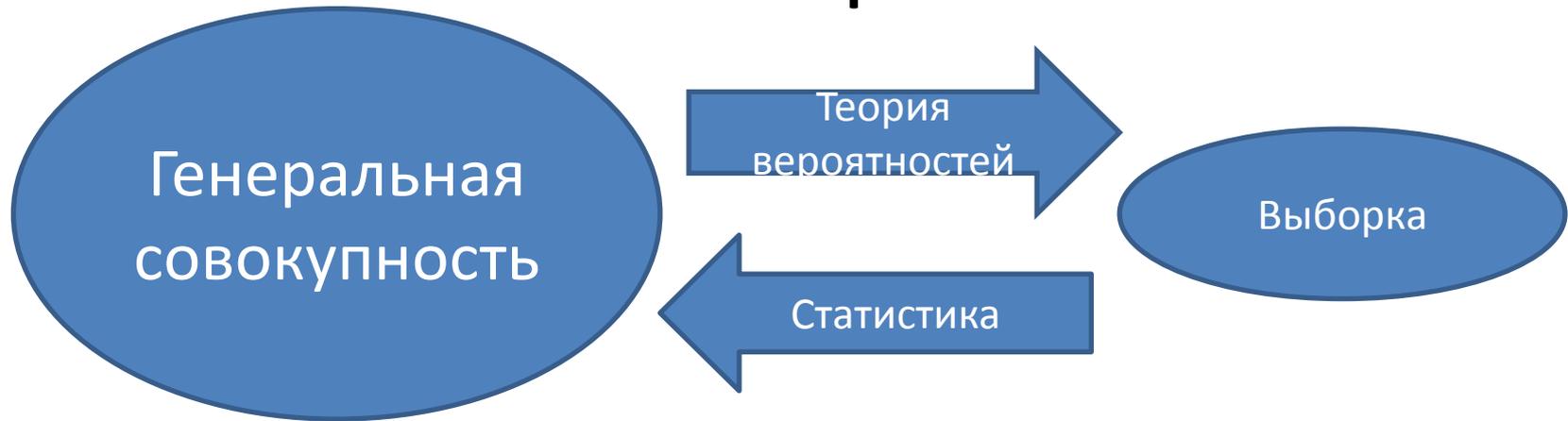


Статистические методы в школьных проектах по биологии

Трубянов Алексей Борисович

Генеральная совокупность и выборка



Генеральная совокупность (от лат. *generis* — *общий, родовой*) (в англ. терминологии — *population*) — совокупность всех объектов (единиц), относительно которых учёный намерен делать выводы при изучении конкретной проблемы.

Выборка или **выборочная совокупность** — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).

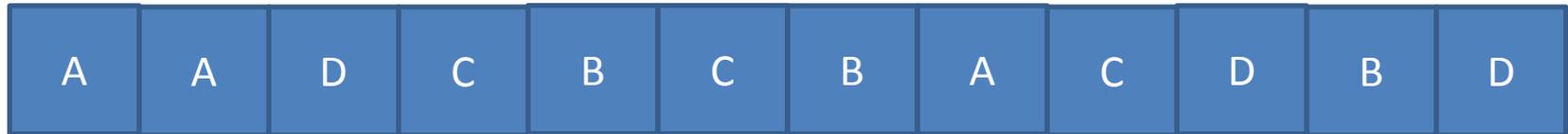
Репрезентативность выборки

Репрезентативность — соответствие характеристик выборки характеристикам популяции или генеральной совокупности в целом. Репрезентативность определяет, насколько возможно обобщать результаты исследования с привлечением определённой выборки на всю генеральную совокупность, из которой она была собрана.

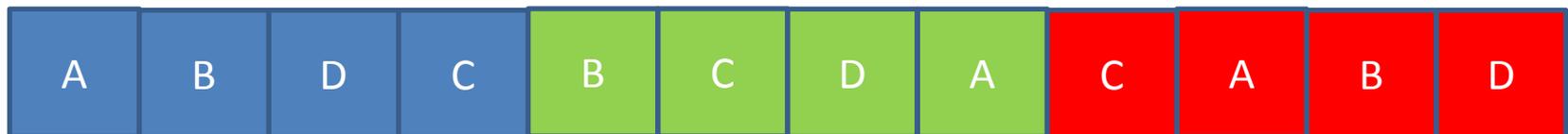
Candidate 1936	The <i>Literary Digest</i> poll	Results on election day
Alf Landon	55%	37%
Franklin Roosevelt	41%	61%

Фиксированная рандомизация

Простая рандомизация предполагает равновероятное распределение испытуемых в группы.



Блочная рандомизация – последовательность формируется из блоков заданной длины, внутри которых проводится случайное распределение



Изучение природных или «искусственных» популяций

«Индексы»:

Флуктуирующая асимметрия

Сухая масса единицы площади листа – LMA
(leaf mass per area)

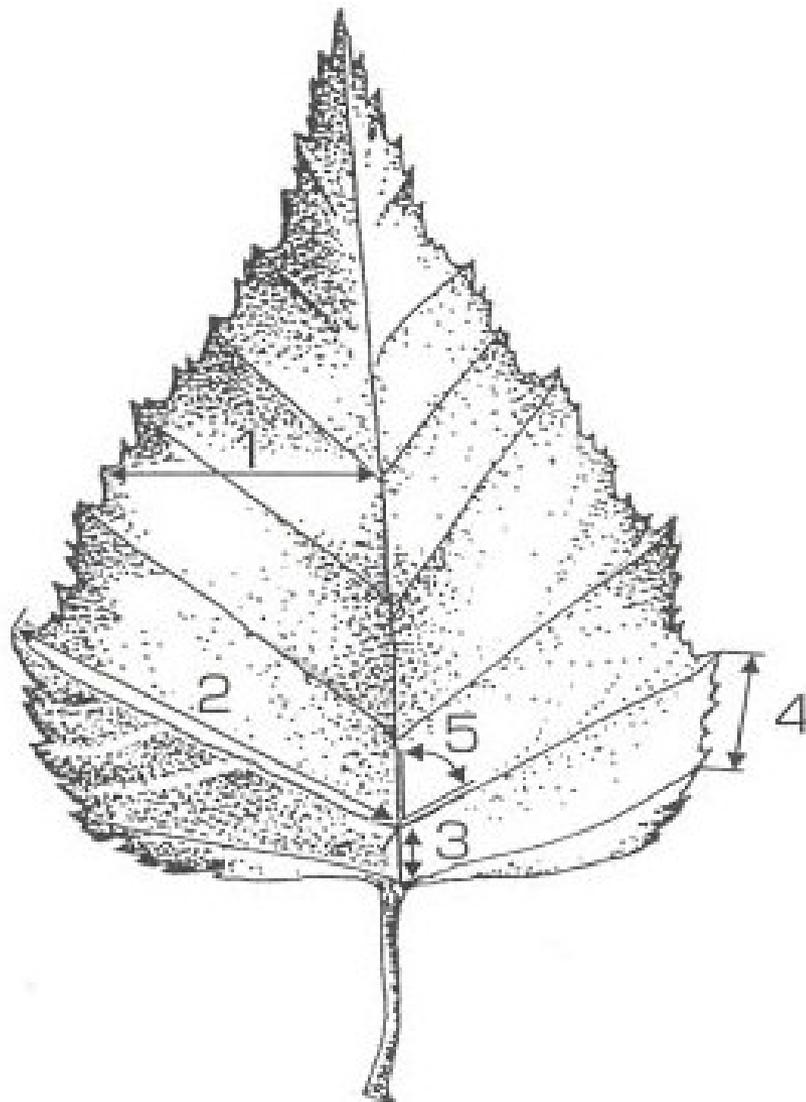
и т.д.

Морфологические признаки

Флуктуирующая асимметрия

Схема промеров листа березы повислой (*Betula pendula*)

(Методические рекомендации..., 2003)



Leaf Mass per Area (LMA) and Its Relationship with Leaf Structure and Anatomy in 34 Mediterranean Woody Species along a Water Availability Gradient

Enrique G. de la Riva, Manuel Olmo, Hendrik Poorter, José Luis Ubera, Rafael Villar

Published: February 11, 2016

<https://doi.org/10.1371/journal.pone.0148788>

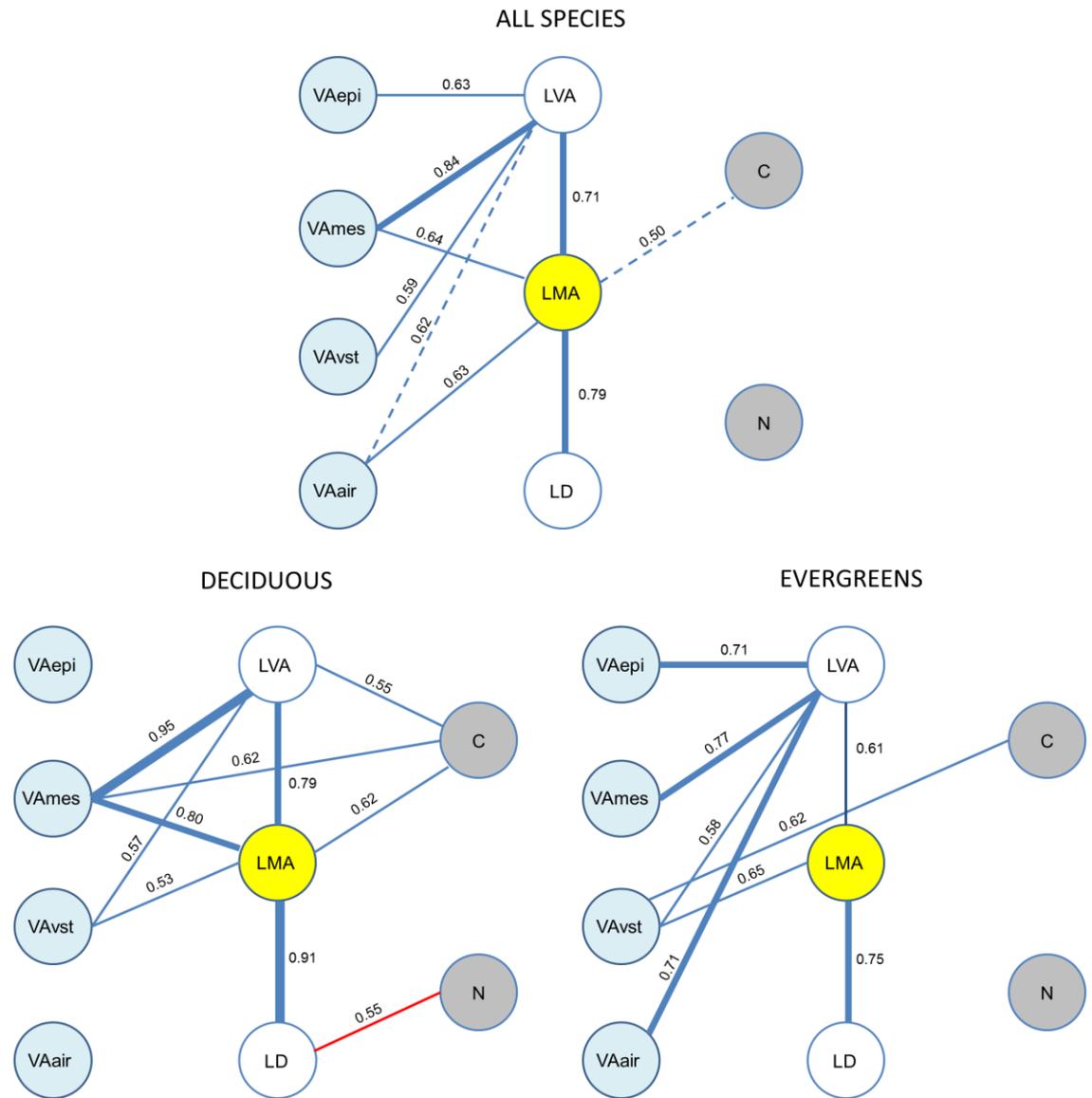
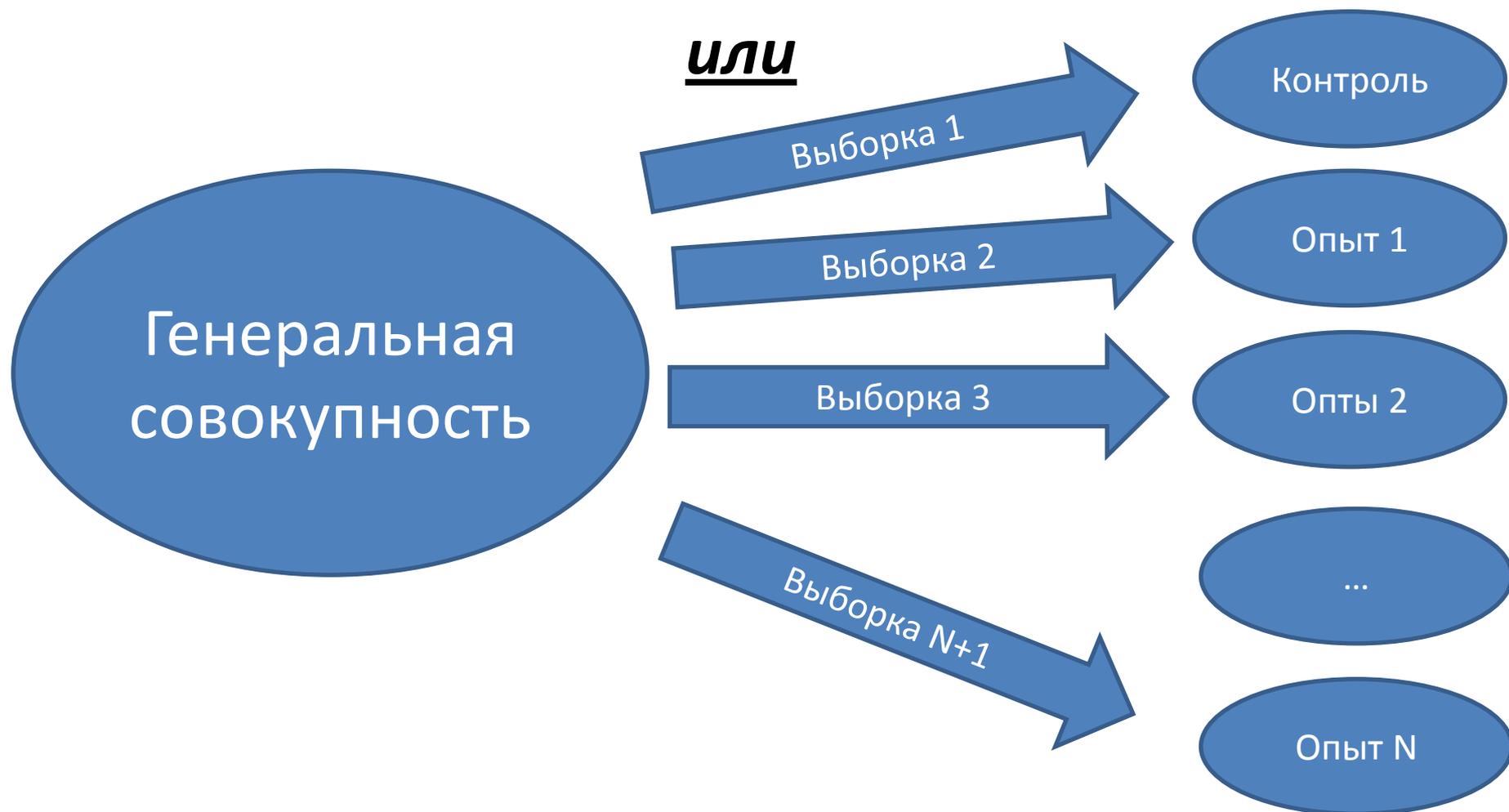


Fig 4. Correlation network for morphological (white), anatomical (blue) and chemical traits (grey), describing the interrelations with LMA (yellow).

Постановка лабораторного эксперимента

Опыт – Контроль



Статистическая гипотеза

Статистическая гипотеза – предположения о параметрах, виде или свойствах распределения.

Задача о леди, дегустирующей чай (Ronald Fisher, *The Design of Experiments*, 1935)



Вероятность указать правильную чашку $P = 0,5$

Ошибки I и II рода

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (ошибка второго рода)
	H_1	H_0 неверно отвергнута (ошибка первого рода)	H_0 верно отвергнута

Пример. H_0 : Пациент болен, H_1 : Пациент здоров

		Верная гипотеза	
		Пациент болен	Пациент здоров
Результат применения критерия	Назначить лечение	Лечить больного	Лечить здорового (ошибка второго рода)
	Не назначать лечение	Не лечить больного (ошибка первого рода)	Не лечить здорового

Статистическая гипотеза

H_0 - проверяемая (основная, нулевая) гипотеза

H_1, H_2, \dots - альтернативные гипотезы

Уровень значимости – мера согласия
экспериментальных данных с проверяемой
гипотезой

$\alpha=0,05$

$\alpha=0,01$

$\alpha=0,001$

Статистические методы

Сравнение двух групп

- Критерий Манна-Уитни
- Критерий Вилкоксона (парные выборки – «до» - «после» на одних и тех же объектах)

Сравнение более двух групп

- Критерий Краскала-Уоллиса

Критерий Манна-Уитни

Исследовалась зависимость размера листа кустов ежевики в разных условиях затенения

	Ширина листа/ см							
На солнце	6.0	4.8	5.1	5.5	4.1	5.3	4.5	5.1
В тени	6.5	5.5	6.3	7.2	6.8	5.5	5.9	5.5

H_0 : ширина листа одинаковая

H_A : ширина листа различная

Критерий Манна-Уитни

Исследовалась зависимость размера листа кустов ежевики в разных условиях затенения

	Ширина листа/ см							
На солнце	6.0 (12)	4.8 (3)	5.1 (4.5)	5.5 (8.5)	4.1 (1)	5.3 (6)	4.5 (2)	5.1 (4.5)
В тени	6.5 (14)	5.5 (8.5)	6.3 (13)	7.2 (16)	6.8 (15)	5.5 (8.5)	5.9 (11)	5.5 (8.5)

Объемы выборок $N_1 = 8, N_2 = 8$

Сумма рангов $R_1 = 41.5, R_2 = 94.5$

Статистика Манна-Уитни $U_1 = N_1 \cdot N_2 + \frac{N_1(N_1+1)}{2} - R_1$

$U_2 = N_1 \cdot N_2 + \frac{N_2(N_2 + 1)}{2} - R_2$

$U_1 = 58.5$ $U_2 = 5.5$ Выбираем меньшую и сравниваем с табличным

Критерий Манна-Уитни

Исследовалась зависимость размера листа кустов ежевики в разных условиях затенения

	Ширина листа/ см							
На солнце	6.0 (12)	4.8 (3)	5.1 (4.5)	5.5 (8.5)	4.1 (1)	5.3 (6)	4.5 (2)	5.1 (4.5)
В тени	6.5 (14)	5.5 (8.5)	6.3 (13)	7.2 (16)	6.8 (15)	5.5 (8.5)	5.9 (11)	5.5 (8.5)

$$U_2 = 5.5$$

$$U_{\text{табл.}} = 13$$

Правило: если $U > U_{\text{табл.}}$, то принимаем нулевую гипотезу

если $U \leq U_{\text{табл.}}$, то принимаем альтернативную гипотезу

Листья ежевики, произрастающие в условиях затенения крупнее, чем на солнце.

Критерий Вилкоксона

Ф. Эшли и соавт. сравнивали два средства для предупреждения образования зубного налета: хлоргексидин и хлорид аммония. Каждый из участников исследования в течение 48 часов полоскал рот одним из средств, после чего налет оценивали визуально. Через некоторое время опыт повторяли с другим средством (очередность определялась случайным образом). Эффективно ли полоскание хлоридом аммония?

Критерий Вилкоксона

Участник	1	2	3	4	5	6	7	8	9	10
Хлорид аммония	32	60	25	45	65	60	68	83	120	110
Хлоргексидин	14	39	24	13	9	3	10	14	1	36
Разница	18 (2)	21 (3)	1 (1)	32 (4)	56 (5)	57 (6)	58 (7)	69 (8)	119 (10)	74 (9)

Ранжируем без учета знака!

Статистика Вилкоксона – сумма рангов для отрицательных разностей минус сумма рангов для положительных разностей, затем берем модуль получившегося числа

$$W=55$$

Критерий Вилкоксона

$W=55$

Определяем табличное значение $W_{\text{табл}} = 8$

Правило: если $W > W_{\text{табл}}$, то принимаем альтернативную гипотезу

если $W \leq W_{\text{табл}}$, то принимаем нулевую гипотезу

Полоскание хлоргексидином значительно более эффективно для предупреждения образования зубного налета в отличие от хлорида аммония

Критерий Краскала-Уоллиса

Испытывали четыре диеты. Приведены потери веса испытуемых, кг.

Диета А	2.70	3.60	1.35	2.25	2.70
Диета Е	4.95	5.85	6.75		
Диета В	9.45	9.00	7.65	7.20	
Диета Г	2.25	4.05	4.50	3.15	3.15

Критерий Краскала-Уоллиса

Испытывали четыре диеты. Приведены потери веса испытуемых, кг.

Диета А 2.70(5) 3.60(9) 1.35(1) 2.25(2.5) 2.70(5) 2.70(5)

Диета Е 4.95(12) 5.85(13) 6.75(14)

Диета В 9.45(18) 9.00(17) 7.65(16) 7.20(15)

Диета Г 2.25(2.5) 4.05(10) 4.50(11) 3.15(7.5) 3.15(7.5)

Объемы выборок $N_1 = 6, N_2 = 3, N_3 = 4, N_4 = 5$

Общий объем выборки $N = 18$

Средние ранги $\bar{R}_1 = 4.58, \bar{R}_2 = 13, \bar{R}_3 = 16.5, \bar{R}_4 = 7.7$

Статистика Краскала-Уоллиса

$$H = \frac{12}{N(N+1)} \left(N_1 \bar{R}_1^2 + \dots + N_4 \bar{R}_4^2 \right) - 3(N+1) = 13.82$$

Критерий Краскала-Уоллиса

$H=13.82$

Табличное значение можно подсчитать в Excel

The screenshot shows the Microsoft Excel interface with the 'Формулы' (Formulas) ribbon active. The 'Вставить функцию' (Insert Function) dialog box is open, displaying the 'Статистические' (Statistical) category. The 'ХИ2.ОБР.ПХ' (CHI2.OBR.PX) function is selected. A tooltip for this function is visible, stating: 'ХИ2.ОБР.ПХ(вероятность;степени_свободы) Возвращает значение обратное к правосторонней вероятности распределения хи-квадрат. Для получения дополнительных сведений нажмите клавишу F1.'

Критерий Краскала-Уоллиса

$$H=13.82$$

Табличное значение можно подсчитать в Excel

Степени свободы = число групп минус 1

The screenshot shows the 'Arguments of Function' dialog box for the CHI2.OBR.PX function. The title bar reads 'Аргументы функции'. The function name 'ЧИ2.ОБР.ПХ' is displayed at the top. There are two input fields: 'Вероятность' (Probability) with a value of '0,05' and 'Степени_свободы' (Degrees of Freedom) with a value of '3'. Each field has a small icon to its right. Below the fields, the calculated result is shown as '= 7,814727903'. A descriptive text block states: 'Возвращает значение обратное к правосторонней вероятности распределения хи-квадрат.' and 'Степени_свободы число степеней свободы - число от 1 до 10^10, исключая 10^10.'. At the bottom, the 'Значение:' (Value) is '7,814727903'. There is a link 'Справка по этой функции' (Help for this function) and two buttons: 'ОК' and 'Отмена' (Cancel).

$$H_{\text{табл}} = 7.81$$

Критерий Краскала-Уоллиса

$$H=13.82$$

$$H_{\text{табл}} = 7.81$$

Правило: если $H > H_{\text{табл}}$, то принимаем альтернативную гипотезу

если $H \leq H_{\text{табл}}$, то принимаем нулевую гипотезу

Диеты различаются!

Но какая диеты от какой отличается???

Попарные сравнения критерием Манна-Уитни!!!

Поправка на множественные сравнения!!!

Благодарю за внимание!