

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Белгородский государственный технологический университет
им. В.Г. Шухова

Информационные технологии в строительной индустрии

Методические указания к выполнению лабораторных работ
для студентов направления 08.04.01– Строительство,
профиля «Дорожно-строительное материаловедение»

Белгород
2017

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Белгородский государственный технологический университет
им. В.Г. Шухова
Кафедра автомобильных и железных дорог

Утверждено
научно-методическим советом
университета

Информационные технологии в строительной индустрии

Методические указания к выполнению лабораторных работ
для студентов направления 08.04.01– Строительство,
профиля «Дорожно-строительное материаловедение»

Белгород
2017

УДК 004.9:625.1(075.8)

ББК У311-21-с 51

И 74

Составители: канд. техн. наук, доц. С.А. Гнездилова

ст. преп. А.С. Погромский

Рецензент канд. техн. наук, доц. А.И. Траутвайн

И74 **Информационные** технологии в строительной индустрии: методические указания к выполнению лабораторных работ / сост.: С.А. Гнездилова, А.С. Погромский. – Белгород: Изд-во БГТУ, 2017. –23 с.

В методических указаниях приведены цель и содержание лабораторных работ, а также указания по их выполнению.

Методические указания предназначены для студентов направления 08.04.01–Строительство, профиля «Дорожно-строительное материаловедение».

Публикуется в авторской редакции.

УДК 004.9:625.1(075.8)

ББК У311-21-с 51

© Белгородский государственный
технологический университет
(БГТУ) им. В.Г. Шухова, 2017

Лабораторная работа № 1

Понятие математической статистики

Цель работы: построить эмпирическое распределение содержания SiO_2 в неогеновых лавах для выборки.

Исходные данные: исходные данные приведены в табл. 1 [1].

Таблица 1

Частотное распределение содержаний SiO_2 в неогеновых лавах

Содержание SiO_2		Класс, от – до	Частота	Частость, %	Накопленная частость, %
целые числа	десятые доли				
56	6				
57		56,0–58,9	1	3	3
58					
59	5				
60	5,7	59,0–61,9	6	20	23
61	2,2,1				
62	9,4				
63	7,1,6,8	62,0–64,9	8	27	50
64	6,6				
65	8,3				
66	8,3	65,0–67,9	7	23	73
67	8,7,5				
68					
69	2,3,9	68,0–70,9	3	10	83
70					
71	6,4				
72	5	71,0–73,9	5	17	100
73	2,2				

Порядок выполнения работы

1. В ячейку A1 введите заголовок *Наблюдения*, а в диапазон A2:A31 – значения (в %) содержания SiO_2 в неогеновых лавах. В ячейке A32 подсчитайте максимальное значение, в ячейке A33 – минимальное значение [2].

2. Выберите ширину интервала 1%. Тогда при крайних значениях 56% и 74% получится 18 интервалов. В ячейку D1 введите название *Границы интервалов*. В диапазон D2:D20 введите через меню **Прав-**

ка/Заполнить/Прогрессия... числовой ряд от 56 до 74 (56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74).

3. Введите заголовки создаваемой таблицы: в ячейки E1 – *Абсолютные частоты*, в ячейки F1 – *Относительные частоты*, в ячейки G1 – *Накопленные частоты*.

4. Заполните столбец абсолютных частот. Для этого выделите для них блок ячеек E2:E20 (используемая функция ЧАСТОТА задается в виде формулы массива). С панели инструментов **Стандартная** вызовите **Мастер функций** (кнопка f_x). В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию ЧАСТОТА, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно ЧАСТОТА необходимо за серое поле мышью отодвинуть вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши в рабочее поле **Массив_данных** введите диапазон данных наблюдений (A2:A31). В рабочее поле **Двоичный_массив** мышью введите диапазон интервалов (D2:D20). Последовательно нажмите комбинацию клавиш Ctrl+Shift+Enter. В столбце E2:E20 появится массив абсолютных частот.

5. В ячейке E21 найдите общее количество наблюдений. Табличный курсор установите в ячейку E21. На панели инструментов **Стандартная** нажмите кнопку **Автосумма**. Убедитесь, что диапазон суммирования указан правильно (E2: E20), и нажмите клавишу Enter. В ячейке E21 появится число 30.

6. Заполните столбец относительных частот. В ячейку F2 введите формулу для вычисления относительной частоты: =E2/E\$21. Нажмите клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) скопируйте введенную формулу в диапазон F3:F20. Получите массив относительных частот.

7. Заполните столбец накопленных частот. В ячейку G2 скопируйте значение относительной частоты из ячейки F2. В ячейку G3 введите формулу: =G2+F3. Нажмите клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) скопируйте введенную формулу в диапазон G3:G20. Получите массив накопленных частот.

8. В результате, после форматирования получим таблицу, представленную на рис. 1.

9. Постройте диаграмму относительных и накопленных частот. Щелчком указателя мыши по кнопке на панели инструментов вызовите **Мастер диаграмм**. В появившемся диалоговом окне выберите вкладку **Нестандартные** и тип диаграммы **График/гистограмма2**.

После нажатия кнопки **Далее** укажите диапазон данных – F1:G20 (с помощью мыши). Проверьте положение переключателя **Ряды в: столбцах**. Выберите вкладку **Ряд** и с помощью мыши введите в рабочее поле **Подписи оси X** диапазон подписей оси X: D2:D20. Нажав кнопку **Далее**, введите названия осей X и Y в рабочее поле. **Ось X** (категорий) – *Наблюдения*, **Ось Y** (значений) – *Относ. частота*, **Вторая ось Y** (значений) – *Накоплен. частота*. Нажмите кнопку **Готово**.

D	E	F	G
Границы интервалов	Абсолютные частоты	Относительные частоты	Накопленные частоты
56	0	0	0
57	1	0,033333333	0,033333333
58	0	0	0,033333333
59	0	0	0,033333333
60	1	0,033333333	0,066666667
61	2	0,066666667	0,133333333
62	3	0,1	0,233333333
63	2	0,066666667	0,3
64	4	0,133333333	0,433333333
65	2	0,066666667	0,5
66	2	0,066666667	0,566666667
67	2	0,066666667	0,633333333
68	3	0,1	0,733333333
69	0	0	0,733333333
70	3	0,1	0,833333333
71	0	0	0,833333333
72	2	0,066666667	0,9
73	1	0,033333333	0,933333333
74	2	0,066666667	1
	30		

Рис. 1. Результат вычислений относительных и накопленных частот

Диаграмма будет иметь такой вид, как на рис. 2.

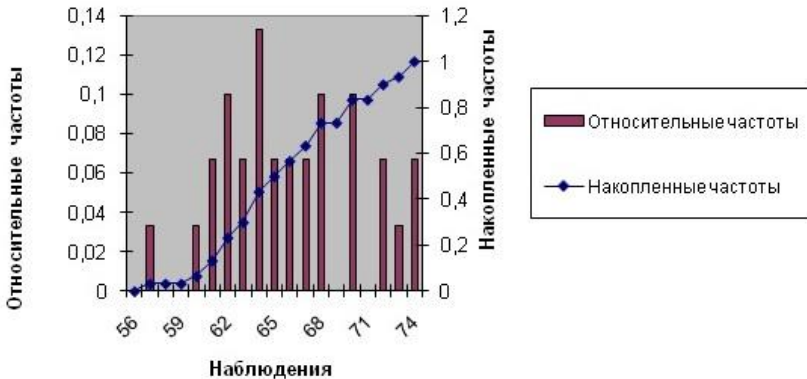


Рис. 2. Диаграмма относительных и накопленных частот

Лабораторная работа № 2

Определение основных статистических характеристик

Цель работы: используя мастер функций *Excel* вычислить выборочные характеристики распределения экспериментальных данных.

Исходные данные: используя исходные данные Лабораторной работы №1, найти характеристики распределения этих данных.

Порядок выполнения работы

1. В ячейку A1 введите заголовок *Наблюдения*, а в диапазон A2:A31 – значения содержания SiO_2 в неогеновых лавах. Отметим, что рассматриваемая группа данных со статистической точки зрения являются выборкой.

2. При статистическом анализе прежде всего необходимо определить характеристики выборки, и важнейшей характеристикой является среднее значение. Для определения среднего значения в контрольной группе необходимо установить табличный курсор в свободную ячейку (A34). На панели инструментов нажмите кнопку **Вставка функции (f_x)**. В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию СРЗНАЧ, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно СРЗНАЧ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных группы для определения среднего значения (A2:A31). Нажмите кнопку **ОК**. В ячейке A34 появится среднее значение выборки – 65,5667.

3. Следующей по важности характеристикой выборки является мера разброса элементов выборки от среднего значения. Такой мерой является среднее квадратичное или стандартное отклонение. Для определения стандартного отклонения в контрольной группе необходимо установить табличный курсор в свободную ячейку (A35). На панели инструментов нажмите кнопку **Вставка функции (f_x)**. В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию СТАНДОТКЛОНП, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно СТАНДОТКЛОНП за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных контрольной группы для определения стандартного отклонения (A2:A31). Нажмите кнопку

ОК. В ячейке A35 появится стандартное отклонение выборки – 4,307. В ячейке B35 по формуле $=3*\text{СТАНДОТКЛОНП}(A2:A31)$ подсчитайте значение $3*\sigma$: 12,922, где σ – стандартное отклонение. Существует правило, согласно которому для распределения близкого к нормальному данные должны лежать в диапазоне $\bar{x} \pm 3\sigma$ (в примере $65,567 \pm 12,922$). Это правило называется правилом трех сигм. Сущность этого правила: *если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения* [1]. Согласно этому правилу, в пределах $\bar{x} \pm 3\sigma$ находится 99,7% всех вариант.

4. В свободных ячейках A36 – A40 рассчитайте выборочную дисперсию, медиану, моду, асимметрию, эксцесс.

5. Сформулировать основные выводы по работе.

Лабораторная работа № 3

Проверка статистических гипотез

Цель работы: проверить соответствие выборочных данных нормальному закону распределения, используя пакет *Excel* с использованием критерия согласия хи-квадрат, построить гистограммы абсолютных и теоретических частот.

Исходные данные: используя исходные данные Лабораторной работы №1, проверить соответствие выборочных данных эмпирического распределения содержания SiO_2 в неогеновых лавах нормальному закону распределения, используя пакет *Excel*.

Порядок выполнения работы

1. В ячейку A1 введите заголовок *Наблюдения*, а в диапазон A2:A31 – значения (в %) содержания SiO_2 в неогеновых лавах. В ячейке A32 подсчитайте максимальное значение, в ячейке A33 – минимальное значение.

2. Выберите ширину интервала 1%. Тогда при крайних значениях 56% и 74% получится 18 интервалов. В ячейку D1 введите название *Границы интервалов*. В диапазон D2:D20 введите через меню **Прав-**

ка/Заполнить/Прогрессия... числовой ряд от 56 до 74 (56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74).

3. Введите заголовки создаваемой таблицы: в ячейки E1 – *Абсолютные частоты*, в ячейки F1 – *Относительные частоты*.

4. Заполните столбец абсолютных частот. Для этого выделите для них блок ячеек E2:E20 (используемая функция ЧАСТОТА задается в виде формулы массива). С панели инструментов **Стандартная** вызовите **Мастер функций** (кнопка f_x). В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию ЧАСТОТА, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно ЧАСТОТА необходимо за серое поле мышью отодвинуть вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши в рабочее поле **Массив_данных** введите диапазон данных наблюдений (A2:A31). В рабочее поле **Двоичный_массив** мышью введите диапазон интервалов (D2:D20). Последовательно нажмите комбинацию клавиш Ctrl+Shift+Enter. В столбце E2:E20 появится массив абсолютных частот.

5. В ячейке E21 найдите общее количество наблюдений. Табличный курсор установите в ячейку E21. На панели инструментов **Стандартная** нажмите кнопку **Автосумма**. Убедитесь, что диапазон суммирования указан правильно (E2: E20), и нажмите клавишу Enter. В ячейке E21 появится число 30.

6. Заполните столбец относительных частот. В ячейку F2 введите формулу для вычисления относительной частоты: =E2/E\$21. Нажмите клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) скопируйте введенную формулу в диапазон F3:F20. Получите массив относительных частот.

7. Найдите теоретические частоты нормального распределения. Для этого предварительно необходимо найти среднее значение и стандартное отклонение выборки.

В ячейке A34 с помощью функции СРЗНАЧ найдите среднее значение для данных из диапазона A2:A31 (65,57). В ячейке A35 с помощью функции СТАНДОТКЛОН найдите стандартное отклонение для этих же данных (4,38). В ячейки G1 введите название столбца – *Теоретические частоты*. Затем с помощью функции НОРМРАСП найдите теоретические частоты. Установите курсор в ячейку G2, вызовите указанную функцию и заполните ее рабочие поля: x – D2; *Среднее* – \$A\$34; *Стандартное_откл* – \$A\$35. *Интегральный* – 0. Получим в ячейке G2 – 0,0083. Далее протягиванием скопируйте содержимое ячейки G2 в диапазон ячеек G3:G20. Затем в ячейки H1 введите назва-

ние нового столбца – *Теоретические частоты*. Установите курсор в ячейку H2 и введите формулу =E\$21*G2. Далее протягиваем скопированное содержимое ячейки H2 в диапазон ячеек H3:H20. В ячейке H21 с помощью функции СУММ найдите среднее значение для данных из диапазона H2:H20 (29,07). В результате у вас получится как на рис. 3.

8. С помощью функции ХИ2ТЕСТ определите соответствие данных нормальному закону распределения. Для этого установите табличный курсор в свободную ячейку I21. На панели инструментов **Стандартная** нажмите кнопку **Вставка функции (fx)**. В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию ХИ2ТЕСТ, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно ХИ2ТЕСТ отодвиньте вправо на 1–2 см от данных. Указателем мыши в рабочие поля введите фактический E2:E20 и ожидаемый H2:H20 диапазоны частот. Нажмите кнопку **ОК**. В ячейке I21 появится значение вероятности того, что выборочные данные соответствуют нормальному закону распределения – 0,56.

9. Поскольку полученная вероятность соответствия экспериментальных данных $p = 0,56$ много больше, чем уровень значимости $\alpha = 1 - p = 0,05$, то можно утверждать, что нулевая гипотеза не может быть отвергнута и, следовательно, данные не противоречат нормальному закону распределения.

D	E	F	G	H	I
Границы интервалов	Абсолютные частоты	Относительные частоты	Теоретические частоты	Теоретические частоты	
56	0	0	0,008393124	0,251793729	
57	1	0,033333333	0,013460805	0,403824141	
58	0	0	0,020492354	0,614770612	
59	0	0	0,029613253	0,888397588	
60	1	0,033333333	0,040621302	1,218639063	
61	2	0,066666667	0,052892612	1,586778369	
62	3	0,1	0,06537469	1,961240712	
63	2	0,066666667	0,076700417	2,301012512	
64	4	0,133333333	0,085419939	2,562598176	
65	2	0,066666667	0,09030135	2,709040497	
66	2	0,066666667	0,090615539	2,718466176	
67	2	0,066666667	0,086314661	2,589439821	
68	3	0,1	0,07804407	2,341322092	
69	0	0	0,066983633	2,009509	
70	3	0,1	0,054572135	1,637164053	
71	0	0	0,042203322	1,266099674	
72	2	0,066666667	0,030981026	0,929430769	
73	1	0,033333333	0,021588297	0,647648921	
74	2	0,066666667	0,014279547	0,428386424	
	30			29,06556233	0,55955

Рис. 3. Результат вычислений относительных и теоретических частот

10. Построить гистограммы абсолютных и теоретических частот (см. рис.4).

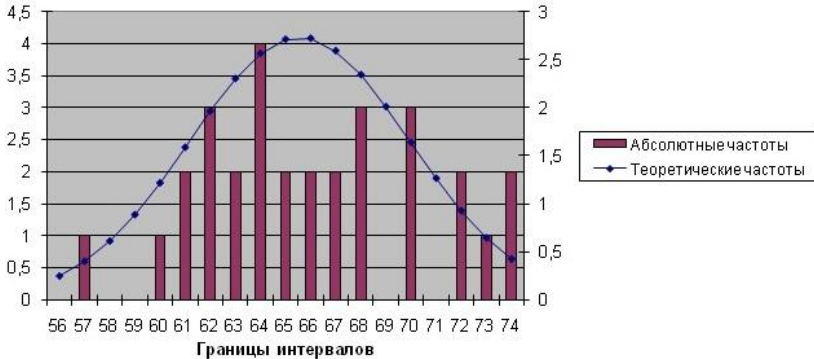


Рис.4. Гистограммы абсолютных и теоретических частот

11. Сформулировать основные выводы по работе.

Лабораторная работа № 4

Принятие статистических решений (анализ одной выборки, анализ двух выборок)

Цель работы: выполнить анализ двух выборок с помощью t -теста Стьюдента.

Исходные данные: в районе широко развиты интрузии гранитов средне- и поздне-палеозойского возраста. Со среднепалеозойскими гранитами связаны месторождения редкометалльных пегматитов. Среднепалеозойские граниты отличаются от нерудоносных верхнепалеозойских повышенным содержанием Na_2O и Ti_2O и пониженным содержанием K_2O (табл. 2). Распределение содержаний Na_2O и K_2O в гранитах обоих комплексов соответствует нормальному закону, а содержание Ti_2O – логнормальному.

Таблица 2

Исходные данные оценки рудоносности гранитов
неизвестного возраста (в %)

Возраст гранитов	Число проб	Na ₂ O		K ₂ O		Ti ₂ O	
		\bar{x}	S ²	\bar{x}	S ²	lg \bar{x}	S ² _{lg x}
Средний палеозой	100	3,90	1,21	4,51	1,42	-0,886	0,268
Поздний палеозой	100	3,46	1,52	5,02	1,65	-1,426	0,321
Неизвестен	30	3,38	1,83	4,83	1,88	-1,352	0,225

Порядок выполнения работы

В нижней строке табл. 2 приведены числовые характеристики содержания оксидов по одной из интрузий неизвестного возраста. Ее сопоставление с интрузиями определенного возраста по содержанию Na₂O и K₂O можно провести с помощью критерия Стьюдента, а по содержанию Ti₂O с помощью критерия Родинона.

В данной задаче проверяемая гипотеза H_0 состоит в том, что \bar{X} и \bar{Y} равны, то есть $H_0: \bar{x} = \bar{y}$, при альтернативе $H_1: \bar{x} \neq \bar{y}$, то есть $\bar{x} > \bar{y}$ или $\bar{x} < \bar{y}$. Поэтому уровень значимости α принимается для двусторонних критериев t_α и $Z_{1-\alpha/2}^2$.

При сравнении интрузий неизвестного возраста с рудоносными среднепалеозойскими гранитами ошибка первого рода будет состоять в том, что фактически перспективная интрузия того же возраста будет признана нерудоносной, и поисковые работы на этом участке будут прекращены, что может привести к пропуску месторождения. Ошибка второго рода произойдет, если фактически нерудоносная интрузия другого возраста будет отнесена к рудоносному комплексу, поисковые работы будут продолжены, но не дадут положительного результата. Экономический ущерб от этой ошибки в данном случае заключается в непроизводительных затратах на поисковые работы в пределах оцениваемой локальной площади. Определить ущерб от ошибки первого рода сложнее, но он может быть гораздо значительнее, так как пропуск месторождения обычно можно восполнить только путем изучения гораздо больших по площади территорий в другом районе. Поэтому при проверке данной гипотезы уровень значимости, то есть вероятность ошибки первого рода, целесообразно принять довольно низким – 0,05,

несмотря на то, что при этом увеличивается вероятность ошибки второго рода.

Сравнивая ту же интрузию с заведомо нерудоносными верхнепалеозойскими гранитами, наоборот следует стремиться снизить вероятность ошибки второго рода. Поэтому уровень значимости целесообразно увеличить до 0,1.

Сравнение интрузий неизвестного возраста со среднепалеозойскими гранитами производится по формулам (1) и (2):

$$\text{по Na}_2\text{O} \left(t = |3,90 - 3,38| / \sqrt{\frac{1,21}{100} + \frac{1,83}{30}} = 1,92 \right), \quad (1)$$

$$\text{по K}_2\text{O} \left(t = |4,51 - 4,83| / \sqrt{\frac{1,42}{100} + \frac{1,88}{30}} = 1,15 \right), \quad (2)$$

$$\text{по Ti}_2\text{O} \left(Z = \frac{\{|-0,886 + 1,352| + 1,153|0,268 - 0,225|\}}{\sqrt{\frac{0,268}{100} + \frac{0,225}{30} + 2,65 \left(\frac{0,268^2}{100-1} + \frac{0,225^2}{30-1} \right)}} = 3,98 \right).$$

Для уровня значимости $\alpha = 0,05$ критическое значение критерия Стьюдента при числе степеней свободы $n_1 + n_2 - 2 = 128$ равно 1,98.

Таким образом, гипотеза об идентичности средних содержаний Na_2O отвергается, в то время как различие по содержанию K_2O можно признать несущественным. Критическое значение Z для тех же условий равно 1,96 [1], поэтому гипотеза о равенстве содержаний также отвергается. Последняя гипотеза отвергается даже в том случае если уровень значимости снизить до 0,001 ($Z_{0,9995} \cong 3,3$).

Критическое значение Z может быть найдено с помощью формулы =НОРМСТОБР(0,025).

Сравнение гранитов неизвестного возраста с верхнепалеозойскими гранитами производится по формулам (1) и (2):

$$\text{по Na}_2\text{O} \left(t = |3,46 - 3,38| / \sqrt{\frac{1,52}{100} + \frac{1,83}{30}} = 0,29 \right);$$

$$\text{по K}_2\text{O} \left(t = |5,02 - 4,83| / \sqrt{\frac{1,65}{100} + \frac{1,88}{30}} = 0,67 \right);$$

$$\text{по Ti}_2\text{O} \left(Z = \frac{\{|-1,426 + 1,352| + 1,153|0,321 - 0,225|\}}{\sqrt{\frac{0,321}{100} + \frac{0,225}{30} + 2,65 \left(\frac{0,321^2}{100-1} + \frac{0,225^2}{30-1} \right)}} = 1,37 \right).$$

Критическое значение критерия Стьюдента (t) и критерия Z для уровня значимости $\alpha = 0,1$ равны: $t_{0,1;128} \cong 1,66$; $Z_{0,95} \cong 1,645$.

Поэтому гипотеза об идентичности химического состава изучаемой интрузии с верхнепалеозойскими гранитами принимается для всех трех компонентов.

По совокупности проверенных гипотез можно достаточно уверенно дать отрицательную оценку перспектив рудоносности в пределах данной интрузии.

Вычисления следует организовать в электронных таблицах *Excel*. Для вычисления модуля числа используйте формулу $=\text{ABS}(\text{число})$, которая возвращает модуль (абсолютную величину) числа. Абсолютная величина числа – это число без знака.

Лабораторная работа № 5

Использование инструмента Пакет анализа для выявления различий между выборками

Цель работы: выявить различия между двумя выборками, используя непараметрический критерий Ван-дер-Вардена.

Исходные данные: Для снижения затрат на разведку на одном из участков россыпного месторождения золота часть шурфов (примерно каждый второй) была заменена скважинами ударно-канатного бурения. Необходимо убедиться в том, что результаты опробования скважин не имеют систематической ошибки. Так как число скважин и шурфов на опытном участке невелико – 13 и 10, для сопоставления вычисленных по ним содержаний золота можно использовать непараметрический критерий Ван-дер-Вардена. Результаты опробования шурфов и скважин приведены в табл. 3. По этим данным составлен общий вариационный ряд (табл. 4).

Таблица 3

Результаты опробования разведочных выработок на россыпном месторождении золота

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

Проверяемая гипотеза H_0 в данной задаче заключается в том, что систематических расхождений в определении содержаний золота по шурфам и скважинам нет (то есть $H_0: X=0, A=B$), при альтернативе $H_1: X \neq 0, A \neq B$ (то есть $A > B$ или $A < B$). Поэтому для нахождения критического значения X нужно воспользоваться таблицами двустороннего критерия Ван-дер-Вардена [3].

Таблица 4

Проверка гипотезы о равенстве средних содержаний золота по критерию Ван-дер-Вардена

№ п/п	Содержание	Выборка	$i/(n+1)$	$\psi(i/(n+1))$
1	2	3	4	5
1	192	А		
2	198	А		
3	221	Б	0,125	-1,15
4	225	А		
5	250	А		
6	251	Б	0,250	-0,67
7	293	А		
8	299	Б	0,333	-0,43
9	315	А		
10	317	А		
11	322	А		
12	348	А		

Окончание табл.4

1	2	3	4	5
13	375	А		
14	381	А		
15	397	Б	0,625	0,32
16	399	А		
17	431	Б	0,708	0,55
18	457	Б	0,750	0,67
19	462	Б	0,792	0,81
20	478	Б	0,833	0,97
21	538	А		
22	541	Б	0,917	1,39
23	548	Б	0,958	1,73

$$\chi=4,19$$

Ошибка первого рода в данной задаче заключается в том, что правильная гипотеза об отсутствии систематического расхождения между результатами опробования шурфов и скважин будет отвергнута, а это не позволит снизить затраты на дальнейшую разведку за счет использования более дешевых скважин. Ошибка второго рода, то есть принятие неверной гипотезы H_0 , состоит в признании несущественным (случайным) расхождения между данными по шурфам и скважинам, в то время как на самом деле оно носит систематический характер. Ошибка второго рода в данной ситуации может привести к неправильной геолого-экономической оценке месторождения, что нанесет гораздо больший экономический ущерб по сравнению с дополнительными затратами за счет ошибки первого рода. Поэтому для уменьшения риска ошибки второго рода уровень значимости при проверке гипотезы H_0 целесообразно принять не слишком высоким, например 0,1.

Порядок выполнения работы

В столбец В введите значения порядковых номеров скважины А: от 1 до 13 (ячейки В1:В13), в столбец С – содержание Au ($\text{м}^2/\text{м}^3$) скважины А (ячейки С1:С13) – см. табл. III.11, затем значения порядковых номеров скважины Б: от 1 до 10 (ячейки В14:В23), в столбец С – содержание Au ($\text{м}^2/\text{м}^3$) скважины Б (ячейки С14:С23) – см. табл. III.11. В ячейки А1:А13 введите символ А, в ячейки А14:А23 – символ Б. Выделите диапазон А1:С23 и через пункт меню **Данные/Сортировка...** в диалоговом окне **Сортировка диапазона** укажите **Сортировать по**

«Столбец С» – по возрастанию. Нажмите кнопку **ОК**. Столбец D заполните через пункт меню **Правка/Заполнить/Прогрессия...** значениями от 1 до 23. В ячейке B24 запишите формулу =СЧЕТ(B1:B23). В ячейке B24 появится значение 23. В ячейку E1 запишите логическую функцию =ЕСЛИ(A1="Б";D1/(\$B\$24+1);0). Скопируйте ее во все ячейки диапазона E1:E23. Для скважины B в столбце E получите аргумента $i/(n+1)$. В ячейку G15 введите формулу =1–E15. Скопируйте ее в ячейки G17:G20, G22:G23. Это вспомогательные значения для определения значений функции $\psi(i/(n+1))$ для положительных Z. По [1] для скважины B найдите значения функции $\psi(i/(n+1))$ при этих значениях аргумента $i/(n+1)$ [1]. Введите их в столбец F. В ячейке F24 подсчитайте сумму записанных значений функции $\psi(i/(n+1))$ – число 4,19.

Может быть предложен другой порядок расчета. В ячейку F3 ввести формулу =НОРМСТОБР(E3). Эту формулу следует скопировать в ячейки F6, F8, F15, F17–F20, F22–F23 для определения значений функции $\psi(i/(n+1))$.

Допустимое значение X-критерия Ван-дер-Вардена для уровня значимости 0,1, числа наблюдений 23 и разности между объемами сравниваемых выборок 3, равно 3,12 [1,2].

Таким образом, гипотеза о равенстве средних значений содержания золота по скважинам и шурфам отвергается. До выяснения причин систематического занижения содержания золота по скважинам использовать их для разведки данного месторождения не рекомендуется.

Непараметрический критерий Вилкоксона (W) также основан на процедуре ранжирования и представляет собой сумму рангов R_i членов меньшей выборки в общем ранжированном ряду из обеих выборок:

$$W = \sum_{i=1}^{n_1} R_i, \quad n_1 \leq n_2. \quad (3)$$

Если гипотеза о равенстве средних по совокупностям A и B верна, то есть $H_0: \bar{x}_1 = \bar{x}_2$, математическое ожидание статистики Вилкоксона (MW) и величины возможных отклонений от нее выборочных оценок (W) зависят только от объемов выборок n_1 и n_2 .

Для случаев, когда n_1 и $n_2 < 25$ значения удвоенного математического ожидания критерия Вилкоксона ($2MW$) и его нижнего критического значения W_1 для заданного уровня значимости α приведены в специальных таблицах [3]. Верхнее критическое значение критерия W_2 определяется из уравнения $W_2 = 2MW - W_1$. Уровень значимости для W_1 в этих таблицах дан для альтернативы $H_1: \bar{x}_1 \neq \bar{x}_2, \bar{x}_1 < \bar{x}_2$. Поэтому

при альтернативе $H_1: \bar{x}_1 \neq \bar{x}_2, \bar{x}_1 < \bar{x}_2$ или $\bar{x}_1 > \bar{x}_2$ уровень значимости для нахождения W_1 необходимо уменьшить в два раза.

Если n_1 или $n_2 > 25$, критические значения критерия Вилкоксона можно определить по следующим приближенным формулам:

$$W_1 \cong 0,5[n_1(n_1 + n_2 + 1) - 1] - Z_{1-\alpha/2} \sqrt{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)}, \quad (4)$$

$$W_2 \cong n_1(n_1 + n_2 + 1) - W_1, \quad (5)$$

где $Z_{1-\alpha/2}$ – значения функции нормального распределения с параметрами 0,1.

При наличии в объединенной выборке совпадающих значений им дается одинаковый средний ранг, равный среднему арифметическому из всех рангов, приходящихся на данную группу повторяющихся значений, а формула принимает следующий вид:

$$W_1 \cong \frac{[n_1(n_1 + n_2 + 1) - 1]}{2} - Z_{1-\alpha/2} \times \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \left[1 - \frac{\sum_{i=1}^k t_i^3 - t_i}{(n_1 + n_2 + 1)(n_1 + n_2)(n_1 + n_2 - 1)} \right]^2}, \quad (6)$$

где k – число групп из повторяющихся значений, принадлежащих разным выборкам; t_i – число совпавших значений в группе с номером i ($i = 1, 2, 3, \dots, k$).

Группы повторяющихся значений, состоящие полностью из значений выборки A или B , можно не учитывать при введении поправки.

Лабораторная работа № 6

Корреляционный анализ и регрессионный анализ

Цель работы: определить наличие корреляционной связи между переменными, построить регрессионное уравнение, описывающее связь между переменными, используя процедуру *Регрессия пакета Анализ данных*.

Исходные данные: Определить наличие корреляционной связи между золотом и свинцом в рудах на неизученном участке месторождения по выборке, представленной в таблице 2.

Таблица 5

Содержание свинца и золота в рудах полиметаллического месторождения

№ проб	Pb	Au	№ проб	Pb	Au	№ проб	Pb	Au
1	2,05	3,76	19	1,21	0,61	37	5,16	0,87
2	5,03	2,09	20	2,92	0,40	38	0,37	1,15
3	0,80	1,98	21	0,74	0,27	39	0,44	0,91
4	0,31	0,20	22	1,53	2,57	40	2,21	4,25
5	0,77	3,10	23	3,70	0,90	41	4,67	2,03
6	4,01	1,67	24	2,71	1,69	42	1,44	4,31
7	1,19	2,59	25	1,90	4,32	43	3,13	0,25
8	1,26	1,70	26	1,51	2,30	44	1,35	0,39
9	0,68	0,23	27	0,21	1,22	45	0,81	1,35
10	0,91	1,21	28	4,81	1,05	46	1,32	3,51
11	4,33	0,91	29	1,38	2,09	47	0,99	1,62
12	2,38	1,68	30	3,96	2,54	48	2,41	3,98
13	0,98	2,44	31	1,96	1,58	49	1,03	0,35
14	0,42	0,50	32	0,52	0,82	50	1,55	2,80
15	1,71	1,21	33	2,95	0,20	51	3,39	0,41
16	3,51	1,15	34	1,10	1,44	52	1,23	1,58
17	1,11	2,30	35	0,93	3,15	53	1,48	4,22
18	2,10	3,48	36	1,78	1,21	54	4,03	1,19

Порядок выполнения работы

Откройте новый лист. Введите в ячейку A1 название *Pb*. Затем в ячейки A2:A55 – соответствующие значения концентраций в пробах. В ячейки B1:B55 введите название *Au* и значения для золота. Затем вычисляется значение коэффициента корреляции между выборками. Для этого табличный курсор установите в свободную ячейку (A56). На панели инструментов нажмите кнопку **Вставка функции (fx)**. В появившемся диалоговом окне **Мастер функций** выберите категорию **Статистические** и функцию **КОРРЕЛ**, после чего нажмите кнопку **ОК**. Появившееся диалоговое окно **КОРРЕЛ** за серое поле мышью отодвиньте вправо на 1–2 см от данных (при нажатой левой клавише).

Указателем мыши введите диапазон данных *Pb* в поле **Массив 1** (A2:A55). В поле **Массив 2** введите диапазон данных *Au* (B2:B55). Нажмите кнопку **ОК**. В ячейке A56 появится значение коэффициента корреляции – число –0,04918. Значение коэффициента корреляции близко к нулю. Значит, можно говорить о том, что наличие корреляционной связи между золотом и свинцом в рудах на неизученном участке месторождения отсутствует.

При отсутствии значимой корреляционной связи между золотом и свинцом, проверить гипотезу о том, что такая связь может существовать только для бедных руд. Для этого из выборочных данных таблицы необходимо убрать пробы со значениями свинца более 1,5 %. Это можно проделать, используя функцию =ЕСЛИ(). В ячейку C2 запишите формулу =ЕСЛИ(A2<=1,5;A2;ЛОЖЬ), а в ячейку D2 – формулу =ЕСЛИ(A2<=1,5;B2;ЛОЖЬ). Скопируйте эти формулы методом автозаполнения в диапазоны C3:C55 и D3:D55 соответственно. В ячейку C56 поместите функцию =КОРРЕЛ(C2:C55;D2:D55). Значение коэффициента корреляции 0,54359. Значит, можно говорить о том, что наличие корреляционной связи между золотом и свинцом в рудах на неизученном участке месторождения существует только для бедных руд (со значениями свинца не более 1,5 %).

Необходимо построить регрессионное уравнение, описывающее связь между золотом и свинцом в рудах на неизученном участке месторождения по выборке из таблицы 2, используя процедуру **Регрессия** пакета **Анализ данных**. Для этого:

1. Откройте рабочую книгу с результатами предыдущего ПРИМЕРА V.2.

2. В пункте меню **Сервис** выберите строку **Анализ данных** и далее укажите курсором мыши на строку **Регрессия**.

3. В появившемся диалоговом окне задайте **Входной интервал Y**. Для этого наведите указатель мыши на верхнюю ячейку столбца зависимых данных (G1), нажмите левую кнопку мыши и, не отпуская ее, протяните указатель мыши к нижней ячейке (G28), затем отпустите левую кнопку мыши. (Обратите внимание, что зависимые данные – это те данные, которые предполагается вычислять).

4. Так же укажите **Входной интервал X**, то есть введите ссылку на диапазон независимых данных F1:F28. (Независимые данные – это те данные, которые будут измеряться или наблюдаться).

5. Установите флажок в поле **Метки в первой строке**. Установите флажок в поле **График подбора**.

6. Далее укажите выходной диапазон. Для этого поставьте переключатель в положение **Выходной интервал** (наведите указатель мыши и щелкните левой кнопкой), затем наведите указатель мыши на правое поле ввода **Выходной интервал** и, щелкнув левой кнопкой мыши, указатель мыши наведите на левую верхнюю ячейку выходного диапазона (К3). Щелкните левой кнопкой мыши. Нажмите кнопку **ОК**.

Результаты анализа. В выходном диапазоне появятся результаты и графики подбора и остатков.

Интерпретация результатов. В таблице **Дисперсионный анализ** оценивается общее качество полученной модели ее достоверность по уровню значимости критерия Фишера – p , который должен быть меньше, чем 0,05 (строка **Регрессия**, столбец **Значимость F**, в примере 0,00338372, то есть $p = 0,00338372$ – модель значима, и степень точности описания моделью процесса – **R-квадрат** (вторая строка сверху в таблице **Регрессионная статистика**, в примере **R-квадрат** = 0,2954899. Поскольку **R-квадрат** < 0,95, можно говорить о невысокой точности аппроксимации – линейная модель не очень хорошо описывает явление.

Далее необходимо определить значения коэффициентов модели. Они определяются из таблицы в столбце **Коэффициенты** – в строке **Y-пересечение** приводится свободный член, в строках соответствующих переменных приводятся значения коэффициентов при этих переменных. В столбце **p-значение** приводится достоверность отличия соответствующих коэффициентов от нуля. В случаях, когда $p > 0,05$, коэффициент может считаться нулевым. Это означает, что соответствующая независимая переменная практически не влияет на зависимую переменную и коэффициент может быть убран из уравнения.

Отсюда выражение для определения концентрации золота в пробе по концентрации свинца будет иметь следующий вид: $0,0440856 + 1,7633983 * \text{концентрация Pb}$. Или, учитывая, что коэффициент для Y-пересечения незначим, – вид: $1,7633983 * \text{концентрация Pb}$.

Воспользовавшись полученным уравнением, можно рассчитать ожидаемую концентрацию золота в пробе по концентрации свинца. Например, для расчета концентрации золота при концентрации свинца равной 0,5 необходимо поставить табличный курсор в любую свободную ячейку (G35); ввести с клавиатуры знак =, щелкнуть указателем мыши по ячейке L19, ввести с клавиатуры знак +, щелкнуть по ячейке L20, ввести с клавиатуры знак * и число 0,5. В результате в ячейке G35 будет получена ожидаемая концентрация золота – 0,9257.

Результат приблизительно можно определить непосредственно по графику подбора. Наведите указатель мыши на график подбора и по появившейся надписи (см. рис. 5) определите приблизительно значение ожидаемой концентрации золота – 0,9610527.

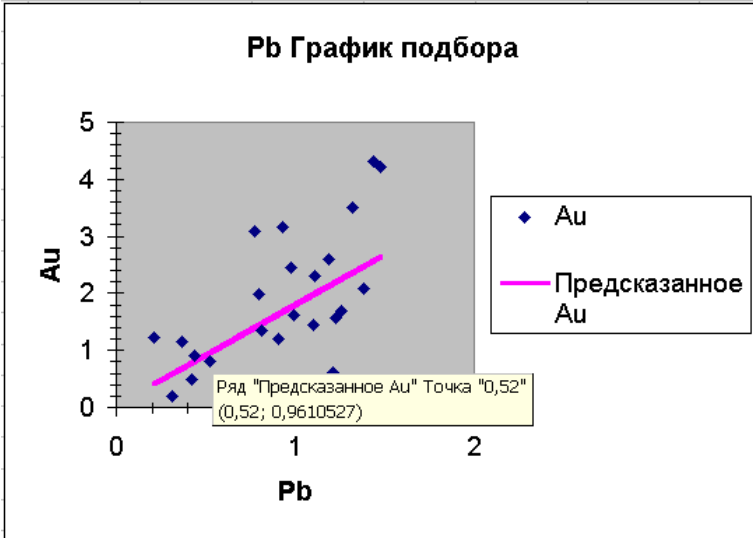


Рис. 5. Определение приближенного результата по графику подбора

Библиографический список

1. Мартыянова, А.Е. Математические методы моделирования в геологии. Часть I: Учебное пособие для студентов направления 650100 «Прикладная геология». – Астрахань: АГТУ, 2008. – 218 с.
2. Мартыянова, А.Е. Математические методы моделирования в геологии. Часть II: Учебное пособие для студентов направления 650100 «Прикладная геология». – Астрахань: АГТУ, 2008. – 190 с.
3. Решение математических задач средствами Excel: практикум / В.Я. Гельман. – СПб.: Питер, 2003. – 240 с.

Оглавление

Лабораторная работа № 1. Понятие математической статистики.....	3
Лабораторная работа № 2. Проверка статистических гипотез	6
Лабораторная работа № 3. Принятие статистических решений	7
Лабораторная работа № 4. Понятие математической статистики (анализ одной выборки, анализ двух выборок).....	10
Лабораторная работа № 5. Использование инструмента Пакет анализа для выявления различий между выборками.....	13
Лабораторная работа № 6. Корреляционный анализ и регрессионный анализ.....	17
Библиографический список.....	21

Учебное издание

Информационные технологии в строительной индустрии

Методические указания к выполнению лабораторных работ
для студентов направления 08.04.01– Строительство,
профиля «Дорожно-строительное материаловедение»

Составители: **Гнездилова** Светлана Александровна
Погромский Алексей Сергеевич

Подписано в печать Формат 60x84/16. Усл.печ.л. 1,2. Уч-изд.л. 1,3.

Тираж экз. Заказ Цена

Отпечатано в Белгородском государственном технологическом университете
им. В.Г. Шухова

308012, г. Белгород, ул. Костюкова, 46