

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное научное учреждение  
«Федеральный научный центр биологических систем и агротехнологий  
Российской академии наук»  
(ФГБНУ ФНЦ БСТ РАН)

## МЕТОДИЧЕСКИЕ УКАЗАНИЯ

ПО ДИСЦИПЛИНЕ

*«Б1.О.09 Информационные технологии в науке и производстве»*

Уровень высшего образования

МАГИСТРАТУРА

Направление подготовки

36.04.02 Зоотехния

(код и наименование направления подготовки)

Питание сельскохозяйственных животных

и кормопроизводство

(наименование направленности (профиля) образовательной программы)

Квалификация

Магистр

Форма обучения

Заочная

Методические указания рассмотрены и утверждены на заседании отдела кормления сельскохозяйственных животных и технологии кормов им. проф. С.Г. Леушина, «10» января 2020 г., протокол № 1

Зав.отделом  Г.К. Дускаев

Год набора 2020

## Содержание

Введение	3
Практическое занятие № 1	5
Практическое занятие № 2	8

## Введение

Методические указания посвящены изучению теоретических основ и практических инструментов, применяемых в информационных технологиях в науке и производстве по 36.04.02 Зоотехния при изучении дисциплины «Информационные технологии в науке и производстве».

Практическое занятие – одна из ключевых форм учебного занятия, направленная на развитие самостоятельности студентов, приобретение умений и формирования общих и профессиональных компетенций. Выполнение практических заданий способствует углубленному пониманию и усвоению пройденного материала профессионального модуля, развитию мышления и творческого потенциала.

Целью практических занятий по учебным дисциплинам является закрепление студентами пройденного теоретического материала и выработка навыков самостоятельной профессиональной и научно-исследовательской деятельности.

Задачи практических занятий обусловлены необходимостью получения обучающимися знаний, умений, навыков согласно требованиям ФГОС ВО, на основе которых формируются соответствующие профессиональные компетенции. Ведущей дидактической целью практических занятий является формирование умений, необходимых в последующей учебной деятельности по общепрофессиональным, профессиональным дисциплинам.

Методические указания по выполнению работ на практических занятиях разработаны по разделам рабочей программы. Подготовка к практическим работам заключается в самостоятельном изучении теории по рекомендуемой литературе, предусмотренной рабочей программой.

Для эффективного выполнения заданий студенты должны знать теоретические материалы и уметь применять эти знания для приобретения навыков при выполнении практических заданий.

Условия и порядок выполнения работы:

1 Изучить методические рекомендации по выполнению практических занятий.

2 Изучить содержание заданий и приступить к выполнению.

3 Консультацию по выполнению работы получить у преподавателя.

4 Работа оценивается в целом, по итогам выполнения работы выставляется оценка.

Защита проводится путем индивидуальной беседы или выполнения зачетного задания.

Процесс выполнения практических заданий направлен на формирование компетенции ОПК-4 – способность использовать в профессиональной деятельности методы решения задач с использованием современного оборудования при разработке новых технологий и использовать современную профессиональную методологию для проведения экспериментальных исследований и интерпретации их результатов.

## **Практическое занятие № 1. Проверка нормальности распределения**

**Цель занятия:** научиться выполнять проверку нормальности выборочных данных.

### **Краткие теоретические сведения**

Одним из самых часто встречающихся вопросов, с которым сталкивается исследователь – применять ли параметрические процедуры или необходимы непараметрические аналоги.

Параметрические процедуры применяются в том случае, когда известно, что исследуемый массив данных имеет нормальное распределение.

Нормальное распределение приобрело очень большую популярность, в связи с разработкой и доказательством центральной предельной теоремы.

Центральная предельная теорема, в теории вероятностей, - это теорема, согласно которой практически любой набор независимых и случайных переменных быстро сходится к нормальному распределению; при этом выборочное среднее и отклонения для выборочных данных будут справедливы и для всей возможной совокупности данных (так называемой генеральной совокупности).

Потому в большинстве работ, особенно если количество данных достаточно велико, исследователи пренебрегают проверкой на нормальность распределения, априори предполагая, что нормальность присутствует.

Однако большинство отечественных и зарубежных статистиков считают, что нормальность необходимо проверять вне зависимости от количества наблюдений [6, 7, 8].

При проверке нормальности стоит учитывать множество различных факторов. Как известно, достаточно трудно определить нормальность, имея небольшой размер выборки. Необходимо учесть также, что данные могут быть искажены. Проверка же нормальности дает ответ на вопрос о том, какие процедуры использовать: параметрические или непараметрические [9].

Существует множество критериев для проверки нормальности эмпирических данных. Но, согласно А. И. Орлову «... мировая наука сошлась на том, что надо использовать критерий Шапиро-Уилка» [10].

Существует ограничение на применение критерия Шапиро-Уилка – не более 50 эмпирических данных. Поскольку медико-биологические исследования обходятся, как правило, много меньшим количеством исследований, данный критерий является наиболее удовлетворительным для поставленных выше целей.

Статистика критерия имеет вид:

$$W = \frac{1}{S^2} \left[ \sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i) \right]^2, \quad (9)$$

где

$$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2, \quad (10)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (11)$$

При этом, если  $n$  четное,  $k=n/2$ , если  $n$  нечетное,  $k=(n-1)/2$ .

Коэффициенты  $a_{n-i+1}$  берутся из таблиц (или рассчитывается с приемлемой точностью по специальным формулам). Из таблиц также берется критическое значение  $W(a)$ .

Если  $W \geq W(a)$ , нулевую гипотезу не бракуют, т.е. распределение считают нормальным.

Проверка нормальности распределения должна быть произведена вне всякой зависимости от исходных целей исследования.

### **Методические указания к работе**

Проверку нормальности необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

## Задание

Необходимо выполнить проверку нормальности выборки, для чего нужно в отдельной таблице вывести следующие данные:

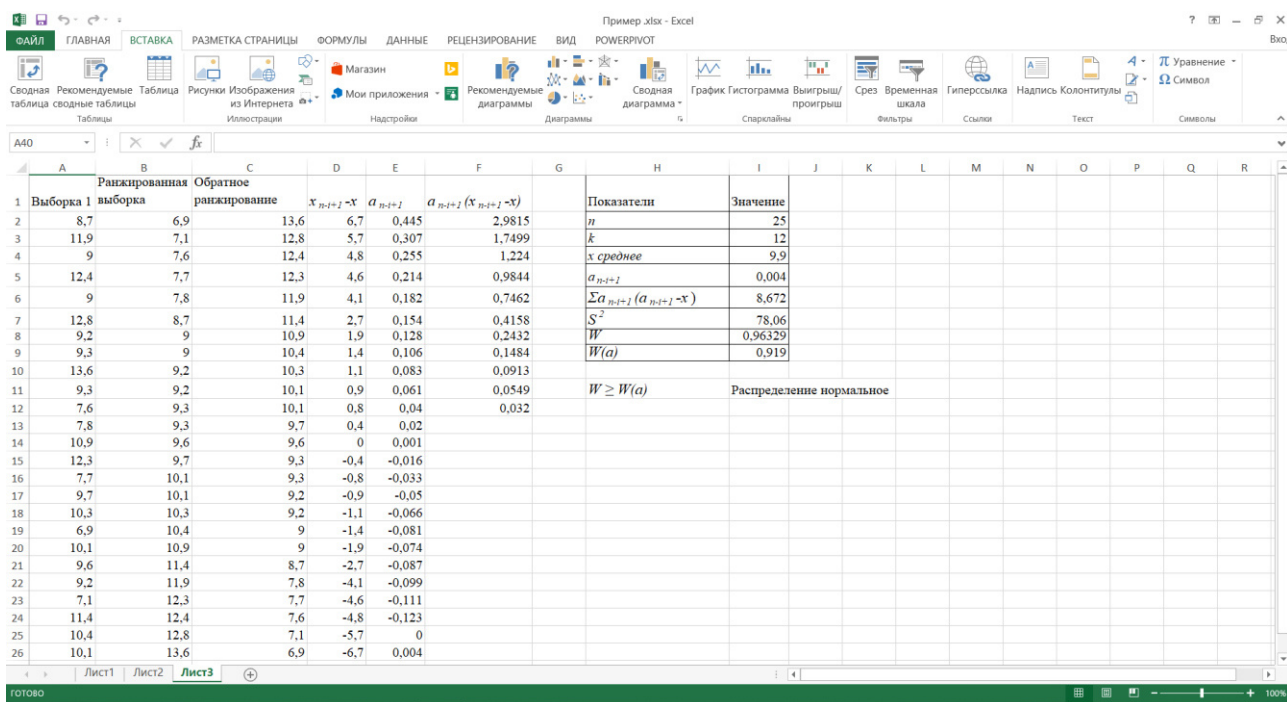
- расчет или значения всех показателей, входящих в формулу (9).
- табличные (или рассчитанные значения) критерия  $W(a)$ .

Затем необходимы выполнить сравнение фактического и табличного значения критерия и сделать вывод о принадлежности или не принадлежности распределения выборки нормальному закону.

Выполнить расчеты не менее, чем на трех выборках.

## Пример

Все параметры внесены в таблицу, отображенную на рисунке 3.



Выборка 1	Ранжированная выборка	Обратное ранжирование	$x_{n+1} - x$	$a_{n+1}$	$a_{n+1} (x_{n+1} - x)$	Показатели	Значение
8,7	6,9	13,6	6,7	0,445	2,9815	$n$	25
11,9	7,1	12,8	5,7	0,307	1,7499	$k$	12
9	7,6	12,4	4,8	0,255	1,224	$x$ среднее	9,9
12,4	7,7	12,3	4,6	0,214	0,9844	$a_{n+1}$	0,004
9	7,8	11,9	4,1	0,182	0,7462	$\sum a_{n+1} (a_{n+1} - x)$	8,672
12,8	8,7	11,4	2,7	0,154	0,4158	$S^2$	78,06
9,2	9	10,9	1,9	0,128	0,2432	$W$	0,96329
9,3	9	10,4	1,4	0,106	0,1484	$W(a)$	0,919
13,6	9,2	10,3	1,1	0,083	0,0913		
9,3	9,2	10,1	0,9	0,061	0,0549	$W \geq W(a)$	Распределение нормальное
7,6	9,3	10,1	0,8	0,04	0,032		
7,8	9,3	9,7	0,4	0,02			
10,9	9,6	9,6	0	0,001			
12,3	9,7	9,3	-0,4	-0,016			
7,7	10,1	9,3	-0,8	-0,033			
9,7	10,1	9,2	-0,9	-0,05			
10,3	10,3	9,2	-1,1	-0,066			
6,9	10,4	9	-1,4	-0,081			
10,1	10,9	9	-1,9	-0,074			
9,6	11,4	8,7	-2,7	-0,087			
9,2	11,9	7,8	-4,1	-0,099			
7,1	12,3	7,7	-4,6	-0,111			
11,4	12,4	7,6	-4,8	-0,123			
10,4	12,8	7,1	-5,7	0			
10,1	13,6	6,9	-6,7	0,004			

Рисунок 3 – Пример решения задачи 3

Все необходимые показатели рассчитаны.

## **Практическое занятие № 2. Критерии различия между группами в несвязных выборках (параметрический)**

**Цель занятия:** научиться выполнять расчет параметрического критерия различия между группами в несвязанных выборках.

### **Краткие теоретические сведения**

Рассмотрим исследование несвязных групп при помощи параметрических процедур. Ряд исследователей-статистиков предлагают предпочесть критерию Стьюдента использовать критерий Крамера-Уэлча [11], в рамках проверки гипотезы о равенстве математического ожидания двух несвязных групп случайных величин при попарном сравнении, основанный на статистике

$$T = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}} \quad (12)$$

где  $n, m$  – количество значений в первой и второй выборке,

$s_x^2, s_y^2$  – выборочные дисперсии соответствующих выборок.

Значение модуля статистики  $T$  Крамера-Уэлча надо сравнивать с граничным значением, равным 1,96. Если значение  $T > 1,96$ , то считается, что выборки значимо различаются.

### **Методические указания к работе**

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### **Задание**

Необходимо провести сравнение выборок путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (12).

Выполнить расчеты не менее, чем на трех выборках.



## Пример

Все параметры внесены в таблицу, отображенную на рисунке 4.

Выборка 1	Выборка 2	Показатели	Значение
8,7	8,9	$n$	25
11,9	6,9	$m$	25
9	5,2	$\bar{x}$ среднее	9,9
12,4	9,6	$\bar{y}$ среднее	8,4
9	7	$Sx^2$	3,3
12,8	11,4	$Sy^2$	5,1
9,2	5,8	$T$	2,43822
9,3	12,4		
13,6	7	$T > 1,96$	Выборки неоднородны
9,3	6,5		
7,6	8,3		
7,8	4,6		
10,9	11,2		
12,3	8,2		
7,7	7,7		
9,7	6,1		
10,3	12,3		
6,9	6,7		
10,1	12,2		
9,6	10,6		
9,2	8,3		
7,1	7,1		
11,4	9,4		
10,4	8,9		
10,1	8,8		

Рисунок 4 – Пример решения задачи 4

Все необходимые показатели рассчитаны.

## **Критерий различия между группами в несвязных выборках (непараметрический)**

**Цель занятия:** научиться выполнять расчет непараметрического критерия различия между группами в несвязанных выборках.

### **Краткие теоретические сведения**

Приведем непараметрические аналоги решения задачи выявления различия в несвязных группах. Самым распространенным и простым способом непараметрической проверки различия между группами в независимых выборках является U критерий Манна-Уитни [15].

Для того, чтобы применить U-критерия Манна-Уитни необходимо произвести следующие действия.

1 Составить общий ранжированный ряд из обеих изучаемых выборок, расставив конкретные их значения по степени нарастания признака и присвоив меньшему значению меньший ранг. Общее количество рангов равно:

$$N = n_1 + n_2 \quad (13)$$

где  $n_1$  – количество значений в первой, а  $n_2$  – количество значений во второй выборке.

2 Разделить полученный общий ранжированный ряд на два, которые будут состоять соответственно из элементов первой и второй выборок.

3 Отдельно подсчитать сумму рангов, пришедшихся на долю каждой из изучаемых выборок.

4 Определить наибольшую из двух полученных ранговых сумм ( $T_x$ ), которая соответствует выборке с  $n_x$  элементами.

5 Далее определить значение U-критерия Манна – Уитни по формуле:

$$U = n_1 n_2 + \frac{n_x(n_x+1)}{2} - T_x \quad (14)$$

б Далее, согласно методам проверки гипотез, определить, значимо ли значение  $U$  для соответствующих  $n_1$  и  $n_2$  [16].

Стоит отметить, что при сравнении полученного эмпирического значения критерия с критическим, важно помнить, что гипотеза о значимости различий принимается тогда, когда эмпирическое значение *меньше* табличного.

### Методические указания к работе

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### Задание

Необходимо провести сравнение выборок путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (14). Выполнить расчеты не менее, чем на трех выборках.

### Пример

Все параметры внесены в таблицу, отображенную на рисунке 5.

Выборка 1	Выборка 2	Ранжированная совокупная выборка	Ранг	Принадлежность	Ранги по $n_1$	Ранги по $n_2$	Показатели	Значение
8,7	8,9	4,6	1	2	0	1	$n_1$	25
11,9	6,9	5,2	2	2	0	2	$n_2$	25
9	5,2	5,8	3	2	0	3	$\Sigma x_{n1}$	771,5
12,4	9,6	6,1	4	2	0	4	$\Sigma x_{n2}$	518,5
9	7	6,5	5	2	0	5	$T_x$	771,5
12,8	11,4	6,7	6	2	0	6	$U$	178,5
9,2	5,8	6,9	7,5	1	7,5	0	$U_{крит}$	227,0
9,3	12,4	6,9	7,5	2	0	7,5		
13,6	7	7	9,5	2	0	9,5	$U < U_{крит}$	Выборки неоднородны
9,3	6,5	7	9,5	2	0	9,5		
7,6	8,3	7,1	11,5	1	11,5	0		
7,8	4,6	7,1	11,5	2	0	11,5		
10,9	11,2	7,6	13	1	13	0		
12,3	8,2	7,7	14,5	1	14,5	0		
7,7	7,7	7,7	15,5	2	0	15,5		
9,7	6,1	7,8	17	1	17	0		
10,3	12,3	8,2	18	2	0	18		
6,9	6,7	8,3	19,5	2	0	19,5		
10,1	12,2	8,3	19,5	2	0	19,5		
9,6	10,6	8,7	21	1	21	0		
9,2	8,3	8,8	22	2	0	22		
7,1	7,1	8,9	23,5	2	0	23,5		
11,4	9,4	8,9	23,5	2	0	23,5		
10,4	8,9	9	25,5	1	25,5	0		
10,1	8,8	9	25,5	1	25,5	0		

Рисунок 5 – Пример решения задачи 5

Все необходимые показатели рассчитаны.

## **Критерий различия между группами в связанных выборках (параметрический)**

**Цель занятия:** научиться выполнять расчет параметрического критерия различия между группами в связанных выборках.

### **Краткие теоретические сведения**

Если имеется определенная необходимость в сравнении двух переменных, относящихся к одной и той же выборке данных (например, какой-либо параметр с разницей во времени), то обычно используется t-критерий Стьюдента для зависимых выборок.

Данный t -критерий для зависимых выборок дает преимущество в том случае, когда важный источник внутригрупповой вариации (или ошибки) может быть легко определен и исключен из анализа. В частности, это относится к экспериментам, в которых две сравниваемые группы наблюдений основываются на одной и той же выборке наблюдений (субъектов), которые тестировались дважды [18].

Вместо исследования каждой группы по-отдельности и анализа исходных значений можно рассматривать разности между двумя измерениями (например, «до теста» и «после теста») для каждого конкретного элемента.

Теоретические предположения t-критерия Стьюдента для независимых выборок также применяются и к критерию зависимых выборок. Это означает, что разности пар должны быть нормально распределены.

Для данного критерия статистика имеет следующий вид:

$$t = \frac{|Md|}{Sd/\sqrt{n}} \quad (15)$$

где  $Md$  – средняя разность значений,

$Sd$  – стандартное отклонение разностей,

$n$  – количество наблюдений.

Необходимо особо отметить, что данная статистика будет иметь распределение  $t(n-1)$  [18].

### Методические указания к работе

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### Задание

Необходимо провести сравнение выборок путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (15).

Выполнить расчеты не менее, чем на трех выборках.

### Пример

Все параметры внесены в таблицу, отображенную на рисунке 6.

Выборка 1	Выборка 2	Разность значений	Показатели	Значение
8,7	8,9	0,2	$n$	25
11,9	6,9	5	$Md$	2,16
9	5,2	3,8	$Scl$	1,7
12,4	9,6	2,8	$t$	0,26
9	7	2	$t_{крит}$	2,06
12,8	11,4	1,4	$t < t_{крит}$	Выборки однородны
9,2	5,8	3,4		
9,3	12,4	3,1		
13,6	7	6,6		
9,3	6,5	2,8		
7,6	8,3	0,7		
7,8	4,6	3,2		
10,9	11,2	0,3		
12,3	8,2	4,1		
7,7	7,7	0		
9,7	6,1	3,6		
10,3	12,3	2		
6,9	6,7	0,2		
10,1	12,2	2,1		
9,6	10,6	1		
9,2	8,3	0,9		
7,1	7,1	0		
11,4	9,4	2		
10,4	8,9	1,5		
10,1	8,8	1,3		

Рисунок 6 – Пример решения задачи 6

Все необходимые показатели рассчитаны.

## Критерий различия между группами в связанных выборках (непараметрический)

**Цель занятия:** научиться выполнять расчет непараметрического критерия различия между группами в связанных выборках.

### Краткие теоретические сведения

Наиболее простой и доступный для использования критерий известен, как «критерий знаков».

Особенностью этого критерия является его двусторонность, т.е. он может отражать как положительные, так и отрицательные изменения признака.

Алгоритм для определения различия между группами в связанных выборках методом G-критерия знаков следующий.

Сначала из элементов второй выборки (результатов «после») вычитаются элементы первой выборки (результаты «до»). Полученные нули игнорируются, рассматриваются лишь полученные положительные и отрицательные значения. Причем, сами значения не играют никакой роли, как правило, отклонения в ту или иную сторону помечаются просто знаком «+» и «-».

Если в результате получаются преимущественно значения «+», то рассматривается положительное отклонение, иначе отклонение будет отрицательным. Превалирующие значения подсчитывают, это и есть показатель  $G$ .

Для этого критерия также существуют таблицы критических значений, однако для простоты можно воспользоваться теоремой Муавра-Лапласа.

Из теоремы Муавра-Лапласа вытекает правило принятия решений на уровне значимости 5%: если

$$K_g = \left| \frac{2G-n}{\sqrt{n}} \right| \leq 1,96 \quad (15)$$

то гипотезу однородности связанных выборок принимают, в противном случае отклоняют [19].

## Методические указания к работе

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### Задание

Необходимо провести сравнение выборок путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (20).

Выполнить расчеты не менее, чем на трех выборках.

### Пример

Все параметры внесены в таблицу, отображенную на рисунке 7.

Выборка 1	Выборка 2	Разность значений	Показатели	Значение
8,7	8,9	+	$n$	25
11,9	6,9	-	" $+$ "	7
9	5,2	-	" $+$ "	18
12,4	9,6	-	$G$	18
9	7	-	$Kk$	2,20
12,8	11,4	-		
9,2	5,8	-	$Kk > 1,96$	Выборки неоднородны
9,3	12,4	+		
13,6	7	-		
9,3	6,5	-		
7,6	8,3	+		
7,8	4,6	-		
10,9	11,2	+		
12,3	8,2	-		
7,7	7,7	-		
9,7	6,1	-		
10,3	12,3	+		
6,9	6,7	-		
10,1	12,2	+		
9,6	10,6	+		
9,2	8,3	-		
7,1	7,1	-		
11,4	9,4	-		
10,4	8,9	-		
10,1	8,8	-		

Рисунок 7 – Пример решения задачи 7

Все необходимые показатели рассчитаны.

## Критерий зависимости между переменными (параметрический)

**Цель занятия:** научиться выполнять расчет параметрического критерия зависимости между переменными в связанных выборках.

### Краткие теоретические сведения

Критерием зависимости между переменными в статистике выступает корреляционная связь.

Корреляционная связь – это согласованное изменение двух признаков, отражающее тот факт, что изменчивость одного признака находится в соответствии с изменчивостью другого [20].

Задача корреляционного анализа сводится к установлению направления (положительное или отрицательное) связи между варьирующими признаками, измерению ее тесноты, и, наконец, к проверке уровня значимости полученных коэффициентов корреляции [21].

По направлению корреляционная связь может быть положительной («прямой») и отрицательной («обратной»). При положительной прямолинейной корреляции более высоким значениям одного признака соответствуют более высокие значения другого, а более низким значениям одного признака – низкие значения другого. При отрицательной корреляции соотношения обратные [22].

Степень, сила или теснота корреляционной связи определяется по величине коэффициента корреляции. Сила связи не зависит от ее направленности и определяется по абсолютному значению коэффициента корреляции [21].

Сам же коэффициент корреляции Пирсона рассчитывается по формуле:

$$r_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 (Y - \bar{Y})^2}} \quad (16)$$

где

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \quad (17)$$



$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t \quad (18)$$

Коэффициенты корреляции Пирсона используются во многих алгоритмах многомерного статистического анализа. В теоретических рассуждениях часто считают, что случайные вектора имеют двумерное нормальное распределение.

Зная коэффициент корреляции, можно дать качественно-количественную оценку тесноты связи. Как правило, используются специальные табличные соотношения (так называемая шкала Чеддока).

Шкала Чеддока выглядит следующим образом:

Таблица 1 – Шкала Чеддока

Количественная мера тесноты связи	Качественная характеристика силы связи
0,1 - 0,3	Слабая
0,3 - 0,5	Умеренная
0,5 - 0,7	Заметная
0,7 - 0,9	Высокая
0,9 - 0,99	Весьма высокая

Коэффициент корреляции Пирсона является достаточно мощным и широко распространенным инструментом при обработке статистических данных.

### **Методические указания к работе**

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### **Задание**

Необходимо выявить силу и направление связи путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (16).

Выполнить расчеты не менее, чем на трех выборках.

### **Пример**

Все параметры внесены в таблицу, отображенную на рисунке 8.

Выборка 1	Выборка 2	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2 (y_i - \bar{y})^2$	Показатели	Значение
8,7	8,9	-1,2	1,33	0,5	0,21	-0,53	0,28	$n$	25
11,9	6,9	2,0	4,19	-1,5	2,38	-3,16	10,00	$\bar{x}$ среднее	9,85
9	5,2	-0,9	0,73	-3,2	10,52	2,76	7,64	$\bar{y}$ среднее	8,44
12,4	9,6	2,5	6,49	1,2	1,34	2,95	8,68	$cov_{xy}$	32,50
9	7	-0,9	0,73	-1,4	2,09	1,23	1,51	$\sigma x \sigma y$	262,54
12,8	11,4	2,9	8,69	3,0	8,74	8,71	75,94	$r$	0,124
9,2	5,8	-0,7	0,43	-2,6	6,99	1,72	2,97		
9,3	12,4	-0,6	0,30	4,0	15,65	-2,18	4,77	$r=0,124$	Связь прямая, слабая
13,6	7	3,7	14,05	-1,4	2,09	-5,41	29,29		
9,3	6,5	-0,6	0,30	-1,9	3,78	1,07	1,15		
7,6	8,3	-2,3	5,07	-0,1	0,02	0,32	0,11		
7,8	4,6	-2,1	4,21	-3,8	14,78	7,89	62,22		
10,9	11,2	1,0	1,10	2,8	7,60	2,89	8,34		
12,3	8,2	2,4	5,99	-0,2	0,06	-0,60	0,36		
7,7	7,7	-2,2	4,63	-0,7	0,55	1,60	2,56		
9,7	6,1	-0,2	0,02	-2,3	5,49	0,36	0,13		
10,3	12,3	0,4	0,20	3,9	14,87	1,73	2,98		
6,9	6,7	-3,0	8,71	-1,7	3,04	5,15	26,50		
10,1	12,2	0,2	0,06	3,8	14,11	0,93	0,87		
9,6	10,6	-0,3	0,06	2,2	4,65	-0,54	0,30		
9,2	8,3	-0,7	0,43	-0,1	0,02	0,09	0,01		
7,1	7,1	-2,8	7,57	-1,3	1,81	3,70	13,68		
11,4	9,4	1,5	2,40	1,0	0,91	1,48	2,19		
10,4	8,9	0,5	0,30	0,5	0,21	0,25	0,06		
26	10,1	8,8	0,2	0,06	0,4	0,13	0,01		

Рисунок 8 – Пример решения задачи 8

Все необходимые показатели рассчитаны.

## Критерий зависимости между переменными (непараметрический)

**Цель занятия:** научиться выполнять расчет непараметрического критерия зависимости между переменными в связанных выборках.

### Краткие теоретические сведения

Непараметрические аналоги коэффициенту корреляции Пирсона являются статистики Спирмена  $R$ ,  $\tau$  Кендалла и коэффициент гамма.

Сравнивая преимущества и недостатки различных непараметрических коэффициентов корреляции, многие авторы [15, 17, 19, 22] склоняются к тому, что достаточно эффективным и в то же время простым в применении является коэффициент ранговой корреляции Спирмена.

Для расчета непараметрического коэффициента ранговой корреляции Спирмена необходимо сделать следующее [23].

1 Для каждого  $x_i$  рассчитать его ранг  $r_i$  в вариационном ряду, построенном по выборке  $x_1, x_2, \dots, x_n$ .

2 Для каждого  $y_i$  рассчитать его ранг  $q_i$  в вариационном ряду, построенном по выборке  $y_1, y_2, \dots, y_n$ .

3 Для набора из  $n$  пар вычислить линейный коэффициент корреляции. Он называется коэффициентом ранговой корреляции, поскольку определяется через ранги.

Коэффициент ранговой корреляции Спирмена равен

$$R = 1 - \frac{6 \sum_{i=1}^n (r_i - q_i)^2}{n^3 - n} \quad (19)$$

Отметим, что коэффициент ранговой корреляции Спирмена остается постоянным при любом строго возрастающем преобразовании шкалы измерения результатов наблюдений. Другими словами, он является адекватным в порядковой шкале, как и другие ранговые статистики, например, статистики Вилкоксона,

Смирнова, типа омега-квадрат для проверки однородности независимых выборок [19].

### Методические указания к работе

Расчет критерия необходимо реализовать средствами MS Excel. Выборку данных необходимо взять у преподавателя (не воспрещается использовать собственные данные или воспользоваться генератором случайных чисел).

### Задание

Необходимо выявить силу и направление связи путем расчета критерия. Для этого в табличной форме отображаются все показатели, входящие в формулу (19).

Выполнить расчеты не менее, чем на трех выборках.

### Пример

Все параметры внесены в таблицу, отображенную на рисунке 3.

	Выборка 1	Выборка 2	Ранги выборки 1	Ранги выборки 2	Разность рангов	Квадрат разности	Показатели	Значение
1	Выборка 1	Выборка 2	Ранги выборки 1	Ранги выборки 2	Разность рангов	Квадрат разности		
2	8,7	8,9	6	16	10	100	n	25
3	11,9	6,9	21	7	-14	196	Сумма	
4	9	5,2	7	2	-5	25	квадратов	
5	12,4	9,6	23	19	-4	16	разностей	1520
6	9	7	8	8	0	0	R	0,42
7	12,8	11,4	24	22	-2	4		
8	9,2	5,8	9	3	-6	36	R=0,42	Связь прямая, умеренная
9	9,3	12,4	11	25	14	196		
10	13,6	7	25	9	-16	256		
11	9,3	6,5	12	5	-7	49		
12	7,6	8,3	3	13	10	100		
13	7,8	4,6	5	1	-4	16		
14	10,9	11,2	19	21	2	4		
15	12,3	8,2	22	12	-10	100		
16	7,7	7,7	4	11	7	49		
17	9,7	6,1	14	4	-10	100		
18	10,3	12,3	17	24	7	49		
19	6,9	6,7	1	6	5	25		
20	10,1	12,2	15	23	8	64		
21	9,6	10,6	13	20	7	49		
22	9,2	8,3	10	14	4	16		
23	7,1	7,1	2	10	8	64		
24	11,4	9,4	20	18	-2	4		
25	10,4	8,9	18	17	-1	1		
26	10,1	8,8	16	15	-1	1		
27								

Рисунок 3 – Пример решения задачи 3

Все необходимые показатели рассчитаны.

## Список использованных источников

1. Новиков Д.А., Новочадов В.В. Статистические методы в медико-биологическом эксперименте (типовые случаи). Волгоград: Издательство ВолГМУ, 2005. – 84 с.
2. Малая медицинская энциклопедия. – М.: Медицинская энциклопедия. 1991-1996 гг.
3. Айвазян С.А. Прикладная статистика: Исследование зависимостей. Справочное издание/ С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1985. – 487 с.
4. Trochim William M.K. Research Methods Knowledge Base 10/20/2006.
5. Шепель В.Н. Алгоритм определения эмпирической функции плотности  $\hat{f}^{(n)}(x)$  по выборке из генеральной совокупности. Современные инф. техн. в науке и практ. Материалы VIII всеросс. науч.-практ. конф. (с межд. уч.). – Оренбург: ИПК ГОУ ОГУ, 2009, с. 224-226.
6. Орлов А.И. О применении статистических методов в медико-биологических исследованиях: Вестник Академии медицинских наук СССР. 1987 г. №.2 – с.88-94.
7. Литвинов В. А., Оксанич И. Н. Оценка уровня виртуальной интеллектуальности прикладной программно-технической системы на основе анализа эргономической модели // Математические машины и системы, Киев, 2008, №2, С. 100-105.
8. Куприенко Н. В., Статистика. Методы анализа распределений. Выборочное наблюдение: учеб. пособие / Н. В. Куприенко, О. А. Пономарева. – СПб.: Издательство СПбГПУ, 2005. – 128 с.
9. Zar J.H. Biostatistical Analysis (5th edition) // Prentice Hall, 960 p.
10. Орлов А.И. Высокие статистические технологии // Типовые ошибки при вхождении в прикладную статистику: [Электронный ресурс]. Режим доступа <http://forum.orlovs.pp.ru/viewtopic.php?t=97>.

11. Крамер Г. Математические методы статистики / Пер. с англ. / 2-е изд. - М.: Мир, 1975. – 648 с.
12. Макарова Н. В., Трофимец В. Я. Статистика в Excel: Учеб. Пособие // Финансы и статистика. 2002. – 367 с.
13. Архипова ГЛ., Лаврова И.Г., Трошина И.М. Некоторые современные методы статистического анализа в медицине. – М.: Метроснаб, 1971. – 75 с.
14. Платонов А.Е. Статистический анализ в медицине и биологии: задачи, терминология, логика, компьютерные методы. – М.: Издательство РАМН, 2000. - 52 с.
15. Гублер Е. В., Генкин А. А. Применение непараметрических критериев статистики в медико-биологических исследованиях. – Л., 1973.
16. Mann H. B., Whitney D. R. On a test of whether one of two random variables is stochastically larger than the other. // Annals of Mathematical Statistics. – 1947. – № 18. – P. 50-60.
17. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 238 с.
18. StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft. WEB: <http://www.statsoft.ru/home/textbook/default.htm>.
19. Орлов А.И. Прикладная статистика М.: Издательство «Экзамен», 2004. – 656 с.
20. Большая советская энциклопедия. - М.: Советская энциклопедия. 1969-1978.
21. Кузнецов В.М. Основы научных исследований в животноводстве. Киров: Зональный НИИСХ Северо-Востока, 2006.- 568 с.
22. Сидоренко Е.В. Методы математической обработки в психологии // СПб.: ООО «Речь», 2003. – 350 с.
23. Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271-295.