

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Экономический институт
Кафедра экономической кибернетики и прикладной статистики
(наименование кафедры)

УТВЕРЖДАЮ
Директор

« 28 » декабрь 20 20 года



**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по учебной дисциплине (практике)**

Многомерные статистические методы

(наименование учебной дисциплины, практике)

38.03.05 Бизнес-информатика

(код и наименование направления подготовки (специальности))

Экономическая аналитика и бизнес-статистика

(наименование профиля подготовки (специальности, магистерской программы); при отсутствии ставится прочерк)

Разработчик (разработчики):

доцент

(подпись)

(подпись)

Спорняк С.А.

ФОС рассмотрен и одобрен на заседании кафедры экономической кибернетики и прикладной статистики от « 25 » 02 2025 г., протокол № 25

Заведующий кафедрой экономической кибернетики и прикладной статистики

Велигур А.В.

(подпись)

Луганск 2025 г.

Комплект оценочных материалов по дисциплине
«Многомерные статистические методы»

Задания закрытого типа

Задания закрытого типа на выбор правильного ответа

1. Выберите один правильный ответ.

Итеративным методом кластерного анализа является

- А) метод k-средних
- Б) метод одиночной связи
- В) метод средней связи
- Г) метод Уорда

Правильный ответ: А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Выберите один правильный ответ.

Экономическое содержание коэффициентов регрессии многофакторной модели заключается в том, что они показывают

- А) влияние факторов на прогнозируемый показатель
- Б) изменение прогнозируемого показателя
- В) зависимость или независимость факторов друг от друга
- Г) существенность факторов регрессионной модели
- Д) зависимость или независимость факторов от остатков модели

Правильный ответ: А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

3. Выберите один правильный ответ.

Критические значения критерия Фишера определяются

- А) по уровню значимости и двум степеням свободы
- Б) по уровню значимости
- В) по двум степеням свободы
- Г) по уровню значимости и одной степени свободы

Правильный ответ: А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

4. Выберите один правильный ответ.

Уровень значимости определяет

- А) вероятность принятия неверной гипотезы
- Б) вероятность совершения ошибки при вычислении среднего выборки
- В) процент отклонения верной гипотезы
- Г) вероятность совершения ошибки при вычислении выборочной дисперсии

Правильный ответ: А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

5. Выберите один правильный ответ.

При построении многофакторных регрессионных моделей вида $Y = a + f(X_1, \dots, X_n)$ параметр a измеряется

- А) в процентах
- Б) в тех же единицах, что и Y
- В) в тех же единицах, что и X
- Г) в денежных единицах

Правильный ответ: Б.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

6. Выберите один правильный ответ.

Оценка значимости всего уравнения регрессии в целом осуществляется с помощью

- А) F-критерия Фишера
- Б) t-распределения
- В) нормального распределения
- Г) распределения Пуассона

Правильный ответ: А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

Задания закрытого типа на установление соответствия

1. Установите соответствие между иерархической процедурой с их процедурой проведения. Каждому элементу левого столбца соответствует только один элемент правого столбца.

1. Агломеративные процедуры –	A) В начале работы алгоритма все объекты принадлежат одному кластеру, который на последующих шагах делится на меньшие
2. Дивизимные процедуры –	Б) это метод иерархической кластеризации, при котором новые кластеры создаются путём объединения более мелких кластеров
3. Агломеративные пороговые процедуры –	В) Задана монотонная последовательность порогов $t \in \{ \dots, c_1 \}$. На j -шаге объединяются все объекты, расстояние между которыми не превышает $j \in \{ \dots, t \}$
	Г) В начале работы алгоритма все объекты принадлежат двум кластерам, которые на последующих шагах делятся на меньшие

Ответ: 1-Б; 2-А; 3-В.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Установите соответствие между названием видов кластерного анализа и их процессом. Каждому элементу левого столбца соответствует только один элемент правого столбца.

1) Иерархический кластерный анализ	А) Ищет k центров кластеров и назначает объекты ближайшим центрам.
2) K-средних	Б) Создает иерархию кластеров, начиная либо с отдельных объектов (агломеративный подход), либо объединяя все объекты в один кластер (дивизивный подход). разность между выручкой от продажи и затратами на производство и реализацию
3) Метод DBSCAN	В) Моделирует данные как смесь гауссовых распределений.
4) Спектральный кластерный анализ	Г) Группирует точки на основе плотности окрестности.
	Д) Преобразует данные в спектральное пространство и использует евклидово расстояние для разделения точек.

Правильный ответ: 1-Б; 2-А; 3-Г; 4-Д.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

3. Установите соответствие между методами многомерного статистического анализа и их характеристиками. Каждому элементу левого столбца соответствует только один элемент правого столбца.

1) Факторный анализ	А) Метод, который группирует объекты на основе их сходства.
2) Кластерный анализ	Б) Метод, используемый для классификации объектов в заранее определенные группы.
3) Дискриминантный анализ	В) Метод, используемый для выявления скрытых факторов, которые объясняют корреляции между переменными.
4) Анализ главных компонент (PCA)	Г) Метод, который преобразует множество коррелированных переменных в меньшее количество некоррелированных переменных.
5) Многомерное шкалирование	

Д) Метод, который отображает многомерные данные в виде двумерного или трехмерного пространства.

Е) Метод, используемый для прогнозирования вероятности события на основе множества переменных.

Правильный ответ: 1-В; 2-А; 3-Б; 4-Г; 5-Д.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

4. Установите соответствие между названием видов кластерного анализа и их недостатками. Каждому элементу левого столбца соответствует только один элемент правого столбца.

1) Иерархический кластерный анализ	А) чувствителен к выбросам, требует большого объема памяти для больших данных.
2) К-средних	Б) требует предварительного выбора количества кластеров (k), чувствителен к начальному расположению центров.
3) Метод DBSCAN	В) сложен в настройке параметров, требует хорошей предварительной обработки данных.
4) Спектральный кластерный анализ	Г) требует хорошего подбора начальных параметров, чувствителен к выбору числа компонент.
5) Модель смеси Гауссианов	Д) медленный на больших данных, требует предварительной нормализации данных.

Правильный ответ: 1-А; 2-Б; 3-В; 4-Д; 5-Г.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

Задания закрытого типа на установление правильной последовательности

1. В алгоритме вычисления канонических корреляций укажите правильную последовательность шагов. Запишите правильную последовательность букв слева направо:

- Извлечение квадратного корня из полученных собственных значений для получения набора чисел, интерпретируемых как коэффициенты корреляции.
- На каждом шаге поиск следующей пары канонических переменных, имеющих максимальную корреляцию и не коррелированных с предыдущими парами, и вычисление соответствующего ей значения канонического корня.
- Вычисление весов, максимизирующих корреляцию между взвешенными суммами по двум множествам, и определение соответствующего им значения первого корня.

Правильный ответ: В; Б; А.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Установите правильную последовательность этапов процедуры дискриминантного анализа. Запишите правильную последовательность букв слева направо:

- А) Выбор переменных – предикторов.
- Б) Разделение выборки на две части.
- В) Вычисление параметров дискриминантной функции.
- Г) Интерпретация результатов.

Правильный ответ: Б; А; В; Г.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

3. Установите правильную последовательность этапов процедуры факторного анализа. Запишите правильную последовательность букв слева направо:

- А) Определение цели анализа
- Б) Сбор данных
- В) Преобразование данных
- Г) Анализ корреляций
- Д) Выбор метода факторного анализа
- Е) Проведение факторного анализа
- Ж) Интерпретация результатов
- З) Проверка достоверности
- И) Представление результатов

Правильный ответ: Б; А; В; Г.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

Задания открытого типа

Задания открытого типа на дополнение

1. Напишите пропущенное словосочетание.

Среднее отклонение расчетных значений от фактических выявляется путем расчета _____.

Правильный ответ: средней ошибки аппроксимации.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Напишите пропущенное слово.

Длина вектора переменной в факторном анализе равна _____ из общности.

Правильный ответ: корню квадратному.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

3. Напишите пропущенное слово.

Проверка статистической значимости уравнения регрессии в целом осуществляется по критерию _____.

Правильный ответ: Фишера.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

4. Напишите пропущенное слово.

Для факторных нагрузок a_{ij} и остаточных дисперсий l_{it}^2 оценивание производится либо методом максимального правдоподобия, либо так называемым _____ методом.

Правильный ответ: центроидным.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

Задания открытого типа с кратким свободным ответом

1. Напишите пропущенное словосочетание

_____ является самой популярной метрикой в кластерном анализе. Оно попросту является геометрическим расстоянием в многомерном пространстве. Геометрически оно лучше всего объединяет объекты в шарообразных скоплениях.

Правильный ответ: Евклидово расстояние.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Вставьте пропущенный вывод.

При исследовании зависимости балансовой прибыли предприятия торговли (Y , тыс. руб.) от объема продаж по безналичному расчету (X , тыс. руб.) получена следующая модель:

$$Y = 5933,100 + 0,651X + \varepsilon.$$

При факторном признаке X интерпретируется коэффициент: при увеличении только объема продаж по безналичному расчету на 1 тыс. руб. балансовая прибыль предприятия торговли будет _____.

Правильный ответ: увеличиваться на 651 руб.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

3. Запишите уравнение.

Согласно коэффициентов при общих факторах Fr в матричном виде имеет вид: _____.

Правильный ответ: $Zj = a_{j1}F1 + a_{j2}F2 + \dots + a_{jm}Fm + a_{jDj}$

или в матричной форме:

$$Zj = AF + a_{jDj}$$

где Dj – характерный фактор.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

4. Запишите уравнение.

Фундаментальная теорема факторного анализа выглядит следующим образом _____ (где: А – матрица весовых коэффициентов, С – корреляционная матрица стохастических связей между общими факторами, F – матрица индивидуальных значений факторов для каждого объекта исследования, R – матрица парных коэффициентов корреляции, Rh – редуцированная корреляционная матрица)

Правильный ответ: $Rh = ACAT$, $Rh = AAT$.

Компетенции (индикаторы): ПК-6 (ПК-6.2).

Задания открытого типа с развернутым ответом

1. Решите задачу по множественному корреляционно-регрессионному анализу.

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7,0	3,9	10,0	11	9,0	6,0	21,0
2	7,0	3,9	14,0	12	11,0	6,4	22,0
3	7,0	3,7	15,0	13	9,0	6,8	22,0
4	7,0	4,0	16,0	14	11,0	7,2	25,0
5	7,0	3,8	17,0	15	12,0	8,0	28,0
6	7,0	4,8	19,0	16	12,0	8,2	29,0
7	8,0	5,4	19,0	17	12,0	8,1	30,0
8	8,0	4,4	20,0	18	12,0	8,5	31,0
9	8,0	5,3	20,0	19	14,0	9,6	32,0
10	10,0	6,8	20,0	20	14,0	9,0	36,0

Построить линейную модель множественной регрессии.

Привести расширенное решение.

Время выполнения – 45 мин.

Ожидаемый результат:

Для удобства проведения расчетов поместим результаты промежуточных расчетов в таблицу:

№	y	x_1	x_2	yx_1	yx_2	x_1x_2	y^2	x_1^2	x_2^2
1	2	3	4	5	6	7	8	9	10
1	7,0	3,9	10,0	27,3	70,0	39,0	49,0	15,21	100,0
2	7,0	3,9	14,0	27,3	98,0	54,6	49,0	15,21	196,0
3	7,0	3,7	15,0	25,9	105,0	55,5	49,0	13,69	225,0
4	7,0	4,0	16,0	28,0	112,0	64,0	49,0	16,0	256,0
5	7,0	3,8	17,0	26,6	119,0	64,6	49,0	14,44	289,0
6	7,0	4,8	19,0	33,6	133,0	91,2	49,0	23,04	361,0

7	8,0	5,4	19,0	43,2	152,0	102,6	64,0	29,16	361,0
8	8,0	4,4	20,0	35,2	160,0	88,0	64,0	19,36	400,0
9	8,0	5,3	20,0	42,4	160,0	106,0	64,0	28,09	400,0
10	10,0	6,8	20,0	68,0	200,0	136,0	100,0	46,24	400,0
11	9,0	6,0	21,0	54,0	189,0	126,0	81,0	36,0	441,0
12	11,0	6,4	22,0	70,4	242,0	140,8	121,0	40,96	484,0
13	9,0	6,8	22,0	61,2	198,0	149,6	81,0	46,24	484,0
14	11,0	7,2	25,0	79,2	275,0	180,0	121,0	51,84	625,0
15	12,0	8,0	28,0	96,0	336,0	224,0	144,0	64,0	784,0
16	12,0	8,2	29,0	98,4	348,0	237,8	144,0	67,24	841,0
17	12,0	8,1	30,0	97,2	360,0	243,0	144,0	65,61	900,0
18	12,0	8,5	31,0	102,0	372,0	263,5	144,0	72,25	961,0
19	14,0	9,6	32,0	134,4	448,0	307,2	196,0	92,16	1024,0
20	14,0	9,0	36,0	126,0	504,0	324,0	196,0	81,0	1296,0
Сумма	192	123,8	446	1276,3	4581	2997,4	1958,0	837,74	10828,0
Средн. знач.	9,6	6,19	22,3	63,815	229,05	149,87	97,9	41,887	541,4

1. Найдем средние квадратические отклонения признаков:

$$\sigma_y = \sqrt{y^2 - \bar{y}^2} = \sqrt{97,9 - 9,6^2} = 2,396;$$

$$\sigma_{x_1} = \sqrt{x_1^2 - \bar{x}_1^2} = \sqrt{41,887 - 6,19^2} = 1,890;$$

$$\sigma_{x_2} = \sqrt{x_2^2 - \bar{x}_2^2} = \sqrt{541,4 - 22,3^2} = 6,642.$$

2. Вычисление параметров линейного уравнения множественной регрессии.

Для нахождения параметров линейного уравнения множественной регрессии

$$\hat{y} = a + b_1 x_1 + b_2 x_2$$

необходимо решить следующую систему линейных уравнений относительно неизвестных параметров a , b_1 , b_2 :

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 = \sum y; \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum yx_1; \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum yx_2 \end{cases}$$

либо воспользоваться готовыми формулами:

$$b_1 = \frac{\sigma_y \cdot r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}; \quad b_2 = \frac{\sigma_y \cdot r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}; \quad a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2.$$

Рассчитаем сначала парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\text{cov}(y, x_1)}{\sigma_y \cdot \sigma_{x_1}} = \frac{63,815 - 6,19 \cdot 9,6}{1,890 \cdot 2,396} = 0,970;$$

$$r_{yx_2} = \frac{\text{cov}(y, x_2)}{\sigma_y \cdot \sigma_{x_2}} = \frac{229,05 - 22,3 \cdot 9,6}{6,642 \cdot 2,396} = 0,941;$$

$$r_{x_1 x_2} = \frac{\text{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{149,87 - 6,19 \cdot 22,3}{1,890 \cdot 6,642} = 0,943.$$

Находим

$$b_1 = \frac{2,396 \cdot 0,970 - 0,941 \cdot 0,943}{1 - 0,943^2} = 0,946;$$

$$b_2 = \frac{2,396}{6,642} \cdot \frac{0,941 - 0,970 \cdot 0,943}{1 - 0,943^2} = 0,0856;$$

$$a = 9,6 - 0,946 \cdot 6,19 - 0,0856 \cdot 22,3 = 1,835.$$

Таким образом, получили следующее уравнение множественной регрессии:

$$\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2.$$

Ответ: уравнение множественной регрессии $\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2$.

Критерии оценивания:

- расчет промежуточных расчетов и расчет средних квадратических отклонений признаков;
- Вычисление параметров и запись линейного уравнения множественной регрессии.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

2. Решите задачу с оценкой значимости множественного уравнения регрессии

Рассмотрим данные:

Y — продажи (единиц),
X1 — цена (доллары),
X2 — реклама (доллары).

Предположим, что после оценки параметров модели методом наименьших квадратов мы получили следующее уравнение регрессии:

$$Y = 500 - 2X1 + 0,5X2$$

Также допустим, что для F-теста мы получили значение F=10 с p-значением p=0,01, а для t-тестов коэффициентов:

$$\begin{aligned} t\beta_1 &= -3, & p &= 0,05 \\ t\beta_2 &= 2, & p &= 0,1 \end{aligned}$$

Оценить значимости множественного уравнения регрессии и коэффициентов.

Привести расширенное решение.

Время выполнения — 10 мин.

Ожидаемый результат:

1. Значимость модели:

Поскольку F=10 и p=0,01, модель в целом значима. Это означает, что хотя бы один из коэффициентов при независимых переменных значимо отличается от нуля.

2. Значимость коэффициентов:

Коэффициент при цене (β_1) значим ($t=-3, p=0,05$), что указывает на то, что цена отрицательно влияет на продажи.

Коэффициент при рекламе (β_2) менее значим ($t=2, p=0,1$), что может указывать на слабую связь рекламы с продажами.

Таким образом, мы оценили значимость множественного уравнения регрессии и пришли к выводу, что модель в целом значима, а влияние цены на продажи подтверждено статистически.

Ответ: модель в целом значима, влияние цены на продажи подтверждено статистически.

Критерии оценивания:

- приведена интерпретация значимости модели;
- приведена интерпретация значимости коэффициентов.

Компетенции (индикаторы): ПК-6 (ПК-6.1).

Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Многомерные статистические методы» соответствует требованиям ФГОС ВО.

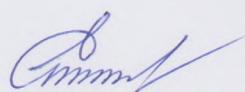
Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 38.03.05 Бизнес-информатика.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки обучающихся по указанному направлению.

Председатель учебно-методической
комиссии экономического института



Шаповалова Е.Н.

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)