

Государственное образовательное учреждение
"Приднестровский государственный университет им. Т.Г. Шевченко"

Инженерно-технический институт

Кафедра информационных технологий
и автоматизированного управления производственными процессами

УТВЕРЖДАЮ

Заведующий кафедрой ИТиАУПП

 Ю.А. Столяренко

«29» августа 2022 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ

Направление подготовки

2.09.03.02 Информационные системы и технологии

Профиль подготовки

Безопасность информационных систем

Квалификация (степень)

выпускника:

бакалавр

Форма обучения:

очная, заочная

Год набора:

2021 г.

Разработал:

к.т.н., доцент кафедры ИТиАУПП,

 /В.С. Попукайло

«29» августа 2022 г.

Тирасполь, 2022

Паспорт фонда оценочных средств по учебной дисциплине

1. В результате изучения дисциплины «Статистическая обработка данных» у обучающегося должны быть сформированы следующие компетенции:

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения универсальной компетенции
Общепрофессиональные компетенции и индикаторы их достижения		
	ОПК-1. Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ИД-1 _{ОПК-1} Знать основы высшей математики, физики, экологии, инженерной графики, информатики и программирования ИД-2 _{ОПК-1} Уметь решать стандартные профессиональные задачи с применением естественнонаучных и общетехнических знаний, методов математического анализа и моделирования ИД-3 _{ОПК-1} Владеть методами теоретического и экспериментального исследования объектов профессиональной деятельности
	ОПК-8. Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем	ИД-1 _{ОПК-8} Знать: методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем. ИД-2 _{ОПК-8} Уметь: применять на практике математические модели, методы и средства проектирования и автоматизации систем на практике. ИД-3 _{ОПК-8} Иметь навыки: моделирования и проектирования информационных и автоматизированных систем.

2. Программа оценивания контролируемой компетенции:

Текущая аттестация	Контролируемые модули, разделы (темы) дисциплины их название	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
РУБЕЖНЫЙ КОНТРОЛЬ	Раздел 1 Раздел 2 Раздел 3	ОПК-1, ОПК-8	Лабораторные работы №1-5
РУБЕЖНАЯ АТТЕСТАЦИЯ	Раздел 4 Раздел 5 Раздел 6		Лабораторные работы №6-9 Тестирование
Промежуточная аттестация		Код контролируемой компетенции (или ее части)	Наименование оценочного средства
№1		ОПК-1, ОПК-8	зачет

3. Показатели и критерии оценивания компетенции по этапам формирования, описание шкал оценивания

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
Первый этап	ИД-1 _{ОПК-1} Знать основы высшей математики, физики, экологии, инженерной графики, информатики и программирования	Не знает	Знает основные понятия но не знает способы использования в профессиональной деятельности	Знает основы высшей математики, физики, экологии, инженерной графики, информатики и программирования, но не может применить знания в полной мере в профессиональной деятельности	Знает основы высшей математики, физики, экологии, инженерной графики, информатики и программирования и может использовать в профессиональной деятельности
Второй этап	ИД-2 _{ОПК-1} Уметь решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования	Не умеет	Уметь решать некоторые стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования	Умеет решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования, но не в полной мере	Умеет решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
Третий этап	ИД-3 _{ОПК-1} Владеть методами теоретического и экспериментального исследования объектов профессиональной деятельности	Не владеет	Владеет методами теоретического и экспериментального исследования объектов профессиональной деятельности, но не владеет ими в междисциплинарном контексте	Владеет методами теоретического и экспериментального исследования объектов профессиональной деятельности, но ошибается в обработке их результатов	Владеет методами теоретического и экспериментального исследования объектов профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте
Первый этап	ИД-1 _{ОПК-8} Знать: методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем	Не знает	Знает основные методы но не знает способы использования в профессиональной деятельности	Знает методологию и основные методы математического моделирования, но не может применить знания в полной мере в профессиональной деятельности	Знает методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем и может использовать в профессиональной деятельности
Второй этап	ИД-2 _{ОПК-8} Уметь: применять на практике математические модели, методы и средства проектирования и автоматизации систем.	Не умеет	Уметь применять на практике некоторые стандартные математические модели, методы и средства проектирования и автоматизации систем	Умеет применять на практике стандартные математические модели, методы и средства проектирования и автоматизации систем, но не в полной мере	Умеет применять на практике математические модели, методы и средства проектирования и автоматизации систем, и решать задачи из профессиональной области
Третий этап	ИД-3 _{ОПК-8} Иметь навыки: моделирования и проектирования инфор-	Не владеет	Владеет методами моделирования и проектирования инфор-	Владеет методами моделирования и проектирования	Владеет методами моделирования и проектирования информационных

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
	мационных и автоматизированных систем.		мационных и автоматизированных систем, но не владеет ими в междисциплинарном контексте	информационных и автоматизированных систем, но ошибается в незнакомых предметных областях	и автоматизированных систем, в том числе в новой или незнакомой среде и в междисциплинарном контексте

4. Шкала оценивания

Согласно Положению «О порядке организации аттестации в ИТИ ПГУ им. Т.Г. Шевченко, итоговая оценка представляет собой сумму баллов, полученных студентом по итогу освоения дисциплины (модуля):

Оценка в традиционной шкале	Оценка в 100-балльной шкале	Буквенные эквиваленты оценок в шкале ЗЕ (% успешно аттестованных)
5 (отлично)	88–100	А (отлично) – 88-100 баллов
4 (хорошо)	70–87	В (очень хорошо) – 80-87баллов
		С (хорошо) – 70-79 баллов
3 (удовлетворительно)	50–69	Д (удовлетворительно) – 60-69 баллов
		Е (посредственно) – 50-59 баллов
2 (неудовлетворительно)	0–49	Гх – неудовлетворительно, с возможной пересдачей – 21-49 баллов
		Г – неудовлетворительно, с повторным изучением дисциплины – 0-20 баллов

Расшифровка уровня знаний, соответствующего полученным баллам, дается в таблице, указанной ниже

А	“Отлично” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.
В	“Очень хорошо” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.
С	“Хорошо” - теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.
Д	“Удовлетворительно” - теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.

Е	“Посредственно” - теоретическое содержание курса освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные программой обучения учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.
FX	“Условно неудовлетворительно” - теоретическое содержание курса освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.
F	“Безусловно неудовлетворительно” - теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, все выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над материалом курса не приведет к какому-либо значимому повышению качества выполнения учебных заданий.

5. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций при изучении учебной дисциплины в процессе освоения образовательной программы

5.1 Образец индивидуального задания к лабораторной работе №1

1. Сгенерировать 100 выборок со стандартным нормальным законом распределения ($M=0$, $s=1$) и объемом выборки $N=(N_{\text{№ по варианту}} + 15)$.
2. Рассчитать для каждой выборки точечные оценки среднего, дисперсии и стандартного отклонения.
3. Рассчитать средние значения для оцененных параметров. Сравнить полученные результаты с генеральными значениями параметров.
4. Рассчитать для каждой выборки интервальные значения среднего, дисперсии и стандартного отклонения, приняв за априорное утверждение, что нам не известны генеральные значения данных параметров. Расчет произвести для вероятностей $P=0,90$, $P=0,95$.
5. Сделать выводы о точности точечных и интервальных оценок неизвестных параметров распределений.

5.2. Образец индивидуального задания к лабораторной работе №2

1. Сгенерировать 1000 выборок со стандартным нормальным законом распределения и объемом выборки N , где $N = N_{\text{№ по списку в журнале}} + 20$.
2. Рассчитать для каждой выборки точечные оценки среднего.
3. Для каждой сотой выборки построить графики, позволяющие оценить вид закона распределения: гистограммы, *qq-plot*, *boxplot*, *violinplot*.
4. Построить аналогичные графики для вектора, состоящего из средних значений.
5. Сделать выводы о проделанной работе.

5.3 Образец индивидуального задания к лабораторной работе №3

1. Загрузить файл «Данные для расчетов».
2. Проверить гипотезу о нормальности для каждой выборки.
3. Построить гистограммы выборок, для которых удалось отвергнуть гипотезу о нормальности.
4. Произвести логарифмическую трансформацию и изучить её влияние на распределение данных, построив гистограммы и проверив гипотезу о нормальности.
5. Для наглядности, сгенерировать выборку из экспоненциального распределения и построить гистограммы по сгенерированным данным и после их логарифмирования.
6. Сделать выводы о проделанной работе.

5.4 Образец индивидуального задания к лабораторной работе №4

1. Загрузить данные из файла «Данные для расчетов».
2. Сделать срез из данных, такой, чтобы в исследуемом `data.frame` остались только столбцы с номерами: $[N_{\text{№ по журналу}}; N_{\text{№ по журналу}}+10]$

3. Проверить предпосылки для применения классического однофакторного дисперсионного анализа.
4. Провести параметрический или непараметрический дисперсионный анализ, в зависимости от полученных результатов на предыдущем шаге.
5. При отклонении гипотезы о равенстве всех математических ожиданий провести попарные сравнения изучаемых групп.
6. Визуализировать полученные различия при помощи boxplot + violinplot..
7. Сделать выводы о проделанной работе.

5.5 Образец индивидуального задания к лабораторной работе №5

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Построить график корреляционного поля для каждого фактора.
3. Построить уравнение парной линейной регрессии для каждого фактора
4. Проверить значимость каждого из полученных уравнений регрессии. Показать уравнения регрессии с заданным в варианте доверительным интервалом на графиках.
5. Построить прогнозы по каждому из уравнений парной регрессии для заданных в варианте значений факторов.
6. Построить уравнение множественной линейной регрессии и получить корреляционную матрицу.
7. Построить прогноз по уравнению множественной регрессии для заданных в варианте значений факторов.

5.6 Образец индивидуального задания к лабораторной работе №6

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Построить модель логистической регрессии для заданной целевой переменной.
3. Оцените качество полученной модели. Определите, какой метрикой лучше пользоваться в данном случае.
4. Постройте ROC кривую для полученной модели.
5. Постройте дерево решений для данных своего варианта.
6. Оцените качество полученной модели. Определите, какой метрикой лучше пользоваться в данном случае.
7. Постройте ROC кривую для полученной модели.
8. Сделать выводы о проделанной работе.

5.7 Образец индивидуального задания к лабораторной работе №7

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Используя пакет caret постройте модели: наивный байесовский классификатор, машины опорных векторов, случайный лес и Adaboost.
3. Оцените качество полученных моделей при помощи метрик Accuracy и Карра.
4. Сделать выводы о проделанной работе.

5.8 Образец индивидуального задания к лабораторной работе №8

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Проведите процедуру кластеризации при помощи метода k-means и иерархической кластеризации.
3. Постройте дендрограмму и определите оптимальное количество кластеров.
4. Сделать выводы о проделанной работе.

5.9 Образец индивидуального задания к лабораторной работе №9

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Проведите разведывательный анализ данных, ответьте на вопросы.
3. Постройте графики, указанные в заданиях.
4. Рассчитайте статистические критерии.
5. Оцените величину корреляционной связи между величинами.
6. Постройте математическую модель по данным.
7. Оформите отчёт в виде RMarkdown файла.

5.10 Образец тестового задания №1

Пример вопросов теста №1:

1. Критерий Стьюдента может применяться:

- a. Для проверки равенства средних значений в двух выборках
- b. Для проверки равенства дисперсий в двух выборках
- c. Для проверки значимости коэффициентов в регрессионных моделях
- d. Для проверки значимости регрессионных моделей в целом

2. В каком случае необходимо применять непараметрические критерии:

- a. Если распределение экспериментальных данных согласуется с семейством нормальных распределений
- b. Если распределение экспериментальных данных не согласуется с семейством нормальных распределений
- c. Если объём экспериментальных данных больше 100
- d. Если необходимо проанализировать количественные данные

3. Оценка называется состоятельной, если:

- a. её математическое ожидание равно оцениваемому параметру
- b. она асимптотически приближается к оцениваемому параметру
- c. её дисперсия является наименьшей среди других оценок

4. Основной целью дисперсионного анализа является:

- a. Исследование значимости различий между средними с помощью анализа дисперсий
- b. Исследование значимости различий между дисперсиями с помощью анализа средних
- c. Исследование значимости различий между дисперсиями с помощью анализа распределений
- d. Исследование значимости различий между распределениями с помощью анализа дисперсий

5. Мощностью критерия называется:

- a. Вероятность ошибки первого рода (ложная тревога)
- b. Вероятность ошибки второго рода (пропуск цели)
- c. Единица минус вероятность ошибки первого рода
- d. Единица минус вероятность ошибки второго рода

6. Для сравнения качества моделей с различным количеством предикторов можно использовать:

- a. Коэффициент корреляции
- b. Коэффициент детерминации
- c. Исправленный коэффициент корреляции
- d. Исправленный коэффициент детерминации
- e. Информационный критерий Акаике
- f. Информационный критерий Шварца
- g. Информационный критерий Стьюдента
- h. Критерий Фишера

7. Какой из пороговых значений p -уровня значимости лучше выбрать, если весьма рискованно отклонить нулевую гипотезу, когда она на самом деле верна:

- a. 0,001
- b. 0,5
- c. 0,05

- d. 0,1
- e. 0,01

5.11 Перечень вопросов к зачёту по предмету

1. Распределение Хи-квадрат.
2. Этапы анализа данных.
3. Рассчитайте точечную и интервальную оценки среднего и дисперсии.
4. Распределение Стьюдента.
5. Однофакторный дисперсионный анализ.
6. Рассчитайте величину линейной корреляционной связи между выборками.
7. Нормальное распределение.
8. Определите, являются ли значимыми различия между выборками.
9. Равномерное распределение.
10. Оценка качества классификатора.
11. Рассчитайте описательные статистики для выборки: среднее арифметическое, коэффициент вариации, выборочную дисперсию, стандартное отклонение.
12. Корреляционный анализ. Методы и подходы.
13. Определите, имеют ли распределения выборок равные дисперсии.
14. Регрессионный анализ. Цели и задачи.
15. Проверьте гипотезу о принадлежности генеральной совокупности к нормальному закону распределения по данным выборки.
16. Идея р-уровня значимости.
17. Оценка качества регрессионной модели.
18. Квантили распределения.
19. Графические способы оценки вида закона распределения.
20. Свойства статистических оценок.
21. Дисперсионный анализ. Цели и задачи.
22. Понятие мощности статистических критериев. Виды ошибок.
23. Гипотеза о наличии в данных выбросов.
24. Центральная предельная теорема.
25. Алгоритмы кластеризации.
26. Алгоритмы классификации.
27. Ограничения метода наименьших квадратов.
28. Логистическая регрессия.
29. Задача обработки естественного языка.