

Приложение 2 к РПД
Статистический анализ данных
39.03.01 Социология
Направленность (профиль)
Цифровая и экспертно-аналитическая социология
Форма обучения – очная
Год набора – 2022

**ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

1. Общие сведения

1.	Кафедра	Математики, физики и информационных технологий
2.	Направление подготовки	39.03.01 Социология
3.	Направленность (профиль)	Цифровая и экспертно-аналитическая социология
4.	Дисциплина (модуль)	Б1.В.ДВ.04.02 Статистический анализ данных
5.	Форма обучения	очная
6.	Год набора	2022

2. Перечень компетенций

ПК-3: Способен выполнить описание, объяснение, прогнозирование социальных явлений и процессов на основе результатов социологических исследований и экспертно-аналитической деятельности (самостоятельно или под руководством)
ПК-4: Способен к научно-исследовательской деятельности социальных явлений и процессов в цифровом обществе

3. Критерии и показатели оценивания компетенций на различных этапах их формирования

Этап формирования компетенции (разделы, темы дисциплины)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
1. Параметрическая проверка гипотез	ПК-3 ПК-4	Методы, которыми необходимо пользоваться в той или иной ситуации в зависимости от типа данных и от исследовательской задачи	Подготовить экспериментальные данные для анализа, подобрать способы и методы анализа данных, интерпретировать полученные в результате данных результаты	Методологией теоретических и экспериментальных исследований в области социологии	Контрольная работа, домашнее задание
2. Непараметрическая проверка гипотез		Методологию теоретических и экспериментальных исследований в области социологии	осуществлять предварительную подготовку данных для анализа	Методологией теоретических и экспериментальных исследований в области социологии	Контрольная работа, домашнее задание
3. Дисперсионный анализ		Методы анализа и оценки современных научных достижений	Анализировать результаты научных исследований, определять тематику исследований, представлять результаты своих исследований в современных формах	Методами статистического анализа данных для исследования социальных процессов	Контрольная работа, домашнее задание
4. Анализ зависимостей		Методы работы российских и международных исследовательских коллективов по решению научных и научно-образовательных задач	решать поставленные задачи с помощью методов статистического анализа	Методами статистического анализа данных для исследования социальных процессов	Контрольная работа, домашнее задание
5. Линейный регрессионный анализ		Методы анализа и оценки современных научных достижений	использовать компьютер для статистического анализа данных	Методами статистического анализа данных для исследования социальных процессов	Контрольная работа, домашнее задание

Шкала оценивания в рамках балльно-рейтинговой системы:

«неудовлетворительно» – 60 баллов и менее; «удовлетворительно» – 61-80 баллов; «хорошо» – 81-90 баллов; «отлично» – 91-100 баллов

4. Критерии и шкалы оценивания

1. Контрольная работа

Баллы	Критерии оценивания
8	контрольная работа выполнена полностью, в решении нет математических ошибок (возможна одна неточность, описка, не являющаяся следствием непонимания материала)
6	контрольная работа выполнена полностью, но обоснования шагов решения недостаточны, допущена одна негрубая ошибка или два-три недочета в выкладках или графиках, если эти виды работы не являлись специальным объектом проверки
4	студент допустил более одной грубой ошибки или более двух-трех недочетов в выкладках и графиках, но студент владеет обязательными умениями по проверяемой теме.
0	студент показал полное отсутствие обязательных знаний и умений по проверяемой теме

Примечание:

К *грубым* ошибкам относятся незнание студентом формул, правил, основных свойств, теорем и неумение их применять, незнание приемов решения задач, а также вычислительные ошибки, если они не являются опiskой.

К *негрубым* ошибкам относятся вычислительные ошибки, если они являются опiskой, потеря решения уравнения или сохранение в ответе постороннего корня.

К *недочетам* относятся нерациональное решение, описки, недостаточность или отсутствие пояснений, обоснований в решении задания.

2. Индивидуальное домашнее задание (ИДЗ)

Баллы	Характеристика индивидуального домашнего задания
4	Уровень расчетно-графической работы отвечает всем требованиям, предъявляемым к выполнению ИДЗ, теоретическое содержание раздела дисциплины «Статистический анализ данных» освоено полностью, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения задания ИДЗ выполнены без замечаний.
3	Уровень расчетно-графической работы отвечает всем требованиям, предъявляемым к выполнению ИДЗ, теоретическое содержание раздела дисциплины «Статистический анализ данных» освоено полностью, при этом некоторые практические навыки работы с освоенным материалом сформированы недостаточно, но все предусмотренные программой обучения задания ИДЗ выполнены, некоторые из них содержат негрубые ошибки.
2	Уровень расчетно-графической работы не отвечает большинству требований, предъявляемым к выполнению ИДЗ, теоретическое содержание раздела дисциплины «Статистический анализ данных» освоено частично, некоторые практические навыки работы не сформированы, отдельные предусмотренные программой обучения задания ИДЗ выполнены с грубыми ошибками.
0	Уровень выполнения ИДЗ показывает, что теоретическое содержание раздела дисциплины «Статистический анализ данных» не освоено, необходимые практические навыки работы не сформированы, все выполненные задания ИДЗ содержат грубые ошибки, дополнительная самостоятельная работа над материалом не приведет к какому-либо значимому повышению качества выполнения заданий ИДЗ.

Требования, предъявляемые к выполнению ИДЗ:

– ИДЗ должно базироваться на знаниях теоретических и методических вопросах дисциплины «Статистический анализ данных». Работа должна содержать элементы творчества, новизны, направленные на эффективное решение заданий ИДЗ;

– ИДЗ должно отразить глубину теоретической подготовки студента, понимание контролируемого учебного материала по дисциплине «Статистический анализ данных»: умение связывать теоретические положения с их практическим применением, способность самостоятельно формировать и обосновывать собственные выводы, логически и грамотно излагать свои мысли;

– в ИДЗ не допускается переписывание учебников, учебных пособий и других источников;

– Студент – автор ИДЗ полностью отвечает за предложенные решения заданий и правильность всех данных, приведенных в ИДЗ;

– ИДЗ должно быть сдано в назначенный руководителем срок.

5. Типовые контрольные задания и методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

5.1. Типовая контрольная работа

Задача 1.

В результате выборочного обследования стажа работы профессорско-преподавательского состава получены следующие данные:

Стаж работы (лет)	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24	24 - 28	28 - 32
Число преподавателей	3	8	25	40	46	31	6	2

Выясните, является ли распределение стажа работы нормальным при уровне значимости $\alpha = 0,01$.

Решение.

Перейдем от заданного интервального распределения к распределению равноотстоящих вариантов и вычислим выборочную среднюю и выборочное среднее квадратическое отклонение.

$$n = 161$$

$$\bar{x} = \frac{1}{161} (2 \cdot 3 + 6 \cdot 8 + 10 \cdot 25 + 14 \cdot 40 + 18 \cdot 46 + 22 \cdot 31 + 26 \cdot 6 + 30 \cdot 2) \approx 16$$

$$\sigma^* = \sqrt{D_x^*} \approx 5,43$$

$$n_i' = \frac{n \cdot h}{\sigma^*} \varphi(u_i) = 118,6 \varphi(u_i)$$

Составим расчетную таблицу, заполняя ее последовательно по столбцам слева направо

i	x_i	$u_i = \frac{x_i - \bar{x}}{\sigma^*}$	$\varphi(u_i)$	$n_i' = 118,6 \varphi(u_i)$	n_i	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	2	-2,58	0,014	1,66	3	1,8	1,08
2	6	-1,84	0,073	8,66	8	0,44	0,05
3	10	-1,1	0,218	25,85	25	0,72	0,033
4	14	-0,37	0,373	44,24	40	17,98	0,41
5	18	0,37	0,373	44,24	46	3,1	0,07
6	22	1,1	0,218	25,85	31	26,52	1,03
7	26	1,84	0,073	8,66	6	7,08	0,82
8	30	2,58	0,014	1,66	2	0,12	0,07
Σ					161		$\chi_{\text{эмп}}^2 = 3,56$

По таблице критических точек распределения по уровню значимости $\alpha = 0,01$ и числу степеней свободы $k=s-3=8-3=5$ находим критическую точку правосторонней критической области

$$\chi_{\text{кр}}^2(0,01;5) = 15,1$$

Так как $\chi_{\text{эмп}}^2 = 3,56 < 15,1 = \chi_{\text{кр}}^2$, то гипотезу о нормальном распределении генеральной совокупности принимаем.

Задача 2.

В механическом цехе анализируется структура себестоимости продукции и доля покупных комплектующих. Было отмечено, что стоимость комплектующих зависит от времени их поставки. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Провести регрессионный анализ данных о поставках:

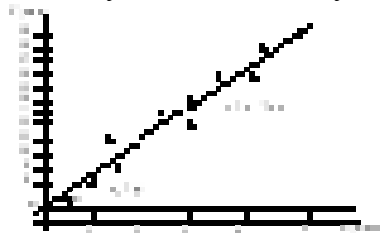
Расстояние, миль	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Время, мин	16	13	19	18	12	11	8	14	9	16

Для проведения регрессионного анализа:

1. построить график исходных данных, приближенно определить характер зависимости;
2. выбрать вид функции регрессии и определить численные коэффициенты модели методом наименьших квадратов и направление связи;
3. оценить силу регрессионной зависимости с помощью коэффициента детерминации;
4. оценить значимость уравнения регрессии;
5. сделать прогноз (или вывод о невозможности прогнозирования) по принятой модели для расстояния 2 мили.

Решение.

1. Построенные точки не находятся точно на линии: помимо расстояния на время поставки влияют пробки на дорогах, время суток, дорожные работы, погода, квалификация водителя, вид транспорта. Но эти точки собраны вдоль прямой линии, поэтому можно предположить линейную положительную связь между параметрами.



2. Вычислим суммы, необходимые для расчета коэффициентов уравнения линейной регрессии и коэффициента детерминации R^2 :

№	x_i	y_i	x_i^2	$x_i y_i$	\hat{y}_i	$(\hat{y}_i - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	12,25	56,00	15,22	2,63	5,76
2	2,4	13	5,76	31,20	12,30	1,70	0,36
3	4,9	19	24,01	93,10	18,95	28,59	29,16
4	4,2	18	17,64	75,60	17,09	12,15	19,36
5	3,0	12	9,00	36,00	13,89	0,08	2,56
6	1,3	11	1,69	14,30	9,37	17,88	6,76
7	1,0	8	1,00	8,00	8,57	25,27	31,36
8	3,0	14	9,00	42,00	13,89	0,09	0,16
9	1,5	9	2,25	13,50	9,90	13,67	21,16
10	4,1	16	16,81	65,60	16,82	10,36	5,76
Σ	28,9	136	99,41	435,30	–	112,42	122,40

$$\bar{x} = \frac{\sum n_i x_i}{n} = 2,89$$

$$\bar{y} = \frac{\sum n_i y_i}{n} = 13,6$$

$$b = \frac{13,6 \cdot 9,941 - 2,89 \cdot 43,53}{9,941 - 2,89^2} = 5,91$$

$$a = \frac{43,53 - 2,89 \cdot 13,6}{9,941 - 2,89^2} = 2,66$$

Искомая регрессионная зависимость имеет вид:

$$\hat{y} = 2,66x + 5,91$$

Определяем направление связи между переменными: знак коэффициента регрессии положительный, следовательно, связь также является положительной, что подтверждает графическое предположение.

$$R^2 = \frac{112,42}{122,40} = 0,92$$

3. Вычислим коэффициент детерминации:

или 92%.

Таким образом, линейная модель объясняет 92% вариации времени поставки, что означает правильность выбора фактора (расстояния). Не объясняется 8% вариации времени, которые обусловлены остальными факторами, влияющими на время поставки, но не включенными в линейную модель регрессии.

4. Проверим значимость уравнения регрессии:

$$F_{\text{набл}} = \frac{0,92^2}{1-0,92^2} \cdot \frac{10-1-1}{1} = 44,1$$

Т.к. $F_{\text{набл}} = 44,1 > F_{\text{кр}}(0,05; 1; 10-1-1) = 5,32$ – уравнение регрессии (линейной модели) статистически значимо.

5.3. Вопросы к зачету

1. Понятия простой выборки и статистики. Примеры статистик: моменты, асимметрия и эксцесс, вариационный ряд и порядковые статистики, эмпирическое распределение.
2. Статистические точечные оценки и их свойства: несмещённость, состоятельность, оптимальность, робастность.
3. Интервальные оценки, понятия доверительного интервала и уровня доверия. Доверительные интервалы для среднего и медианы.
4. Часто используемые распределения: нормальное, хи-квадрат, Фишера, Стьюдента, Бернулли, биномиальное, Пуассона.
5. Проверка статистических гипотез, основные понятия: уровень значимости, достигаемый уровень значимости, ошибки I и II рода. Односторонние и двусторонние альтернативы.
6. Свойства критериев: несмещённость, состоятельность, мощность.
7. Критерии нормальности: критерий хи-квадрат (Пирсона)
8. Критерии, основанные на различиях между эмпирической и теоретической функциями распределения, критерий Колмогорова-Смирнова.
9. Гипотезы о средних: t- и z-критерии Стьюдента для одной и двух выборок, связанные выборки
10. Гипотезы о дисперсиях: критерии хи-квадрат и Фишера.
11. Критерии знаков: одновыборочный, для связанных выборок.
12. Вариационный ряд, ранги и связки.
13. Ранговые критерии: критерий Уилкоксона-Манна-Уитни.
14. Однофакторная модель. Независимые выборки: критерии Фишера, Краскела-Уоллиса
15. Проверка гипотезы о равенстве дисперсий
16. Корреляция Пирсона. Значимость коэффициента корреляции
17. Многомерная линейная регрессия. Примеры прикладных задач.
18. Метод наименьших квадратов.
19. Предположения Гаусса-Маркова. Статистические свойства МНК-оценок в отсутствие предположения нормальности.
20. Анализ структуры линейной регрессионной модели. Значимость коэффициентов линейной регрессии