Министерство образования Республики Беларусь
Учреждение образования «Витебский государственный
университет имени П.М. Машерова»
Кафедра географии

В.Н. Лабовкин, И.А. Красовская

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЙ В ГЕОГРАФИИ

Методические рекомендации

Витебск
ВГУ имени П.М. Машерова
2013

УДК 004(075.8)
ББК 32.97я73
Л68

Печатается по решению научно-методического совета учреждения образования «Витебский государственный университет имени П.М. Машерова». Протокол № 1 от 24.10.2013 г.

Авторы: доцент кафедры информатики и информационных технологий ВГУ имени П.М. Машерова, кандидат технических наук **В.Н. Лабовкин;** доцент кафедры географии ВГУ имени П.М. Машерова, кандидат геолого-минералогических наук **И.А. Красовская**

Р е ц е н з е н т ы :
заведующий кафедрой географии ГГУ имени Ф. Скорины, кандидат географических наук, доцент *А.И. Павловский;*
доцент кафедры геометрии и математического анализа ВГУ имени П.М. Машерова, кандидат физико-математических наук *С.А. Шлапаков*

Лабовкин, В.Н.

Л68 Математические методы исследований в географии : методические рекомендации / В.Н. Лабовкин, И.А. Красовская. – Витебск : ВГУ имени П.М. Машерова, 2013. – 36 с.

В методических рекомендациях рассмотрены теоретические основы методов математической статистики для обработки результатов полевых и экспериментальных измерений с использованием компьютеров в среде электронных таблиц Excel. Приводятся примеры обработки и интерпретации результатов, а также возможные варианты задач. Издание предназначено для выполнения лабораторных работ по математическим методам исследований в географии. Может использоваться студентами естественнонаучных специальностей при обработке результатов собственных экспериментов и проведении научно-исследовательских работ.

УДК 004(075.8)
ББК 32.97я73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Исчерпывающую информацию о природной совокупности можно получить, выяснив значения признака для всех без исключения ее составляющих. В отдельных случаях так и поступают: например, при переписях населения каких-либо территорий (район, область, государство) получают данные о распределении жителей по полу, возрасту, образованию и т.д. Однако, полное наблюдение не всегда возможно провести из-за больших трудовых затрат, времени и средств. В некоторых случаях выполнить такое наблюдение просто невозможно. Например, нельзя выловить всех рыб исследуемого вида из океана или высеять все семена некоторого сорта пшеницы, чтобы определить их всхожесть.

Таким образом, чтобы обойти указанные трудности, изучение всей совокупности заменяют изучением некоторой ее части – выборки. Формируя выборку, исследователь должен организовать ее так, чтобы она наиболее полно характеризовала свойства всей совокупности, т.е. была репрезентативной. В математической статистике доказывается, что выборка будет репрезентативной при условии случайности отбора элементов выборки из всей совокупности. Такая случайность достигается в том случае, если все объекты имеют одинаковую вероятность попасть в выборку.

Разработка методов, позволяющих по результатам изучения тех или иных свойств объектов, составляющих выборку делать обоснованное заключение о распространении этих свойств на всю совокупность, является одной из важнейших задач математической статистики.

Математическая обработка результатов измерений числового значения некоторого признака у элементов выборки связана со значительными вычислениями. Доступность персональных компьютеров и развитого программного обеспечения освобождает исследователя от непроизводительных затрат времени на обработку результатов измерений и позволяет представить информацию в более наглядном графическом виде.

Для записи результатов большого количества однотипных измерений удобно использовать таблицы. С их помощью удается избежать ненужной многократной записи обозначения измеряемой величины, единиц измерения, используемых множителей и т.п. В таблицы, помимо экспериментальных данных, могут быть сведены промежуточные результаты обработки этих данных. Форма таблицы должна быть удобна для записи и дальнейшей обработки экспериментальных данных. Поэтому необходимо предварительно продумать, значения каких физических величин или результаты расчетов будут помещены в таблицу. Отсюда заранее определяют количество столбцов и строк, необходимых в таблице.

Данное учебное издание посвящено рассмотрению методов математической статистики для обработки результатов измерений с использованием персональных компьютеров в среде электронных таблиц Excel. Издание предназначено для выполнения лабораторных работ по математическим методам исследований в географии, предусмотренных учебной программой курса «Географические методы исследований» для студентов специальности 1-31 02 01-02 География (научно-педагогическая деятельность). Может использоваться студентами естественнонаучных специальностей при обработке результатов собственных экспериментов и проведении научно-исследовательских работ.

ЛАБОРАТОРНАЯ РАБОТА № 1

Вариационный ряд и его графическое представление

При большом числе экспериментальных данных простая статистическая совокупность перестает быть удобной формой записи статистического материала – она становится слишком громоздкой и мало наглядной. Представление результатов наблюдений в компактной форме значительно облегчает их восприятие, сравнение с ранее полученными данными и выработку тех или иных гипотез относительно некоторого свойства.

Пусть рассматриваемая дискретная случайная величина **X** в результате n независимых опытов принимает значения

$$x_1, x_2, \ldots, x_n. \qquad (1)$$

Отдельные **различные** значения случайной величины **X** называются **вариантами**, а совокупность всех значений x_i, (i=1,...,n) называется **выборкой.**

Для анализа результатов наблюдений необходимо сгруппировать полученные значения, располагая их в порядке возрастания вариант α_j, (j=1,...,m, m≤n) и указать число опытов f_j, в которых наступило событие $\mathbf{X}=\alpha_j$, т.е. **частоту** f_j появления каждой из вариант.

Построенная таблица частот

α_1	α_2	…	α_m
f_1	f_2	…	f_m

называется **вариационным рядом.**

Если изучается непрерывная случайная величина, то группировка заключается в разбиении интервала наблюдаемых значений случайной величины на L частичных интервалов (классов) равной длины $(c_0;c_1)$, $(c_1;c_2)$, …, $(c_{L-1};c_L)$ и подсчете количества f_j элементов выборки x_i попавшим в j-й интервал. Значения x_i лежащие на границе между интервалами, относятся к правому интервалу, а $x_i=x_{max}$ – к левому интервалу. Числа f_j называются **частотами**, а $p_j=f_j/n$ – относительными частотами попадания в j-й интервал или **частостями**. Очевидно, что $\sum_{j=1}^{L} f_j = n$, а $\sum_{j=1}^{L} p_j = 1$.

Количество интервалов L обычно вычисляют по следующей эмпирической формуле

$$L = 1 + 3{,}32 \cdot \lg n$$

с округлением до большего целого. Длина каждого интервала определяется из выражения

$$\delta = \frac{x_{\max} - x_{\min}}{L}$$

с округлением до большего целого, где x_{max} и x_{min} – соответственно наибольшее и наименьшее значения случайной величины X в выборке объема n.

Таким образом, $c_0 = x_{min}$, $c_j = c_0 + \delta \cdot j$ (j = 1,…,L).

Произведенная группировка позволяет представить полученную выборочную информацию в виде вариационного ряда.

Границы интервалов	Середина интервала	Частота	Частость, %
c_0; c_1	$\frac{c_0 + c_1}{2}$	f_1	p_1
.	.	.	.
.	.	.	.
.	.	.	.
c_{L-1}; c_L	$\frac{c_{L-1} + c_L}{2}$	f_L	p_L

Иногда интервальный вариационный ряд строят и для дискретных случайных величин, если число наблюдающихся вариант слишком велико. В свою очередь, интервальный вариационный ряд может быть условно заменен дискретным. В этом случае срединное значение интервала принимается за варианту, а соответствующая интервальная частота рассматривается как дискретная частота этой варианты.

Для большей наглядности и улучшения визуального восприятия вариационные ряды удобно представлять в графической форме. Распределение значений дискретного признака отображается с помощью **полигона**. Для его построения в прямоугольной системе координат наносятся точки с координатами (α_j, f_j), где α_j – значения вариант, откладываемые по горизонтальной оси абсцисс, а f_j – частота появления j-го значения, откладываемая по вертикальной оси ординат. Построенные таким образом точки соединяют отрезками прямых.

Интервальные вариационные ряды изображают с помощью **гистограмм**. По горизонтальной оси которых откладывают отрезки, изображающие интервалы варьируемого признака, на каждом из которых, как на основании, строится прямоугольник, с высотой равной частоте f_j или частости p_j данного интервала.

Построение гистограмм и их анализ является самым простым и широко распространенным методом анализа экспериментальных данных. При визуальном восприятии гистограмм исследователь может выдвинуть гипотезу о законе распределения исследуемой случайной

величины X, сделать выводы о сосредоточении основной части полученных данных вокруг средних значений и характере рассеивания наблюдений вокруг средних значений.

Контрольные вопросы:

1. Что такое выборка и варианта?
2. Как распределяются варианты в вариационном ряду?
3. Каковы принципы группировки данных для дискретных и непрерывных случайных величин?
4. На сколько классов разбивается интервал наблюдаемых значений случайной величины?
5. В чем разница между гистограммой и полигоном распределения?
6. Какую информацию можно получить в результате анализа гистограмм?

Пример решения задачи:

При испытании нового рациона кормления цыплят была взята выборка, состоящая из веса 50 цыплят в граммах (таблица). Требуется составить вариационный ряд распределения веса цыплят, представить его графически и сделать выводы.

145	131	120	100	115	184	161	147	129	167
149	173	147	180	195	157	149	177	163	113
152	170	157	122	174	151	164	146	133	148
150	153	160	190	157	147	159	154	153	159
146	112	180	157	161	171	152	179	170	181

Т.к. число наблюдающихся вариант дискретных случайных величин (вес цыплят) слишком велико, строим интервальный вариационный ряд. Для этого заполним в Excel таблицу с исходными данными и подсчитаем в ней значения x_{max}, x_{min}, L, δ, необходимые для построения вариационного ряда.

В ячейке B8 вычисляется x_{min} по формуле =МИН(A2:J6), в ячейке B9 вычисляется x_{max} по формуле =МАКС(A2:J6), в ячейке B10 вычисляется L по формуле =ОКРУГЛВВЕРХ(1+3,32*LOG10(50);0), в ячейке B11 вычисляется δ по формуле =ОКРУГЛВВЕРХ((B9-B8)/B10;0).

	A	B	C	D	E	F	G	H	I	J
1	**Распределение веса цыплят в граммах при испытании нового рациона.** **Таблица 1.**									
2	145	131	120	100	115	184	161	147	129	167
3	149	173	147	180	195	157	149	177	163	141
4	152	170	157	122	174	151	164	146	133	148
5	150	153	160	190	157	147	159	154	153	159
6	146	112	180	157	161	171	152	179	170	181
7										
8	X_{min}=	100								
9	X_{max}=	195								
10	L=	7								
11	δ=	14								
12				**Таблица 2**						
13	**Номер интервала**	**Границы интервалов**	**Середина интервала**	**Частота, f**	**Частость, %**					
14	1	100; 114	107	2	4%					
15	2	114; 128	121	3	6%					
16	3	128; 142	135	4	8%					
17	4	142; 156	149	16	32%					
18	5	156; 170	163	12	24%					
19	6	170; 184	177	10	20%					
20	7	184; 198	191	3	6%					
21				50	100%					

Для построения гистограммы распределения веса цыплят необходимо выделить диапазон ячеек B14:B20 и с нажатой клавишей <Ctrl> выделить диапазон D14:D20 и дать команду <Вставка – Диаграмма…>. В окне <Мастер диаграмм> выбрать тип диаграммы <Гистограмма>, ее вид <Обычная> и щелкнуть <Далее>. В появившемся окне <источник данных диаграммы> отображаются диапазон выде-

ленных ячеек и вид будущей гистограммы. После щелчка по кнопке <Далее> в окне <параметры диаграммы> ввести название гистограммы: Гистограмма распределения веса цыплят, Ось X: Границы интервалов, Ось Y: Частота и щелкнуть по кнопке <Готово>. Для окончательного оформления гистограммы щелкните правой кнопкой на любом из рядов данных, выберите <Формат рядов данных… – Параметры>, укажите <Ширина зазора> 0 и щелкните <Ок>. Удалите легенду.

Анализируя распределение вариант в вариационном ряду, представленном в табл. 2 и на рисунке, легко заметить некоторые общие закономерности:

- большинство вариант располагается в средней части вариационного ряда или около середины вариационной кривой, здесь наблюдается максимум вариант, как бы их сгущение;
- распределение вариант в обе стороны от этого максимума более или менее симметрично;
- частота вариант постепенно убывает к краям вариационного ряда.

Варианты заданий:

1. Для определения петрографического типа пород из горизонта неогеновых лав отобрано и проанализировано на содержание кварца (SiO_2) 30 проб.

Содержание SiO_2 в (%) в неогеновых лавах

№ пробы	SiO_2	№ пробы	SiO_2	№ пробы	SiO_2	№ пробы	SiO_2
1	59,5	9	73,2	17	69,3	24	61,1
2	66,8	10	64,6	18	64,6	25	63,8

3	60,5	11	62,9	19	67,8	26	67,5
4	63,7	12	62,4	20	56,6	27	65,3
5	72,5	13	71,6	21	71,4	28	69,9
6	69,2	14	65,8	22	67,6	29	73,2
7	61,2	15	63,1	23	63,6	30	60,7
8	66,3	16	61,2				

Содержание SiO_2 в отдельных пробах меняется от 56,6 (андезито-базальт) до 73,2 % (риолит), что не позволяет оценить состав лав горизонта в целом по единичному наблюдению.

Рассчитайте для определенных интервалов (классов) значений частости и графически изобразите вероятность рассматриваемых случайных событий (содержание SiO_2) в отдельных пробах.

2. На изучаемом месторождении выделяют три типа руд: богатые – с содержанием молибдена больше 200 усл. ед., рядовые – с содержанием от 100 до 200 усл. ед., бедные – с содержанием ниже 100 усл. ед. Отобранные в месторождении пробы руды показали следующее содержание молибдена.

№ обр.	Содержание Mo	№ обр.	Содержание Mo	№ обр.	Содержание Mo	№ обр.	Содержание Mo
1	280	11	92	21	48	31	45
2	300	12	89	22	60	32	47
3	87	13	154	23	230	33	205
4	96	14	97	24	130	34	265
5	275	15	96	25	142	35	124
6	240	16	72	26	54	36	39
7	80	17	201	27	81	37	270
8	48	18	105	28	112	38	46
9	154	19	195	29	78	39	53
10	83	20	30	30	124	40	251

Определить по гистограмме распределения содержания молибдена в отобранных образцах соотношение различных сортов руд изучаемого месторождения.

3. В таблице приведены замеры азимутов падения кварцевых прожилков по документации канав на рудопроявлении золота.

Замеры азимутов падения (в градусах) кварцевых прожилков

№ п/п	Азимут	№ обр.	Азимут	№ обр.	Азимут	№ обр.	Азимут
1	132	13	330	25	178	37	105
2	302	14	88	26	335	38	130
3	304	15	191	27	110	39	144

4	162	16	198	28	112	40	177
5	130	17	325	29	200	41	42
6	58	18	214	30	257	42	190
7	159	19	211	31	270	43	169
8	144	20	199	32	171	4	41
9	315	21	124	33	141	45	205
10	162	22	84	34	260	46	225
11	318	23	181	35	185	47	270
12	92	24	3	36	15	48	260

Определите графически, соответствует ли распределение азимутов падения нормальному закону распределения?

4. В пределах пункта исследований ежемесячно в течение 5 лет проводили измерения суммарного количества выпавших осадков (мм). Определите графически, существуют ли общие закономерности в распределении количества осадков, если за время наблюдений были получены следующие результаты.

Количество осадков в пункте наблюдений (мм)

Месяц / Год	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1 год	105	124	133	128	138	161	129	103	136	110	155	131
2 год	149	144	147	104	169	136	115	105	142	119	150	168
3 год	133	127	13	149	106	135	144	139	125	144	121	140
4 год	125	158	133	179	116	136	112	122	151	131	134	142
5 год	140	110	119	125	124	104	170	117	138	130	160	124

5. В результате замера длины тела у 50 экземпляров леща в возрасте года озера Езерище получена следующая таблица (в мм).

143	155	148	139	137	140	139	142	140	141
142	120	144	130	138	124	127	137	139	129
128	119	120	138	130	114	126	138	117	132
130	145	140	153	138	142	145	137	131	125
127	134	135	137	139	125	137	131	165	136

Составьте вариационный ряд и постройте гистограмму.

6. Месторождение силикатного никеля приурочено к латеритной коре выветривания. Для изучения химического состава коры выветривания и поведения различных элементов в процессе корообразования на одном из участков месторождения были отобраны пробы, по которым выполнены анализы на содержание в том числе NiO и SiO_2.

Содержание NiO									
0,35	1,65	1,36	1,23	0,42	0,27	0,38	1,36	0,90	2,59
1,33	0,41	0,50	1,06	1,35	3,63	1,22	0,44	0,29	1,44
1,30	2,32	1,12	0,32	1,35	0,66	1,17	1,12	1,65	1,41
1,10	2,69	1,16	2,58	2,09	3,28	1,50	2,55	2,70	0,45
Содержание SiO_2									
6,73	2,30	3,98	12,48	36,00	37,52	7,91	3,27	2,32	13,26
38,77	38,96	4,40	7,92	6,32	18,99	33,44	41,30	5,28	4,21
2,02	26,19	34,15	38,21	1,97	6,37	24,20	3,09	10,89	26,51
2,00	11,26	27,04	28,36	26,10	21,94	37,86	31,12	4,45	31,50

Установите графически характер поведения NiO и SiO_2 в процессе корообразования.

ЛАБОРАТОРНАЯ РАБОТА № 2

Оценка основных параметров распределения выборки

Графическое изображение вариационного ряда – это превращение исходных результатов наблюдений в наглядную форму. Для сравнения вариационных рядов разных выборок важно получить их числовые характеристики. Вариационные ряды могут различаться по тому значению признака, вокруг которого концентрируется большинство вариант и по степени отклонения вариант от этого признака. В соответствии с этим, статистические показатели разделяются на показатели, характеризующие центральную тенденцию (средняя арифметическая, средняя геометрическая, мода, медиана) и показатели, измеряющие степень вариации (вариационный размах, среднее абсолютное отклонение, дисперсия, среднее квадратическое отклонение, коэффициенты вариации и асимметрии, эксцесс).

При наличии выборки наблюдений xi, (i=1,…,n) за переменной X **среднее арифметическое значений** данной переменной, называемое также **математическим ожиданием**, находится по формуле:

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

Средняя арифметическая является обобщенной величиной, она отражает уровень совокупности в целом, дает сводную характеристику данного изучаемого признака.

Средняя геометрическая

$$G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

вычисляется только при положительных и не равных нулю значениях наблюдаемого признака. Основным критерием для применения средней геометрической является возрастание данного признака в геометрической прогрессии. На практике обычно используется следующая рабочая формула:

$$G = 10^{\frac{1}{n}\sum_{i=1}^{n} \lg x_i}.$$

Наиболее часто встречающееся значение признака во множестве наблюдений, образующих вариационный ряд, называется **модой** и обозначается Мо. Если все значения в вариационном ряду имеют одинаковую частоту, то он не имеет моды. В зависимости от числа мод бывают унимодальные, бимодальные и тримодальные вариационные ряды.

Если распределение асимметрично, т.е. с одной стороны гистограммы наблюдается резкое изменение частот, а с другой – медленное, при этом большая часть вариант расположена с одной стороны от среднего арифметического, тогда полезно в качестве характеристики центральной тенденции использовать медиану Ме. **Медиана** – это значение варианты, находящейся точно в середине ряда, таким образом, медиана делит распределение на две равные части, содержащиеся по 50% всех наблюдений.

Простейшим показателем изменчивости вариационного ряда является **вариационный размах**, равный разности наибольшей (x_{max}) и наименьшей (x_{min}) вариант ряда

$$R = x_{max} - x_{min}.$$

Вариационный размах не отражает существенных особенностей изменчивости значений в ряду наблюдений, поскольку крайние значения ряда, как правило, ненадежны и могут значительно отличаться при повторных измерениях. Вот почему для характеристики вариаций между членами совокупности используются другие показатели, которые являются мерами рассеяния вариант вокруг средних величин.

Одним из таких показателей служит **среднее абсолютное отклонение** вариант от среднего арифметического

$$Md = \frac{\sum_{i=1}^{n} |x_i - M|}{n}.$$

Однако среднее абсолютное отклонение не улавливает истинного рассеяния вариант в вариационном ряду вокруг средней арифметической, поэтому значительно чаще используют среднее арифметическое квадратов отклонений вариант от их среднего арифметического, называемое **дисперсией** или **вариансой:**

$$D = \frac{\sum_{i=1}^{n}(x_i - M)^2}{n-1}.$$

В случаях, когда требуется учитывать наименование единицы измерения вариант, используют **среднее квадратическое отклонение:**

$$\sigma = \sqrt{D}.$$

Его можно рассматривать как среднюю характеристику самого отклонения.

В качестве безразмерной меры рассеяния случайной величины X используется **коэффициент вариации** V, определяемый в процентах:

$$V = \frac{\sigma}{M} \cdot 100\%.$$

Коэффициент вариации показывает, насколько среднее арифметическое полно представляет вариационный ряд. При одинаковых средних арифметических значениях у двух выборок более представительным является среднее арифметическое значение той из них, коэффициент вариации которой меньше. Т. к. V безразмерная величина, то коэффициент вариации можно использовать для сравнения вариационных рядов различных признаков.

Для оценки вида и степени асимметрии вариационного ряда используется коэффициент асимметрии:

$$As = \frac{\sum_{i=1}^{n}(x_i - M)^3}{n\sigma^3}.$$

При As=0 значения признака распределены симметрично, если As<0 то говорят о левосторонней асимметрии, т.е. распределение вытянуто влево, если As>0 – правосторонняя асимметрия.

К показателям, описывающим форму кривой распределения вариационного ряда, принадлежит **эксцесс:**

$$Ex = \frac{\sum_{i-1}^{n}(x_i - M)^4}{n\sigma^4} - 3.$$

Значение *Ex*<0 указывает на плосковершинность кривой распределения, т. е. частоты появления большинства вариант, лежащих по обе стороны от медианы, отличаются друг от друга незначительно, значение *Ex*>0 – на островершинность, т. е. небольшое число вариант встречается с частотой, значительно превосходящей частоты остальных значений, значение *Ex*=0 характеризует распределение вероятностного признака как нормальное, средневершинное.

Контрольные вопросы:

1. Какие численные показатели характеризуют центральную тенденцию вариационных рядов и степень их изменчивости?
2. По каким формулам вычисляются средняя арифметическая и средняя геометрическая?
3. В каких случаях целесообразно пользоваться средней геометрической?
4. Что такое мода и медиана?
5. Что характеризует вариационный размах?
6. Какие меры рассеяния вариант вокруг средних используются при обработке статистических совокупностей?
7. В каких случаях вместо дисперсии используют среднее квадратическое отклонение?
8. Для чего используется коэффициент вариации?
9. На что указывает коэффициент асимметрии?
10. Как описывает форму кривой распределения вариационного ряда эксцесс?

Пример решения задачи:

Вычислить для задания из лабораторной работы № 1 среднюю арифметическую, среднюю геометрическую, моду, медиану, вариационный размах, среднее абсолютное отклонение, дисперсию, среднее квадратическое отклонение, коэффициенты вариации и асимметрии, эксцесс.

Для удобства работы скопируем исходную таблицу 1 на лист 2.

В ячейки B8:B18 поместим соответственно формулы:
=СРЗНАЧ(A2:J6), =СРГЕОМ(A2:J6), =МОДА(A2:J6),
=МЕДИАНА(A2:J6), =МАКС(A2:J6)-МИН(A2:J6),
=СРОТКЛ(A2:J6), =ДИСП(A2:J6), =КОРЕНЬ(B14),
=B15/B8*100, =СКОС(A2:J6), =ЭКСЦЕСС(A2:J6).

Результаты вычисления приведены в таблице на с.17.

Расчет основных выборочных параметров может производиться с использованием надстройки (опции) «Пакет анализа», которая позволяет оперативно получить значения показателей описательной статистики.

По умолчанию эта опция не установлена, поэтому для ее активации необходимо с помощью команды <Надстройки> из меню <Сервис> открыть окно диалога «Надстройки» и в нем установить флажок для компоненты «Пакет анализа». После нажатия кнопки [ОК] меню <Сервис> будет дополнено командой <Анализ данных>.

Для расчета показателей выполняем последовательность команд *Сервис/Анализ данных/Описательная статистика*/в поле «*Входной интервал*» указываем наши значения (клетки A2:J6), в поле группиро-

вание выбираем «*по строкам*», в «*Параметрах вывода*» выбираем «*Выходной интервал*» и указываем там ячейку A20, отмечаем параметры «*Итоговая статистика*» и «*Уровень надежности*» (значение можно изменять, в нашем случае указываем 95%), нажимаем [OK].

Варианты заданий:

1. Вычислить для своего варианта задания из лабораторной работы № 1 среднюю арифметическую, среднюю геометрическую, моду, медиану, вариационный размах, среднее абсолютное отклонение, дисперсию, среднее квадратическое отклонение, коэффициенты вариации и асимметрии, эксцесс.

2. Сравнить значения основных выборочных параметров, вычисленные с помощью вставки функций, со значениями, полученными с использованием надстройки «Пакет анализа».

	A	B	C	D	E	F	G	H	I	J
1	**Распределение веса цыплят в граммах при испытании нового рациона Таблица 1**									
2	145	131	120	100	115	184	161	147	129	167
3	149	173	147	180	195	157	149	177	163	141
4	152	170	157	122	174	151	164	146	133	148
5	150	153	160	190	157	147	159	154	153	159
6	146	112	180	157	161	171	152	179	170	181
7										
8	M=	154,8								
9	G=	153,4								
10	Mo=	157								
11	Me=	155,5								
12	R=	95								
13	Md=	15,08								
14	D=	396,7								
15	σ =	19,92								
16	V=	12,87								
17	As=	-0,502								
18	Ex=	0,43								

3. Сделать выводы по параметрам.

ЛАБОРАТОРНАЯ РАБОТА № 3

Измерение связи между выборками. Корреляция

Изучение изменений параметров по отдельным признакам в ряде случаев не достаточно. На практике часто приходится устанавливать зависимость между вариациями двух или даже нескольких признаков. Две любые случайные величины могут быть связаны функ-

циональной или статистической (корреляционной) зависимостью либо быть независимыми.

При функциональных зависимостях каждому значению одной переменной величины соответствует одно вполне определенное значение другой величины. Такие зависимости наблюдаются в математике и физике. Случайные величины, определяемые биологическими объектами, как правило, имеют статистические связи, при которых численному значению одной переменной соответствует много значений другой переменной. Например, между ростом и весом определенной группы людей, объединенных по какому-то критерию в совокупность, существует зависимость, но это не значит, что определенному росту соответствует строго определенный вес человека. Ясно, что в силу функциональных особенностей организма каждого отдельного человека, среди людей одинакового роста всегда найдутся люди с разным весом.

Пусть (x_1, y_1), (x_2, y_2), …, (x_n, y_n) – наблюдаемые по выборке пары возможных значений случайных величин X и Y. Безразмерной характеристикой связи двух случайных величин X и Y является **коэффициент корреляции**

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{(n-1)\sigma_x\sigma_y}, \text{ или}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2 \sum_{i=1}^{n}(y_i - M_y)^2}}.$$

Свойства коэффициента корреляции.

1. $|r| \le 1$.

2. Если $r = 0$, то вариация обоих признаков происходит независимо.

3. При $r \neq 0$ вариации обоих признаков взаимосвязаны, т.е. с изменением одного признака меняется и другой (в том же направлении при $r>0$ и в противоположном направлении при $r<0$). О тесной корреляции говорят в случае $|r| \ge 0.7$. Коэффициенты корреляции порядка 0,5 – 0,6 считаются средними, $|r|<0,5$ указывает на слабую связь.

Иногда изучение корреляции между двумя признаками затрудняется тем, что их вариации находятся под влиянием какого-то третьего признака. Чтобы исключить это влияние используют **частный коэффициент корреляции** или **коэффициент частной корреляции**.

Допустим, что три случайные величины X, Y, Z коррелируют друг с другом и их коэффициенты корреляции r_{xy}, r_{xz}, r_{yz}, тогда коэффициент частной корреляции вычисляется по формуле:

$$r_{xy\bullet z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}.$$

В этой формуле признак z выключается из корреляционной зависимости между признаками x и y.

Контрольные вопросы:

1. Что такое корреляционная зависимость? Приведите примеры.
2. Какая разница между корреляционной и функциональной зависимостью?
3. Как определяется коэффициент корреляции?
4. Каковы возможные значения коэффициента корреляции?
5. Какая разница между положительной и отрицательной корреляцией?
6. Какие значения коэффициента корреляции считают большими, а какие средними и малыми? Почему?
7. Напишите формулу коэффициента частной корреляции и объясните ее значение.

Пример решения задачи:

Исследованиями установлено, что на содержание подвижного марганца в почве влияет реакция среды. Необходимо доказать достоверность установленной зависимости. Получены следующие исходные данные (*x* – гидролитическая кислотность, мг-экв. на 100 г почвы; *y* – содержание подвижного марганца, мг/кг почвы):

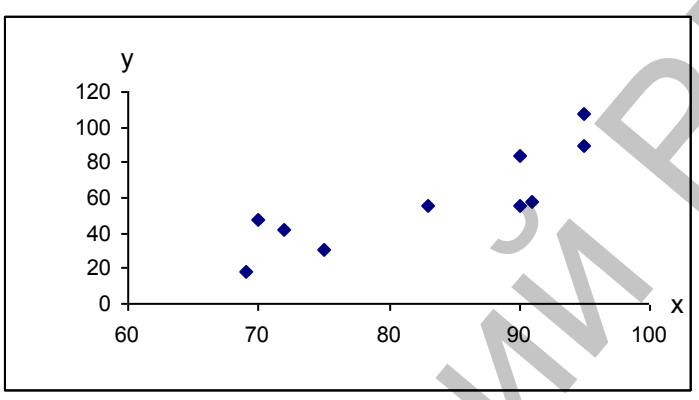
x	83	72	69	90	90	95	95	91	75	70
y	56	42	18	84	56	107	90	58	31	48

1. Вычислите коэффициент корреляции между x_i и y_i.
2. Оцените значимость коэффициента корреляции (***r***) по критерию Стьюдента по формуле $t_r = r\sqrt{N-2}/\sqrt{1-r^2}$ и сравните с табличным (критическим) значением.
3. Сделайте соответствующие выводы.

Вначале строим график, который указывает на существование между исследуемыми показателями положительной линейной зависимости, что требует вычисления коэффициента корреляции. Для вычисления коэффициента корреляции в ячейку B4 помещаем формулу =КОРРЕЛ(B1:K1;B2:K2).

	A	B	C	D	E	F	G	H	I	J	K
1	x	83	72	69	90	90	95	95	91	75	70
2	y	56	42	18	84	56	107	90	58	31	48
3											
4	r =	0,853102									

График зависимости содержания подвижного марганца (*y*) от гидролитической кислотности (*x*)



Поскольку r=0,85>r_m=0,7, то зависимость между содержанием подвижного марганца и гидролитической кислотностью определяется, как достоверная положительная.

Варианты заданий:

1. При проведении гидрогеологических изысканий по одному из профилей буровых скважин были выполнены опытные электроразведочные работы, результаты которых занесены в таблицу. Для оценки эффективности этого метода определите, существует ли зависимость между электрическим сопротивлением пород ($\rho_к$) и относительной мощностью горизонта гравийно-галечных отложений ($m_г$), к которым приурочены основные водоносные горизонты.

Результаты опытных электроразведочных работ

№ скважины	1	2	3	4	5	6	7	8	9	10	11	12
$m_г$, %	67	80	40	24	25	38	18	72	44	51	76	50
$\rho_к$, Ом·м	253	115	126	82	66	25	44	180	32	319	421	51

2. Установите, влияет ли на содержание подвижного марганца в почве реакция среды, если на опытном участке получены экспериментальные данные гидролитической кислотности (x, мг-экв. на 100 г почвы) и концентрациям подвижного марганца (C_{Mn}, мг/кг почвы), которые занесены в таблицу.

Результаты геохимических исследований образцов почвы

Показатель кислотности, мг-экв./100 г	83	72	69	90	90	95	95	91	75	70
C_{Mn}, мг/кг	56	42	18	84	56	107	90	58	31	48

3. Проанализируйте показатели температур (°С), количества осадков (мм) и относительной влажности (%), приведенные в таблице. В каких климатических поясах и областях находятся п.п.1 – 12 (по вариантам)?

Годовой ход температуры, осадков и относительной влажности

п.п.	Климатические показатели	Месяцы												
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Годовые значения
1	Температура, °С	-7,8	-7,3	-2,9	5,0	12,6	16,0	18,0	16,3	11,2	5,2	-0,4	-5,2	5,1
	Осадки, мм	39	33	37	42	58	76	84	82	56	43	49	41	646
	Относительная влажность, %	86	84	79	74	67	69	74	78	82	85	88	88	80
2	Температура, °С	21,6	23,0	26,4	30,1	31,7	30,7	28,1	27,3	28,2	29,7	26,9	22,7	27,2
	Осадки, мм	0	0	2	0	8	21	85	109	52	13	0	0	290
	Относительная влажность, %	29	28	21	19	30	41	61	67	58	42	32	34	38
3	Температура, °С	10,3	11,2	12,4	14,2	15,7	19,3	21,1	21,7	20,2	16,8	13,6	11,0	15,7
	Осадки, мм	92	89	87	66	50	18	4	6	36	83	109	104	744
	Относительная влажность, %	77	72	67	67	64	60	58	57	62	67	73	75	67
4	Температура, °С	-7,0	-6,9	-2,5	4,0	11,6	15,7	18,5	16,5	12,2	5,8	-0,9	-6,2	5,1
	Осадки, мм	36	32	35	40	59	75	81	80	54	42	50	42	626
	Относительная влажность, %	85	83	78	74	66	68	72	77	80	84	87	87	78
5	Температура, °С	25,5	25,8	26,3	26,6	27,0	26,6	26,8	26,5	26,4	26,5	26,1	26,7	26,3
	Осадки, мм	246	181	185	197	166	177	169	198	175	201	256	263	2414
	Относительная влажность, %	85	81	82	82	83	82	81	81	81	82	82	82	82
6	Температура, °С	9,7	11,1	13,8	17,8	22,3	25,6	26,9	26,7	24,4	19,6	14,3	10,7	18,6
	Осадки, мм	76	82	76	65	75	119	172	163	117	81	55	68	1149
	Относительная влажность, %	65	63	62	62	63	65	68	68	67	62	61	65	64

Установите наличие и тесноту корреляционной зависимости показателей относительной влажности от количества осадков и значений температур в одном из пунктов наблюдений.

4. В таблице приведены значения суточных сумм солнечной радиации, приходящей к земной поверхности при абсолютной прозрачности атмосферы на разных широтах северного полушария в дни равноденствий и солнцестояний.

Суточные суммы солнечной радиации (кал/см^2), приходящей к земной поверхности на разных широтах северного полушария

Дата	Широта, °									
	0	10	20	30	40	50	60	70	80	90
21.03	923	909	867	779	707	593	461	316	160	0
22.06	814	900	964	1005	1022	1020	1009	1043	1093	1110
23.09	912	898	857	789	698	586	456	312	158	0
22.12	869	756	624	480	327	181	51	0	0	0

Оцените тесноту зависимости значений суточных сумм солнечной радиации у земной поверхности северного полушария от географических координат в дни весеннего и осеннего равноденствия и летнего и зимнего солнцестояния (по вариантам).

5. На примере Италии установите математически: зависит ли уровень экономического развития административных единиц от долей населения (%), занятого 1) в сельском хозяйстве и рыболовстве; 2) в промышленности и строительстве; 3) в сфере обслуживания.

Области (районы) Италии

Области (районы)	Главные центры областей (районов)	Доля населения, занятого			Величина ВВП на душу населения, (%) (Италия – 100%)
		в сельском хозяйстве и рыболовстве	в промышленности и строительстве	в сфере обслуживания	
Пьемонт	Турин	6,8	41,9	51,4	113
Валле-д'Аоста	Аоста	9,8	26,5	63,8	123
Лигурия	Генуя	3,9	5,0	72,1	115
Ломбардия	Милан	3,2	43,1	53,7	128
Трентино-Альто-Адидже	Тренто	10,2	25,3	64,5	118
Венето	Венеция	7,1	41,6	51,3	111
Фриули-Венеция-Джулия	Триест	5,9	31,8	62,3	116
Эмилия-Романья	Болонья	8,6	35,1	56,3	122
Тоскана	Флоренция	5,5	33,6	60,9	107
Умбрия	Перуджа	9,6	33,4	57,0	98
Марке	Анкона	10,1	38,0	51,8	99
Лацио	Рим	5,0	19,8	75,2	114
Кампания	Неаполь	12,0	24,8	63,2	70
Абруцци	Л'Акуипо	12,4	28,3	59,3	90
Молизе	Компо-бассо	21,0	23,3	55,7	78
Апулия	Бари	15,6	23,8	60,5	73
Базиликата	Потенца	20,3	24,1	55,6	64
Калабрия	Катанд-заро	18,9	20,4	60,7	60
Сицилия	Палермо	14,4	21,8	63,8	70
Италия	Рим	8,5	32,2	59,3	100

6. Сырье, поступающее на завод из карьера, содержит два полезных компонента – минералы A и B. Результаты анализов десяти образцов сырья, поступивших в разное время из разных мест карьера, приведены в таблице, где x и y выражают соответственно процентное содержание минералов A и B в образцах.

x	67	54	72	64	39	22	58	43	46	34
y	24	15	23	19	16	11	20	16	17	13

Оценить коэффициент корреляции величин x и y.

ЛАБОРАТОРНАЯ РАБОТА № 4

Измерение связи между выборками. Регрессия

Коэффициент корреляции указывает лишь на степень связи в вариации двух переменных величин или на меру тесноты этой связи, он не позволяет судить о том, как количественно меняется одна величина по мере изменения другой. На этот вопрос может ответить метод регрессии.

Предположим, что зависимость между величинами X и Y близка к линейной (коэффициент корреляции r близок к +1 или –1). В этом случае естественно ставить вопрос о функции y=ax+b, которая наилучшим образом выражала бы зависимость Y от X. Для нахождения такой функции используется метод наименьших квадратов, согласно которому коэффициенты a и b определяются из условия минимума функции

$$\Phi(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2.$$

Откуда получаем

$$\frac{\partial\Phi}{\partial a}=-2\sum_{i=1}^{n}(y_i-ax_i-b)x_i=0,$$

$$\frac{\partial\Phi}{\partial b}=-2\sum_{i=1}^{n}(y_i-ax_i-b)=0.$$

После преобразований имеем

$$\begin{cases}\sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i^2-b\sum_{i=1}^{n}x_i=0\\ \sum_{i=1}^{n}y_i-a\sum_{i=1}^{n}x_i-bn=0.\end{cases}$$

Решая эту систему уравнений, находим

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} = r\frac{\sigma_y}{\sigma_x},$$

$$b = M_y - aM_{x,}$$

где r – коэффициент корреляции; σ_x, σ_y – средние квадратические отклонения, а M_x, M_y – математические ожидания величин X и Y соответственно.

Если при установлении зависимости между признаками используется больше одной независимой переменной, то применяется множественный регрессионный анализ.

Общее уравнение множественной линейной регрессии, когда один признак зависит от двух факторов, имеет вид:

$$y = a + bx + cz.$$

Согласно методу наименьших квадратов коэффициенты a, b, и c находятся из условия минимума функции

$$\Phi(a,b,c) = \sum_{i=1}^{n} (z_i - ax_i - by_i - c)^2.$$

Откуда получаем систему уравнений:

$$\begin{cases} \sum_{i=1}^{n} x_i z_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i, \\ \sum_{i=1}^{n} y_i z_i = a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i^2 + c\sum_{i=1}^{n} y_i, \\ \sum_{i=1}^{n} z_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i + cn. \end{cases}$$

Решая ее, находим

$$a = \frac{r_{xy}r_{yz} - r_{xz}}{r_{xy}^2 - 1} \cdot \frac{\sigma_z}{\sigma_x},$$

$$b = \frac{r_{xy}r_{xz} - r_{yz}}{r_{xy}^2 - 1} \cdot \frac{\sigma_z}{\sigma_y},$$

$$c = M_z - aM_x - bM_y.$$

Контрольные вопросы:

1. Что такое регрессия?
2. В каких случаях между величинами X и Y существует линейная зависимость?
3. В чем заключается идея метода наименьших квадратов?

4. Напишите систему двух уравнений для определения значений a и b в уравнении y=ax+b.
5. Напишите уравнение множественной линейной регрессии, когда один признак зависит от двух факторов.
6. Напишите формулы для определения коэффициентов a, b и c в уравнении множественной линейной регрессии.

Пример решения задачи:

При изучении зависимости между биомассой трав (z, г/м^2) в агроландшафте, с одной стороны, температурой (x, ^{0}C) и количеством атмосферных осадков (y, мм), с другой, установлена прямая односторонняя зависимость z от x и y. С практической точки зрения целесообразно составить уравнение множественной регрессии, которое можно было бы использовать для прогноза биомассы по температуре и количеству выпавших осадков. Данные по x, y, z представляют собой средние многолетние показатели за период вегетации (май, июнь):

z	300	350	370	420	450	500
x	14,5	15,0	15,6	17,2	18,5	19,3
y	82	95	105	120	130	140

Для вычисления коэффициентов a, b и c уравнения линейной регрессии в Excel в ячейку B4 помещаем формулу =(КОРРЕЛ(B2:G2;B3:G3)*КОРРЕЛ(B3:G3;B1:G1)-КОРРЕЛ(B2:G2;B1:G1))*КОРЕНЬ(ДИСП(B1:G1))/КОРЕНЬ(ДИСП(B2:G2))/(КОРРЕЛ(B2:G2;B3:G3)^2-1), в ячейку B5 - формулу =(КОРРЕЛ(B2:G2;B3:G3)*КОРРЕЛ(B2:G2;B1:G1)-КОРРЕЛ(B3:G3;B1:G1))*КОРЕНЬ(ДИСП(B1:G1))/КОРЕНЬ(ДИСП(B3:G3))/(КОРРЕЛ(B2:G2;B3:G3)^2-1), в ячейку B6 – формулу =СРЗНАЧ(B1:G1)-B4*СРЗНАЧ(B2:G2)-B5*СРЗНАЧ(B3:G3).

	A	B	C	D	E	F	G
1	z	83	72	69	90	90	95
2	x	56	42	18	84	56	107
3	y	82	95	105	120	130	140
4	a=	3,122					
5	b=	3,012					
6	c=	8,914					

Следовательно, общая формула множественной регрессии имеет вид: z=3,122x + 3,012y +8,914.

Уравнения регрессии можно установить также с использованием надстройки (опции) «Пакет анализа». Если по умолчанию эта опция не установлена, то для ее активации необходимо с помощью команды <Надстройки> из меню <Сервис> открыть окно диалога «Надстройки» и в нем установить флажок для компоненты «Пакет анализа». После нажатия кнопки [ОК] меню <Сервис> будет дополнено командой <Анализ данных>.

Выполняют последовательность команд *Сервис/Анализ данных/Регрессия*/в поле «*Входной интервал*» указывают диапазоны значения для Y и X соответственно, в «*Параметрах вывода*» выбирают «*Выходной интервал*» и указывают ячейку на этом же листе, отмечают параметры «*Уровень надежности*» (значение можно изменять, в нашем случае указываем 95%), нажимаем [ОК].

Если удалялся артефакт, то необходимо скопировать первоначальные значения в другие ячейки, поскольку значения во входном интервале должны быть непрерывными.

3. Скопируйте построенный график на другой лист и поэкспериментируйте с различными линиями тренда, посмотрите какие из них наиболее подходят для имеющихся данных.

4. Сделайте выводы по значениям коэффициента корреляции, уравнению регрессии и критерию Стьюдента.

Варианты заданий:

1. Равноточные измерения некоторой величины y, отвечающие ряду значений аргумента x, привели к результатам, приведенным в таблице.

x_i	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5
y_i	1,67	1,32	1,10	0,81	0,48	0,18	-0,10	-0,46	-0,80	-1,15

Предполагая, что теоретически y=ax+b, найти a и b.

2. Путем еженедельного взятия проб с поля изучено изменение высоты растений сои y (в см) с возрастом x (в неделях).

x_i	1	2	3	4	5	6	7
y_i	5	13	16	23	33	38	40

Составьте уравнение регрессии и постройте эмпирическую и теоретическую линии регрессии.

3. Для установления связи между содержанием фосфора в почве и содержанием фосфора в злаковых растениях проведено 9 анализов со следующими результатами.

x_i	1	4	5	9	13	11	22	23	28
y_i	64	71	54	81	93	76	77	95	109

Составьте уравнение регрессии.

4. Методом регрессии провести анализ зависимости толщины ледяного покрова y (см) от суммы среднесуточной температуры воздуха x (^{0}C).

x_i	0,30	0,10	0,08	0,31	0,40	0,50	0,68	0,70
y_i	-10	-5	-4	-10	-12	-16	-30	-31

5. Исследованиями установлено, что на содержание подвижного марганца в почве влияет реакция среды. Получить линейную зависимость содержания подвижного марганца y (мг/кг почвы) от гидролитической кислотности x (мг-экв. на 100 г почвы).

x_i	83	72	69	90	87	95	93	91	75	70
y_i	56	42	18	84	56	107	90	58	31	48

6. Методом регрессии провести анализ зависимости между весом 100 семян ячменя (г) и продолжительностью вегетационного периода ячменя (дни).

y_i	47,5	46,75	45,75	42,85	44,76	41,44	37
x_i	90	85	80	75	70	65	60

ЛАБОРАТОРНАЯ РАБОТА № 5

Установление сходства или различия между выборками

Проведение исследований предполагает не только изучение строения, развития, закономерностей распространения изучаемых объектов, но и установление сходства между одноименными генеральными совокупностями изучаемых систем. Это зависит от условий, в которых протекает один и тот же процесс. Сопряженный анализ одноименных признаков в выборках используется для классификации и районирования по одному или нескольким параметрам.

Если достоверность различия между выборочными совокупностями доказана, то генеральные совокупности, сравниваемые по какому-либо признаку, выделяют как самостоятельные. В случае отсутствия достоверных различий их объединяют в одну группу.

Достоверность различий между генеральными совокупностями может быть определена с помощью критериев достоверности: критерия Стьюдента (*t*) и критерия Фишера (*F*). Сравнение выборочных совокупностей по критерию Стьюдента *t* позволяет утверждать с некоторой долей уверенности сходство или различие между средними выборок по разнице между ними, сравнение по критерию Фишера (F) – между дисперсиями выборок.

Рассмотрим сравнение двух выборочных совокупностей по критерию Стьюдента с использованием формулы:

$$t = d / m_d$$

где d – разность между средними выборок (M_1-M_2), m_d – ошибка разности средних. При расчете разницы между средними из большей величины вычитают меньшую независимо от нумерации выборочных совокупностей.

Выделяют три типа сравниваемых статистических совокупностей: независимые с одинаковым объемом выборок, независимые с разным объемом выборок, сопряженные только с одинаковым объемом выборок.

При одинаковом объеме выборок в случаях независимых статистических совокупностей ошибки средних для каждой выборки в отдельности вычисляют по формуле:

$$m_i = \frac{\sigma}{\sqrt{N}}, \quad где \quad \sigma = \sqrt{\frac{\sum (x_i - M)^2}{N-1}}.$$

Ошибку разности между средними вычисляют по формуле:

$$m_d = \sqrt{m_1^2 + m_2^2},$$

где m_1 – ошибка среднего арифметического первой выборки; m_2 – ошибка среднего арифметического второй выборки.

Число степеней свободы устанавливают следующим образом:

$$\nu = N_1 + N_2 - 2.$$

При разном объеме выборок в сравниваемых совокупностях ошибка разности средних определяется по формуле:

$$m_d = \sqrt{\frac{\sum\left(x_{i_1} - M_1\right)^2 + \sum\left(x_{i_2} - M_2\right)^2}{(N_1 + N_2 - 2)} \frac{N_1 + N_2}{N_1 \cdot N_2}},$$

где $\sum\left(x_{i_1} - M_1\right)^2$ – сумма квадратов отклонений от среднего для первой выборки; $\sum\left(x_{i_2} - M_2\right)^2$ – второй выборки; N_1, N_2 – количество вариант в первой и второй выборках соответственно.

При установлении различий между сопряженными выборками вычисление ошибки разности средних производится по формулам:

$$m_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{N_n(N_n - 1)}}; \quad m_d = \sqrt{\frac{\sum d_i^2 - \left(\sum d_i\right)^2 / N_n}{N_n(N_n - 1)}},$$

где d_i – разность между индивидуальными сопряженными вариантами в выборках; $\bar{d}$ – разность между средними сопряженных выборок; N_n – число сопряженных пар в сопряженных выборках. Число степеней свободы в этом случае находят по равенству $\nu = N_n - 1$.

Сопоставляя рассчитанное значение критерия Стьюдента t_ϕ с теоретическим t_m, приведенным в таблице, устанавливают или отвергают с некоторой долей уверенности различия между средними арифметическими выборок. Если $t_m > t_\phi$, то разность между средними признается несущественной (недостоверной), в противном случае – средние достаточно различны, а различие сравниваемых генеральных совокупностей признается неодинаковым.

Значение критерия Стьюдента t
при различных уровнях вероятности

ν	Уровни вероятности			ν	Уровни вероятности		
	0,95	0,99	0,999		0,95	0,99	0,999
2	4,30	9,93	31,60	21	2,08	2,83	3,82
3	3,18	5,84	12,94	22	2,07	2,82	3,79
4	2,78	4,60	8,61	23	2,07	2,81	3,77
5	2,57	4,03	6,86	24	2,06	2,80	3,75
6	2,45	3,71	5,96	25	2,06	2,79	3,73
7	2,37	3,50	5,41	26	2,06	2,78	3,71
8	2,31	3,36	5,04	27	2,05	2,77	3,69
9	2,26	3,25	4,78	28	2,05	2,76	3,67
10	2,23	3,17	4,59	29	2,04	2,76	3,66
11	2,20	3,11	4,44	30	2,04	2,75	3,65

12	2,18	3,06	4,32	40	2,02	2,70	3,55
13	2,16	3,01	4,22	50	2,01	2,68	3,50
14	2,15	2,98	4,14	60	2,00	2,66	3,46
15	2,13	2,95	4,07	80	1,99	2,64	3,42
16	2,12	2,92	4,02	100	1,98	2,63	3,39
17	2,11	2,90	3,97	120	1,98	2,62	3,37
18	2,10	2,88	3,92	200	1,97	2,60	3,34
19	2,09	2,86	3,88	500	1,96	2,59	3,31
20	2,09	2,85	3,85	∞	1,96	2,58	3,29

Контрольные вопросы:

1. Для решения каких задач применяют критерии Стьюдента и Фишера?
2. По какому статистическому параметру проводится сравнение выборочных совокупностей по критерию Стьюдента *t*?
3. Какие типы статистических совокупностей выделяют при их сравнении по критерию Стьюдента и критерию Фишера? В чем разница в вычислении критериев в каждом возможном случае?

Пример решения задачи:

При исследовании глубины расчленения рельефа (м) в Воложинском районе N_1 и Браславском районе N_2, результаты которого приведены в таблице, необходимо установить, объединять рассматриваемые участки в один геоморфологический район по степени расчленения рельефа или различать их как самостоятельные.

Воложинский район	20	17	16	15	15
Браславский район	17	16	15	14	14

Сравним независимые выборки с одинаковыми объемами (N_1 и N_2) по значениям средних арифметических (M_1 и M_2). Для этого заполним в Excel таблицу с исходными данными и подсчитаем в ней значения M_1 и M_2, $(x_{i1}\text{-}M_1)^2$ и $(x_{i2}\text{-}M_2)^2$, m_1, m_2 и m_d, необходимые для вычисления критерия Стьюдента.

	A	B	C	D	E	F	G	H
1	**Результаты исследования глубины расчленения рельефа (м)**							
2		**N_1, x_i**	$(x_{i1}\text{-}M_1)^2$	**N_2,x_i**	$(x_{i2}\text{-}M_2)^2$			
3	x_1	20	11,56	17	3,24			
4	x_2	17	0,16	16	0,64			
5	x_3	16	0,36	15	0,04			
6	x_4	15	2,56	14	1,44			
7	x_5	15	2,56	14	1,44			

8		M_1=		M_2=				
9		16,6	17,20	15,2	6,80			
10								
11	***d*=**	**1,4**						
12	$\boldsymbol{m_1}$**=**	**0,93**						
13	$\boldsymbol{m_2}$**=**	**0,58**						
14	$\boldsymbol{m_d}$**=**	**1,2**						
15	$\boldsymbol{t_ф}$**=**	**1,17**						

В ячейке B9 вычисляем M_1 по формуле =СРЗНАЧ(B3:B7), в ячейке D9 вычисляем M_2 по формуле =СРЗНАЧ(D3:D7). В ячейках C3:C7 вычисляем значения $(x_{i1}-M_1)^2$ по формулам =СТЕПЕНЬ(B3-B9;2), =СТЕПЕНЬ(B4-B9;2) … =СТЕПЕНЬ(B7-B9;2). В ячейки E3:E7 заносим значения $(x_{i2}-M_2)^2$, вычисленные по формулам =СТЕПЕНЬ(D3-D9;2), =СТЕПЕНЬ(D4-D9;2) … =СТЕПЕНЬ(D7-D9;2). В ячейках C9 и E9 подсчитываем значения сумм $(x_{i1}-M_1)^2$ и $(x_{i2}-M_2)^2$ по формулам =СУММ(C3:C7) и =СУММ(E3:E7).

По средним арифметическим (M_1=16,6 м и M_2=15,2 м) различие по глубине расчленения рельефа можно признать как существенным, так и несущественным. Для объективных выводов используем критерий Стьюдента. В ячейке B11 определим разницу между средними выборок по формуле =B9-D9.

В ячейке B12 и B13 помещаем значение m_1, m_2, равные

$$m_1 = \sqrt{\frac{\sum\left(x_{i_1} - M_1\right)^2}{N_1(N_1 - 1)}}; m_2 = \sqrt{\frac{\sum\left(x_{i_2} - M_2\right)^2}{N_2(N_2 - 1)}}.,$$

вычисленные по формулам =C9/20 и =E9/20, соответственно.

Ошибку разности средних $m_d = \sqrt{m_1^2 + m_2^2}$ вычисляем в ячейке B14 по формуле =КОРЕНЬ(СТЕПЕНЬ(B12;2)+СТЕПЕНЬ(B13;2)).

Фактический критерий Стьюдента, вычисленный по формуле =B11/B14, помещаем в ячейку B15. Он составляет 1,17.

Установив число степеней свободы,

$$(\nu = N_1 + N_2 - 2 = 5 + 5 - 2 = 8),$$

сопоставляем табличные значения критерия Стьюдента t_m=2,32 и 3,36 (см. таблицу) при уровне вероятности P=0,95 и 0,99 для ν=8 с расчетным (1,17). Поскольку $t_т > t_ф$, то разность между средними признается несущественной (недостоверной). Следовательно, при выделении геоморфологических районов по глубине расчленения рельефа, рассматриваемые участки необходимо объединить в один геоморфологический район.

ЛАБОРАТОРНАЯ РАБОТА №5 (дополнение)

Установление сходства или различия между выборками

1. Нахождение сходства или отличия между двумя выборками с помощью t-теста (критерия Стьюдента).

Выбор конкретной команды зависит от типа выборки (зависимая/независимая и от величин дисперсий.

Так для **независимой** выборки с **различными дисперсиями** выполняются следующие действия: *Сервис*/*Анализ данных*/*Двухвыборочный t-тест с различными дисперсиями/ОК*/.

Для **независимой** выборки с **одинаковыми дисперсиями** алгоритм следующий: *Сервис*/*Анализ данных*/*Двухвыборочный t-тест с одинаковыми дисперсиями/ОК*/.

Для **сопряженной** выборки: *Сервис*/*Анализ данных*/*Парный двухвыборочный t-тест для средних/ОК*/.

В поле «*интервал переменной* 1» указываем интервал значений для первой области (B2:H2), в поле «*интервал переменной* 2» – интервал значений для второй области (B3:H3), далее выбираем «*Выходной интервал*» и указываем там ячейку G5, нажимаем [ОК].

Сделать выводы по указанным переменным.

Варианты заданий:

1. Для снижения затрат на разведку на одном из участков россыпного месторождения золота часть шурфов заменили скважинами. Определите, имеют ли результаты опробования скважин систематическую ошибку.

Результаты опробования разведочных выработок на россыпном месторождении золота

Выработка А (скважины)		Выработка А (скважины)		Выработка Б (шурфы)		Выработка Б (шурфы)	
№ п/п	Содержание Au, $м^2/м^3$	№ п/п	Содержание Au, $м^2/м^3$	№ п/п	Содержание Au, $мг/м^3$	№ п/п	Содержание Au, $мг/м^3$
1	322	8	375	1	431	6	221
2	250	9	381	2	397	7	548
3	225	10	538	3	462	8	478
4	315	11	198	4	457	9	299
5	399	12	317	5	251	10	541
6	348	13	293				
7	192						

На основании результатов сравнения средних значений содержания золота по скважинам и шурфам сделайте вывод о возможности использования скважинных методов на данном месторождении.

2. Можно ли отнести сосны и березы, произрастающие в пределах изучаемого участка леса, к одному ярусу по величине средних значений их высоты (в метрах), если были получены значения, занесенные в таблицу.

Средние значения высоты деревьев в пределах изучаемого участка леса (в метрах)

Сосны	4	5	4	5	5	4	5	4	3	5	4	5
	6	1	6	4	4	4	5	5	3	5	5	4
	6	4	6	2	3	4	5	5	5	5	5	5
	4	5	5	6	4	6	2	5	5	3	5	5
	5	4	6	4	5	5	5	5	5	5	5	5
Березы	5	5	4	6	7	6	3	5	5	6	5	5
	5	4	4	2	4	4	6	2	6	5	4	5
	5	5	5	5	4	5	4	6	5	4	7	5
	5	5	6	6	4	4	4	6	5	4	3	5
	5	7	5	5	5	5	4	3	7	6	4	4

3. Можно ли считать симметричными склоны изучаемого участка оврага, если их крутизна (в градусах) составляет следующие значения.

Значения крутизны склонов оврага на изучаемом участке (в градусах)

Правый склон	7	5	6	3	6	7	4	5	8	6	3	3	4	3	7	4	4	4	5	3
	8	10	6	3	3	6	5	2	5	3	11	3	7	4	7	3	5	5	3	4
	1	3	7	2	5	5	5	3	3	4	6	5	6	1	6	4	4	4	6	4
Левый склон	4	2	5	4	8	6	3	4	6	5	2	6	6	1	2	2	2	5	2	2
	5	9	3	5	6	4	6	5	7	1	3	6	5	4	2	8	9	4	5	3
	2	2	11	4	6	6	4	6	2	5	3	5	7	2	6	5	5	1	2	7

4. В результате геоморфологических исследований двух соседних равных по площади участков были получены значения густоты расчленения рельефа (км/км^2).

Густота расчленения рельефа (км/км^2)

Участок № 1									
8,2	9,7	5,6	7,4	8,0	6,4	6,6	6,8	8,4	7,1
9,0	6,0	7,6	8,1	11,8	5,8	9,3	7,3	8,2	7,2
7,2	6,4	7,7	9,0	8,1	7,1	7,1	8,8	7,5	9,2
7,5	6,8	7,0	6,4	7,4	8,2	6,3	7,0	8,1	10,0
7,0	7,1	8,7	6,3	8,6	7,7	7,3	8,0	8,4	9,3
Участок № 2									
7,3	6,0	7,7	6,1	9,6	7,4	7,2	7,2	8,7	7,5
9,1	6,4	8,3	6,5	8,2	7,2	6,9	6,9	8,2	9,0
7,4	8,0	8,4	7,0	7,1	7,4	6,6	6,4	8,3	7,9
8,3	7,2	7,2	6,6	6,6	7,7	8,7	5,6	7,5	5,7
6,9	7,4	7,2	6,2	6,9	6,8	9,2	9,2	7,1	6,5

Можно ли объединить рассматриваемые участки в один геоморфологический район по степени расчленения рельефа?

5. В одном из перспективных на олово районов при вскрытии геохимических аномалий выявлены два участка оловянной минерализации. Оруденение представлено маломощными жилами и прожилками кварца с касситеритом. Для уточнения направления дальнейших геологоразведочных работ необходимо провести статистический анализ замеров элементов залегания жильных образований.

Замеры элементов залегания кварцевых жил с касситеритом

Азимуты падения на первом участке, град.									
190	192	176	230	308	290	198	154	357	338
170	174	88	150	90	336	292	180	110	296
262	204	310	250	142	48	166	214	280	274
194	222	258	22	314	186	114	108	354	308
Азимуты падения на втором участке, град.									
120	114	102	82	82	74	308	158	162	130
68	180	202	128	102	90	96	72	60	56
142	134	150	112	118	142	354	230	178	160
160	112	100	90	88	60	50	4	24	108

6. По распределению сумм активных температур воздуха выше 10°С по территории различных областей Республики Беларусь сравните агроклиматические условия двух соседних областей и сделайте вывод.

Распределение сумм активных температур воздуха выше 10°С областей Республики Беларусь

Брестская	2347	2445	2596	2340	2550	2524	2522	2434	2403
	2549	2455	2339	2605	2521	2435	2482	2472	2340
Гомельская	2425	2427	2442	2479	2469	2456	2454	2419	2482
	2341	2470	2440	2442	2436	2426	2481	2478	2335
Могилевская	2272	2342	2271	2370	2164	2193	2251	2243	2281
	2278	2250	2229	2241	2205	2363	2321	2215	2248
Минская	2302	2252	2242	2201	2189	2328	2329	2207	2197
	2377	2241	2249	2131	2327	2254	2334	2251	2375
Гродненская	2337	2233	2363	2240	2310	2225	2210	2278	2365
	2170	2096	2218	2326	2300	2173	2322	2267	2320
Витебская	2204	2173	2128	2220	2176	2068	2078	2192	2190
	2192	2167	2205	2140	2076	2089	2209	2189	2168

ЛИТЕРАТУРА

1. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. – СПб.: Лань, 1998. – 224 с.
2. Гончаров Л. Excel 7.0 в примерах. – СПб: Питер, 1996.
3. Забелин Н.Н. Сборник задач по теории вероятностей и математической статистике: учебное пособие. – Гродно: Изд-во Гродненского филиала негосударственного института современных знаний, 1998. – 159 с.
4. Зайцев Н.Г. Математическая статистика в экспериментальной ботанике. – М.: Наука, 1984. – 424 с.
5. Пузаченко Ю.Г. Математические методы в экологических и географических исследованиях: учеб. пособие для студ. вузов / Ю.Г. Пузаченко. – М.: Издательский центр «Академия», 2004.
6. Рокицкий П.Ф. Биологическая статистика. – Минск, 1973. – 320 с.
7. Чертко Н.К. Математические методы в географии: учеб.-метод. пособие / Н.К. Чертко, А.А. Карпиченко. – Минск: БГУ, 2009.
8. Microsoft Excel 2002: справ. пособие / В.А. Долженков, Ю.В. Колесников. – СПб.: БХВ-Петербург, 2003. – 1053 с.

Учебное издание

ЛАБОВКИН Владимир Никитович
КРАСОВСКАЯ Ирина Анатольевна

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЙ В ГЕОГРАФИИ

Методические рекомендации

Технический редактор *Г.В. Разбоева*
Компьютерный дизайн *Л.Р. Жигунова*

Подписано в печать . .2013. Формат 60x84 $^{1}/_{16}$. Бумага офсетная.
Усл. печ. л. 2,09. Уч.-изд. л. 1,27. Тираж экз. Заказ .

Издатель и полиграфическое исполнение – учреждение образования
«Витебский государственный университет имени П.М. Машерова».
ЛИ № 02330/110 от 30.01.2013.

Отпечатано на ризографе учреждения образования
«Витебский государственный университет имени П.М. Машерова».
210038, г. Витебск, Московский проспект, 33.