

Энгельсский технологический институт (филиал) федерального государственного
бюджетного образовательного учреждения
высшего образования
«Саратовский государственный технический университет имени Гагарина Ю.А.»

Кафедра «Естественные и математические науки»

Оценочные материалы по дисциплине / практике

Б.1.3.1.2 «Анализ данных»

направления подготовки

09.03.01 «Информатика и вычислительная техника»

профиль

«Программное обеспечение средств вычислительной техники и автоматизированных систем»

Энгельс 2023

1. Перечень компетенций и уровни их сформированности по дисциплинам (модулям), практикам в процессе освоения ОПОП ВО

В процессе освоения образовательной программы у обучающегося в ходе изучения дисциплины «Анализ данных» должны сформироваться компетенции: ПК-2.

Критерии определения сформированности компетенций на различных уровнях их формирования

Индекс компетенции	Содержание компетенции
ПК-2	Способен применять методы проектирования и разрабатывать сопровождающую документацию на ПО

Код и наименование индикатора достижения компетенции	Виды занятий для формирования компетенции	Оценочные средства для оценки уровня сформированности компетенции
ИД-4 ПК-2 Знает структуры и алгоритмы обработки данных и анализирует их для проектирования ПО	Лекции, практические занятия, самостоятельная работа	Практические задания Вопросы к зачету Тестовые задания по дисциплине

Уровни освоения компетенции

Уровень освоения компетенции	Критерии оценивания ¹
Продвинутый (отлично)	Знает: технологию разработки алгоритмов и программ, методы отладки и решения задач на ЭВМ в различных режимах на продвинутом уровне. Умеет: ставить задачу и разрабатывать алгоритм ее решения, использовать прикладные системы программирования, разрабатывать основные программные документы на продвинутом уровне. Владеет: языками процедурного и объектно-ориентированного программирования, навыками разработки и отладки программ не менее, чем на одном из процедурных языков программирования высокого уровня на продвинутом уровне.
Повышенный (хорошо)	Знает: технологию разработки алгоритмов и программ, методы отладки и решения задач на ЭВМ в различных режимах на повышенном уровне. Умеет: ставить задачу и разрабатывать алгоритм ее решения, использовать прикладные системы программирования, разрабатывать основные программные документы на повышенном уровне. Владеет: языками процедурного и объектно-ориентированного программирования, навыками разработки и отладки программ не менее, чем на одном из процедурных языков программирования высокого уровня на повышенном уровне.

Пороговый (базовый) (удовлетворительно)	Знает: технологию разработки алгоритмов и программ, методы отладки и решения задач на ЭВМ в различных режимах на базовом уровне. Умеет: ставить задачу и разрабатывать алгоритм ее решения, использовать прикладные системы программирования, разрабатывать основные программные документы на базовом уровне. Владеет: языками процедурного и объектно-ориентированного программирования, навыками разработки и отладки программ не менее, чем на одном из процедурных языков программирования высокого уровня на базовом уровне.
--	---

2. Методические, оценочные материалы и средства, определяющие процедуры оценивания сформированности компетенций (элементов компетенций) в процессе освоения ОПОП ВО

2.1 Оценочные средства для текущего контроля

Задачи по темам

Тема 1. Методы прикладной статистики

1. Обработать статистические данные с помощью встроенных функций Excel; изучить возможности Пакета анализа и его некоторые инструменты.
2. Ввод данных в PSPP. Вывод результатов
3. Разработать схему кодирования и матрицу данных вопросов анкеты (по вариантам) для обработки в SPSS в файле Excel (см. пример файла excel)
4. Создать в СУБД БД и форму для ввода данных ответов на анкету. Ввести через форму 50 ответов на анкету. Импортировать данные из БД в файл Excel. Структура данных должна соответствовать матрице данных. Загрузить файл данных в SPSS
5. Рассчитать стандартную ошибку среднего, если выборочное среднее равняется 30, дисперсия 25, при $N = 100$
6. Считается, что значение IQ (уровень интеллекта) у людей имеет нормальное распределение со средним значением равным 100 и стандартным отклонением равным 15 ($M = 100, sd = 15$).
Какой приблизительно процент людей обладает $IQ > 105$?
7. Считается, что значение IQ (уровень интеллекта) у людей имеет нормальное распределение со средним значением равным 100 и стандартным отклонением равным 15 ($M = 100, sd = 15$).
Какой приблизительно процент людей обладает IQ на промежутке от 70 до 112

Тема 2. Использование статистических методов в среде PSPP

1. С помощью критерия Стьюдента в PSPP сделать вывод о различии средних двух выборок.
2. На основе файла данных определить отсутствие или наличие линейной корреляционной взаимосвязи между 2 переменными. Построить точечную диаграмму и выдвинуть гипотезу о характере связи между рассматриваемыми переменными. Рассчитать коэффициент корреляции. Проверить значимость коэффициента корреляции. Сделать выводы.
3. Для исследования на основе простой регрессии воспользоваться файлом данных exam.sav. Он содержит 10-балльную оценку нервной возбудимости

(тревожности) 36 студентов и количество решенных ими зачетных тестовых задний (из 20 возможных). Гипотеза о линейности отношения этих двух переменных говорит о том, что чем выше нервная возбудимость студента, тем выше его результативность (например, потому, что спокойных студентов меньше волнуют их знания, а тревожные студенты проводят больше времени за подготовкой к зачету). В качестве зависимой переменной (критерия) выступает переменная тест, а в качестве независимой переменной (предиктора) — переменная трев.

- Используя исходные данные, построить модель парной линейной регрессии для изучения зависимости x от y . Найти коэффициент корреляции и коэффициент детерминации. Сделать прогноз показателя при конкретном значении. Найти коэффициент эластичности, т.е. на какую величину в среднем изменится показатель Y при изменении X на единицу. При описании результатов необходимо предоставить диаграмму рассеяния, уравнение регрессии, оценку модели.
- Построить оптимальную регрессионную модель для целевой переменной, зависящей от нескольких показателей.

2.2 Оценочные средства для промежуточного контроля⁴

Вопросы к зачету

- Нормальное распределение. Зависимость и независимость выборок. Вероятность ошибки.
- Структурирование, ввод и проверка данных.
- Описательный (дескриптивный) анализ. Аналитическая статистика.
- Частотный анализ. Частотные таблицы.
- Медиана для концентрированных данных. Форматы частотных таблиц. Графическое представление.
- Отбор данных. Выбор наблюдений. Операторы отношения. Логические операторы.
- Извлечение случайной выборки. Сортировка наблюдений. Разделение наблюдений на группы.
- Модификация данных. Вычисление новых переменных. Формулировка численных выражений. Функции.
- Подсчет частоты появлений определенных событий. Перекодирование значений. Ручное перекодирование. Автоматическое перекодирование.
- Вычисление новых переменных в соответствии с определенными условиями. Формулировка условий.
- Агрегирование данных. Ранговые преобразования. Типы рангов.
- Весы случаев. Коррекция при отсутствии репрезентативности. Анализ концентрированных данных.
- Статистические характеристики. Описательная статистика. Сводка наблюдений.
- Исследование данных. Обнаружение ошибок ввода. Проверка закона распределения.
- Вычисление характеристик. Анализ без группирующей переменной. Анализ для групп наблюдений.
- Таблицы сопряженности. Создание таблиц сопряженности. Графическое представление таблиц сопряженности.
- Статистические критерии для таблиц сопряженности. Тест хи-квадрат.
- Коэффициенты корреляции. Меры связанности для переменных с номинальной шкалой.
- Меры связанности для переменной с порядковой шкалой. Другие меры связанности.
- Анализ множественных ответов. Дихотомный метод.
- Определение наборов. Частотные таблицы для дихотомических наборов.

22. Категориальный метод. Определение наборов. Частотные таблицы для категориальных наборов.
23. Сравнение средних. Сравнение двух независимых выборок. Сравнение двух зависимых выборок.
24. Сравнение более двух независимых выборок. Разложение на составляющие тренда. Априорные контрасты. Апостериорные тесты.
25. Сравнение более чем двух зависимых выборок. t-тест одной выборки.
26. Непараметрические тесты. Сравнение двух независимых выборок.
27. U-тест по методу Манна и Уитни.
28. Тест Мозеса.
29. Тест Колмогорова-Смирнова.
30. Тест Уальда-Вольфовица.
31. Сравнение двух зависимых выборок. Тест Уилкоксона.
32. Знаковый тест.
33. Тест хи-квадрат по методу МакНемара.
34. Сравнение более чем двух независимых выборок.
35. H-тест по методу Крускала и Уоллиса.
36. Медианный тест.
37. Сравнение более чем двух зависимых выборок.
38. Тест Фридмана. W Кендала. Q Кохрана.
39. Тест Колмогорова-Смирнова для проверки формы распределения.
40. Отдельный тест по критерию хи-квадрат.
41. Биномиальный тест. Анализ последовательностей.
42. Корреляции. Коэффициент корреляции Пирсона.
43. Ранговые коэффициенты корреляции по Спирману и Кендалу.
44. Частная корреляция. Мера корреляции и мера сходства. Внутриклассовый коэффициент корреляции.
45. Регрессионный анализ. Простая линейная регрессия.
46. Расчет уравнения регрессии. Сохранение новых переменных.
47. Множественная линейная регрессия. Нелинейная регрессия.
48. Бинарная логистическая регрессия.
49. Мультиномиальная логистическая регрессия.
50. Порядковая регрессия. Пробит-анализ.
51. Приближение при помощи кривых. Взвешенное оценивание.
52. Двухступенчатый метод наименьших квадратов.
53. Дисперсионный анализ. Одномерный дисперсионный анализ.
54. Одномерный дисперсионный анализ общий многофакторный.
55. Одномерный дисперсионный анализ по методу Фишера.
56. Одномерный дисперсионный анализ с повторным измерением.
57. Ковариационный анализ.
58. Многомерный дисперсионный анализ. Компоненты дисперсии.
59. Дискриминантный анализ.
60. Факторный анализ. Порядок выполнения факторного анализа.
61. Кластерный анализ. Принцип кластерного анализа.
62. Иерархический кластерный анализ.
63. Иерархический кластерный анализ с двумя переменными.
64. Иерархический кластерный анализ с более чем двумя переменными.
65. Иерархический кластерный анализ с предварительным факторным анализом.
66. Меры расстояния и меры сходства. Переменные, относящиеся к интервальной шкале (метрические переменные).
67. Частоты. Бинарные переменные. Методы объединения.
68. Кластерный анализ при большом количестве наблюдений. Кластерный анализ методом k-средних.
69. Анализ пригодности. Задания типа верно-неверно. Задания со ступенчатыми ответами.
70. Экспортирование входных данных. Программирование. Основные синтаксические правила.

71. Применение сценариев. Автоматические сценарии. редактор сценариев.
72. Подготовка данных. Кодирование и кодировочная таблица. Матрица данных.
73. Предварительные условия для проведения статистического теста. Типы статистических шкал. Определение переменных.

Оценивание результатов обучения в форме уровня сформированности элементов компетенций проводится путем контроля во время промежуточной аттестации в форме зачета:

а) оценка «зачтено» – компетенция(и) или ее часть(и) сформированы на базовом уровне;

б) оценка «не зачтено» – компетенция(и) или ее часть(и) не сформированы.

Критерии, на основе которых выставляются оценки при проведении текущего контроля и промежуточной аттестации приведены в табл. 1.

Оценки «Не зачтено» ставятся также в случаях, если обучающийся не приступал к выполнению задания, а также при обнаружении следующих нарушений:

- списывание;
- плагиат;
- фальсификация данных и результатов работы.

Таблица 1 – Критерии выставления оценок при проведении текущего контроля и промежуточной аттестации

Шкала оценки	Оценка	Критерий выставления оценки
Двухбалльная шкала	Зачтено	Обучающийся ответил на теоретические вопросы. Показал знания в рамках учебного материала. Выполнил практические задания. Показал удовлетворительные умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала
	Не зачтено	Обучающиеся при ответе на теоретические вопросы и при выполнении практических заданий продемонстрировал недостаточный уровень знаний и умений при решении задач в рамках учебного материала. При ответах на дополнительные вопросы было допущено множество неправильных ответов

2.3. Итоговая диагностическая работа по дисциплине

ЗАДАНИЯ ДЛЯ ДИАГНОСТИЧЕСКОЙ РАБОТЫ ПО ДИСЦИПЛИНЕ «Анализ данных»

Номер задания	Правильный ответ *	Содержание вопроса	Компетенция	Код и наименование индикатора достижения компетенции
1.	a	Укажите, в какой из выборок наибольшее стандартное отклонение? a. 100 300 250 400 230 280 320 112 b. 15 10 13 7 28 31 20 32 c. 1 3 2 4 5 7 1 8	ПК-2	ИД-4 ПК-2 Знает структуры и алгоритмы обработки данных и анализирует их для проектирования ПО
2.	a	Как соотносятся стандартные (среднеквадратические) отклонения двух выборок: (1) 2 4 5 8 9 10 14 16 (2) 6 12 15 24 27 30 42 48 a) $SD_2 = SD_1 * 3$ b) $SD_2 = SD_1 * 9$ c) $SD_2 = SD_1$ d) $SD_2 = SD_1 / 3$	ПК-2	ИД-4 ПК-2
3.	a	Если увеличить размер выборки, то сильные отклонения выборочных средних от истинного среднего будут возникать a. Реже, распределение выборочных средних станет более узким b. Чаще, распределение выборочных средних станет более широким c. Это не повлияет на характер распределения выборочных средних	ПК-2	ИД-4 ПК-2

4.	a	В центре 95% доверительного интервала, рассчитанного по выборочным значениям, находится: a. Выборочное среднее значение b. Среднее значение генеральной совокупности c. Значение стандартной ошибки среднего	ПК-2	ИД-4 ПК-2
5.	a	Если в определенной ситуации весьма рискованно отклонить нулевую гипотезу, когда она на самом деле верна, то лучше использовать показатель α равный a. 0,001 b. 0,05 c. 0,1 d. 0,5	ПК-2	ИД-4 ПК-2
6.	Качественные и количественные	Назовите два основных типа переменных в статистике	ПК-2	ИД-4 ПК-2
7.	нет	Может ли показатель стандартного отклонения принимать отрицательные значения?	ПК-2	ИД-4 ПК-2
8.	0,2	Чему равна стандартная ошибка среднего, если выборочное среднее равняется 10, дисперсия 4, при $N = 100$	ПК-2	ИД-4 ПК-2
9.	Станет более узким	Как изменится 95% доверительный интервал, если увеличить объем выборки в два раза (при условии, что показатель стандартного отклонения остался неизменным)	ПК-2	ИД-4 ПК-2
10.	совершили ошибку второго рода α	Предположим, нулевой гипотезой вашего исследования являлось предположение, что среднее в генеральной совокупности равняется 100. Вы получили $p = 0,12$ и не смогли отклонить нулевую гипотезу. Однако позже выяснилось, что среднее в генеральной совокупности равняется 114. Как можно описать результаты данного исследования? Совершили ли вы ошибки первого и второго рода?	ПК-2	ИД-4 ПК-2
11.	Коэффициент корреляции равен 0,5 или -0,5	Если при исследовании взаимосвязи двух переменных коэффициент детерминации оказался равен 0,25 то чему равен коэффициент корреляции	ПК-2	ИД-4 ПК-2
12.	Нормальное распределение переменных, линейная взаимосвязь двух переменных	Укажите основные требования к данным при использовании коэффициента корреляции Пирсона	ПК-2	ИД-4 ПК-2