

Санкт-Петербургский государственный университет

**Р А Б О Ч А Я П Р О Г Р А М М А
У Ч Е Б Н О Й Д И С Ц И П Л И Н Ы**

Методы статистической обработки информации
Methods of Statistical Information Processing

Язык(и) обучения

русский

Трудоемкость в зачетных единицах: 3

Регистрационный номер рабочей программы: 001123

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Цель дисциплины – получение базовых знаний о курсе статистического анализа с навыком применением пакетов прикладных статистических программ для получения результатов. Изучаются средства среды Excel и других пакетов, включая пакеты R и Python, для проведения статистических вычислений, позволяющих исследовать при помощи компьютера, эконометрические модели. В рамках курса выполняются тестовые и лабораторные работы, показывающие уровень понимания студентом методов статистической обработки данных, включая анализ данных медицинского и экономического характера.

1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Для понимания материала дисциплины необходимы знания по курсам: теория вероятностей и математическая статистика. Владение навыками работы с Microsoft Excel. Знание одного из высокоуровневых языков программирования: R или python.

1.3. Перечень результатов обучения (learning outcomes)

Дисциплина способствует формированию следующих компетенций: ПК-2 способен творчески применять базовые знания математических и естественных наук, программирования и информационных технологий; ПК-5 способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования. В результате изучения дисциплины студент должен знать фундаментальные методы статистического анализа данных, основные статистические алгоритмы; уметь реализовывать статистические алгоритмы анализа данных в решении прикладных задач, уметь экспериментально (с помощью компьютера) исследовать эффективность результатов статистического исследования, применять средства Excel и других пакетов для анализе данных. владеть навыками применения статистических алгоритмов в существующем программном обеспечении, приобрести практические навыки самостоятельной работы в решении задач прикладного анализа статистических данных.

1.4. Перечень и объём активных и интерактивных форм учебных занятий

Все занятия необходимо проводить с привлечением интерактивных методов: работа в малых группах, групповое обсуждение материалов лекций, представление самостоятельно выполненных индивидуальных заданий и коллективное обсуждение полученных результатов – 32 часа.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

2.1.1

Основной

курс

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																		
Код модуля в составе дисциплины, практики и т.п.	Контактная работа обучающихся с преподавателем											Самостоятельная работа				форм учебных занятий	Трудоёмкость	
	лекции	семинары	консультации	практические занятия	лабораторные работы	контрольные работы	коллоквиумы	текущий контроль	промежуточная аттестация	итоговая аттестация	под руководством преподавателя	в присутствии преподавателя	сам. раб. с использованием методических материалов	текущий контроль (сам.раб.)	промежуточная аттестация (сам.раб.)			итоговая аттестация (сам.раб.)
ОСНОВНАЯ ТРАЕКТОРИЯ																		
Форма обучения: очная																		
Семестр 1					32				2				67		7		32	3
					2-8				10-25				1-1		1-1			
ИТОГО					32				2				67		7			3

Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации						
Код модуля в составе дисциплины, практики и т.п.	Формы текущего контроля успеваемости		Виды промежуточной аттестации		Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)	
	Формы	Сроки	Виды	Сроки	Виды	Сроки
ОСНОВНАЯ ТРАЕКТОРИЯ						
Форма обучения: очная						
Семестр 1			зачёт, устно, традиционная форма	по графику промежуточной аттестации		

2.2. Структура и содержание учебных занятий

Введение.

Статистическое оценивание. Описательная статистика, основные законы распределения. Точечные оценки: метод моментов, метод максимального правдоподобия, другие методы. Статистическое оценивание в R и Excel.

Проверка статистических гипотез. Общая схема проверки статистических гипотез, критерии согласия для простых и сложных гипотез (критерии хи-квадрат, Колмогорова). Критерии однородности (критерии Колмогорова-Смирнова, ранговый критерий Вилкоксона, критерий Манна-Уитни). Критерии равенства средних и дисперсий (Стьюдента, Фишера). Критерии о значении параметров распределений. Однофакторный и двухфакторный дисперсионный анализ и их непараметрические аналоги.

Регрессионный анализ. Линейный регрессионный анализ. Бинарная регрессия. Квантильная регрессия. Ридж-регрессия. Реализация методов в R и Excel.

Статистическая классификация. Дискриминантный анализ, кластерный анализ.

Временные ряды. Выделение тренда и периодической компоненты. Прогнозирование.

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1 Методические указания по освоению дисциплины

Самостоятельная работа студентов включает в себя изучение учебников, учебных пособий и иных материалов. Время и место самостоятельной работы выбираются студентами по своему усмотрению с учетом рекомендаций преподавателя. Самостоятельную работу над дисциплиной следует начинать с изучения учебно-методического комплекса, который содержит основные требования к знаниям, умениям, навыкам. Необходимо также вспомнить рекомендации преподавателя, данные в ходе консультаций, затем приступить к изучению отдельных разделов и тем по рекомендованным учебникам и учебным пособиям. После чего приступить к выполнению лабораторных работ по прикладной статистике и к подготовке отчетов.

3.1.2 Методическое обеспечение самостоятельной работы

Комплект заданий, представляющих собой описание лабораторных работ с указаниями для самостоятельного выполнения студентами индивидуальных заданий. Рекомендованная учебная литература.

3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Для успешного прохождения промежуточной аттестации необходимо выполнить 18 лабораторных работ и 4 теста, необходимо представить результаты проведенных вычислений (Excel-файлы или файлы вычислений в R или в Python) и 22 отчета (word – файлы по лабораторным работам и по тестам), отчеты по лабораторным работам могут быть представлены в электронном виде, отчеты по тестам также в электронном виде и, кроме того, они должны быть распечатаны. Преподавателем могут быть заданы теоретические вопросы по темам лабораторных работ и тестов, знание которых необходимо для правильного понимания проведенных статистических вычислений. Теоретические вопросы не предполагают обязательного доказательства теорем, они задаются с целью проверки правильного понимания алгоритмов и статистических методов, реализованных в ходе выполнения лабораторных работ и тестов. При условии выполнения всех вычислительных заданий (лабораторные работы и тесты) и при условии правильного понимания примененных методов и алгоритмов студент получает положительную оценку во время прохождения аттестации.

3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

Список основных тем лабораторных работ и тестов (теоретические вопросы могут быть заданы только по перечисленным ниже темам):

1. Описательная статистика, основные законы распределения.
2. Точечные оценки: метод моментов, метод максимального правдоподобия, другие методы.

3. Статистическое оценивание в R и Excel.
4. Общая схема проверки статистических гипотез, критерии согласия для простых и сложных гипотез (критерии хи-квадрат, Колмогорова).
5. Критерии однородности (критерии Колмогорова-Смирнова, ранговый критерий Вилкоксона, критерий Манна-Уитни, критерий знаков).
6. Критерии равенства средних и дисперсий (Стьюдента, Фишера).
7. Критерии о значении параметров распределений.
8. Однофакторный и двухфакторный дисперсионный анализ и их непараметрические аналоги.
9. Линейный регрессионный анализ. Бинарная регрессия. Квантильная регрессия. Ридж-регрессия. Реализация методов в R и Excel.
10. Дискриминантный анализ, кластерный анализ.
11. Выделение тренда и периодической компоненты. Прогнозирование.

Примеры лабораторных работ

ЛАБОРАТОРНАЯ РАБОТА: Интервальное оценивание

Описание работы соответствует функциям Excel, в R-project используйте аналогичные функции.

Интервальная оценка параметров распределения. Доверительные интервалы

1. Смоделируйте генеральную совокупность из 200 значений нормально распределенной случайной величины. Запишите использованные при этом значения математического ожидания μ и среднеквадратического отклонения σ . Построенная генеральная совокупность является выборкой из нормального распределения объема $n=200$. Постройте гистограмму и полигон относительных частот для смоделированной генеральной совокупности (для 8 и 12 интервалов разбиения).

2. Вычислите дисперсию σ_g^2 и среднеквадратическое отклонение σ_g полученной генеральной совокупности, с помощью функций **ДИСПР** и **СТАНДОТКЛОНП** соответственно. Вычислите математическое ожидание смоделированной генеральной совокупности μ_g .

3. Вычислите вероятность попадания случайной величины $X \in N(\mu, \sigma^2)$ в интервал $[5;11]$: $P\{5 < X < 11\} = F(11) - F(5)$. Для нахождения значений функции распределения $F(x)$ нормальной случайной величины воспользуйтесь функцией **НОРМРАСП**, аргументами которой являются:

1. **X** — аргумент функции распределения (граница интервала);
2. **Среднее** — математическое ожидание случайной величины μ ;
3. **Стандартное_откл** — среднеквадратическое отклонение σ ;
4. **Интегральный** — параметр, указывающий функцию для вычисления (при вычислении значения функции распределения значение поля должно быть 1).

Вычислите аналогичную вероятность (относительную частоту) для смоделированной генеральной совокупности. Сравните вероятности.

4. Сгенерируйте две случайные выборки объемом 30 значений из исходной нормально распределенной генеральной совокупности (с математическим ожиданием a и среднеквадратическим отклонением σ).

5. Для каждой выборки постройте точечные оценки параметров распределения (приведены названия функций):

1. Выборочное среднее \bar{X}_o (**СРЗНАЧ**);
2. Выборочная дисперсия (несмещенная) \tilde{S}^2 (**ДИСП**);
3. Выборочное среднеквадратическое отклонение \tilde{S} (**СТАНДОТКЛОН**).

В дальнейшем понадобится объем выборки. Его можно вычислить с помощью функции **СЧЕТ**.

6. Для каждой из выборок постройте гистограммы относительных частот (5 и 7 интервалов)

7. Найдите доверительный интервал для оценки неизвестного математического ожидания с уровнем доверия $\gamma = 0,95$, предполагая, что известно среднеквадратическое отклонение генеральной совокупности σ (здесь и далее в пп 9 и 10 доверительные интервалы строятся по трем выборкам и в отчете располагаются рядом, сделайте выводы о размахе полученных доверительных интервалов)

Для этого вычислите значение Z_{tabl} , входящее в формулу доверительного интервала с помощью функции Excel **НОРМСТОБР** (можно также воспользоваться функцией **НОРМОБР**, указав в качестве среднего значения 0, а в качестве стандартного отклонения 1). Для того чтобы правильно воспользоваться функцией **НОРМСТОБР**, необходимо учесть, что значение поля **Вероятность** в этой функции представляет собой вероятность $F(z) = \Pr\{X \leq z\}$, т.е. равно площади под графиком плотности от $-\infty$ до Z . Поэтому следует взять значение вероятности равное $0,95 + 0,05/2 = 0,975$. Это значение указывается в поле **Вероятность**.

Вычислите границы доверительного интервала по формуле

$$\bar{X}_o - Z_{tabl} \frac{\sigma}{\sqrt{n}} < a < \bar{X}_o + Z_{tabl} \frac{\sigma}{\sqrt{n}}.$$

Проверьте полученное значение точности оценки $\delta = Z_{tabl} \frac{\sigma}{\sqrt{n}}$ с помощью функции **ДОВЕРИТ**.

8. Постройте доверительный интервал для оценки математического ожидания при неизвестном значении генерального среднеквадратического отклонения по формуле

$$\bar{X}_o - t_{tabl} \frac{\tilde{S}}{\sqrt{n}} < a < \bar{X}_o + t_{tabl} \frac{\tilde{S}}{\sqrt{n}}$$

где \tilde{S} — среднеквадратическое отклонение, вычисленное по несмещенной выборочной дисперсии, а t_{tabl} находится по заданному значению уровня доверия 0,95 по таблице распределения Стьюдента для $n-1$ степеней свободы или с помощью функции **СТЮДРАСПОБР**.

9. Найдите доверительный интервал с уровнем доверия 0,95 для дисперсии при неизвестном значении генерального среднего по формуле

$$\frac{(n-1)\tilde{S}^2}{t_2} \leq \sigma^2 \leq \frac{(n-1)\tilde{S}^2}{t_1},$$

и для генерального среднеквадратического отклонения по формуле

$$\frac{\sqrt{n-1} \tilde{S}}{\sqrt{t_2}} \leq \sigma \leq \frac{\sqrt{n-1} \tilde{S}}{\sqrt{t_1}},$$

где \tilde{S} — среднеквадратическое отклонение, вычисленное по несмещенной выборочной дисперсии, t_2 находится по заданному значению уровня доверия 0,025 по таблице распределения Пирсона хи-квадрат для $n-1$ степеней свободы или с помощью функции

хи2обр, t_1 находится по заданному значению уровня доверия 0,975 по таблице распределения Пирсона хи-квадрат для $n-1$ степеней свободы или с помощью функции

хи2обр.

10. Смоделируйте выборку из биномиального распределения объема $n = 500$ для выбранного значения вероятности успеха P , пользуясь интегральной предельной теоремой Муавра-Лапласа постройте приближенный 95% доверительный интервал для параметра P , заменяя вероятности успеха P и неудачи q в знаменателе относительными частотами, рассчитанными по выборке, или воспользуйтесь асимптотическими интервалами из лекции.

Сформируйте отчет по всем выполненным заданиям.

ЛАБОРАТОРНАЯ РАБОТА: Проверка статистических гипотез

Описание работы соответствует функциям Excel, в R-project используйте аналогичные функции)

I. Критерии однородности выборок

1. Смоделируйте две выборки (X_1, \dots, X_n) , (Y_1, \dots, Y_m) из нормального распределения с произвольными математическими ожиданиями a_x , a_y и произвольными дисперсиями σ_x^2 , σ_y^2 (Положите, например, $n=50$, $m=30$).

2. Проверьте гипотезу однородности $H_0: a_x = a_y$, считая дисперсии известными.

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Примените для проверки двух выборочный z -тест, вычислите статистику

Найдите $t_{табл}$ соответствующее вероятности ошибки первого рода 0,05, с помощью таблицы стандартного нормального распределения или с помощью функции Excel **НОРМСТОБР** (можно также воспользоваться функцией **НОРМОБР**, указав в качестве среднего значения 0, а в качестве стандартного отклонения 1). Для того чтобы правильно

воспользоваться функцией **НОРМСТОБР**, необходимо учесть, что значение поля **Вероятность** в этой функции представляет собой вероятность $F(t) = \Pr\{X < t\}$, т.е. равно площади под графиком плотности от $-\infty$ до t , следует взять значение вероятности равное $0,95+0,05/2=0,975$. Это значение указывается в поле **Вероятность**.

3. Сравните значение вычисленной статистики с найденным значением. По результатам сравнения примите или отвергните проверяемую гипотезу.

4. Проверьте полученные результаты с помощью пакета **Анализ данных** в Excel.

5. Для смоделированных выборок проверьте гипотезу о равенстве дисперсий $H_0: \sigma_x^2 = \sigma_y^2$ с помощью двухвыборочного F-теста для дисперсий. Вычислите статистику:

$$F = \frac{\tilde{S}_x^2}{\tilde{S}_y^2}.$$

Найдите $t_{табл}$ соответствующее вероятности ошибки первого рода 0,05, с помощью таблиц распределения Фишера с числами степеней свободы $n-1$ и $m-1$ или с помощью функции **Фраспобр** в Excel.

6. Сравните вычисленное значение статистики с найденным табличным значением и сделайте вывод о принятии или непринятии гипотезы равенства дисперсий.

7. Проверьте полученные результаты с помощью пакета **Анализ данных** в Excel.

8. Смоделируйте две выборки $(X_1, \dots, X_n), (Y_1, \dots, Y_m)$ из нормального распределения с произвольными математическими ожиданиями a_x, a_y и равными дисперсиями $\sigma_x^2 = \sigma_y^2 = \sigma^2$. (Положите, например, $n=50, m=30$).

9. Проверьте гипотезу однородности $H_0: a_x = a_y$, с помощью двухвыборочного t-теста, считая дисперсии неизвестными, но одинаковыми. На практике этот тест применяют для проверки гипотезы о равенстве средних, если применение двухвыборочного F-теста к изучаемым выборкам показало, что можно принять гипотезу о равенстве дисперсий. Вычислите статистику:

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}, \quad s^2 = \frac{\tilde{S}_x^2(n-1) + \tilde{S}_y^2(m-1)}{(n-1) + (m-1)}, \text{ где}$$

Найдите $t_{табл}$ соответствующее вероятности ошибки первого рода 0,05, с помощью таблиц распределения Стюдента с числом степеней свободы $n+m-2$ или с помощью функции **Стюдраспобр** в Excel.

10. Сравните вычисленное значение статистики с найденным табличным значением и сделайте вывод о принятии или непринятии гипотезы равенства средних.

11. Проверьте полученные результаты с помощью пакета **Анализ данных** в Excel.

12. Смоделируйте две выборки $(X_1, \dots, X_n), (Y_1, \dots, Y_m)$ из нормального распределения с произвольными математическими ожиданиями a_x, a_y и равными дисперсиями $\sigma_x^2 = \sigma_y^2 = \sigma^2$. (Положите, например, $n=9, m=7$).

13. Примените **критерий Уилкоксона** к проверке гипотезы о статистической однородности двух выборок (двусторонняя альтернатива, 5% уровень значимости).

14. Оформите отчет по всем выполненным заданиям. (Отчет оформляется в MS Word 2003 или в подобном редакторе)

ЛАБОРАТОРНАЯ РАБОТА: Линейный дискриминантный анализ

Описание работы соответствует функциям Excel, в R-project используйте аналогичные функции.

I. Задание. Результаты наблюдений составляют два набора строк, всего $n_1 + n_2$ строк в таблице. В каждой строке содержатся числовые значения k переменных: x_{1ij} , $i = 1, \dots, n_1$, $j = 1, \dots, k$ – первый набор и x_{2ij} , $i = 1, \dots, n_2$, $j = 1, \dots, k$ – второй набор.

$$X_1 = \begin{pmatrix} x_{111} \dots x_{11k} \\ x_{121} \dots x_{12k} \\ \dots \dots \dots \\ x_{1n_1 1} \dots x_{1n_1 k} \end{pmatrix} \quad \text{и} \quad X_2 = \begin{pmatrix} x_{211} \dots x_{21k} \\ x_{221} \dots x_{22k} \\ \dots \dots \dots \\ x_{2n_2 1} \dots x_{2n_2 k} \end{pmatrix}.$$

Зададим две матрицы:

Процедура линейного дискриминантного анализа заключается в нахождении такой линейной комбинации переменных x_1, x_2, \dots, x_k , чтобы как можно лучше разделить два набора данных.

Линейная дискриминантная функция определяется выражением

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk}, \quad i = 1, 2, \quad t = 1, \dots, n_i. \quad (1)$$

Центры двух множеств (векторы средних) задаются формулами:

$$\bar{X}_1 = (\bar{X}_{11}, \bar{X}_{12}, \dots, \bar{X}_{1k})^T, \quad \bar{X}_2 = (\bar{X}_{21}, \bar{X}_{22}, \dots, \bar{X}_{2k})^T,$$

$$\bar{X}_{ij} = \frac{1}{n_i} \sum_{t=1}^{n_i} x_{itj} \quad \text{— среднее по значениям переменной } j \text{ для набора данных } i.$$

$$\bar{Y}_i = \frac{1}{n} \sum_{t=1}^{n_i} y_{it}, \quad i = 1, 2.$$

Сумма квадратов отклонений от среднего является мерой вариации переменной. Тогда внутригрупповой вариацией данных является величина:

$$\begin{aligned} \sum_{i=1}^2 \sum_{t=1}^{n_i} (y_{it} - \bar{Y}_i)^2 &= \sum_{i=1}^2 \sum_{t=1}^{n_i} [\beta_1 (x_{it1} - \bar{X}_{i1}) + \beta_2 (x_{it2} - \bar{X}_{i2}) + \dots + \beta_k (x_{itk} - \bar{X}_{ik})]^2 = \\ &= \beta^T (\tilde{X}_1^T \cdot \tilde{X}_1 + \tilde{X}_2^T \cdot \tilde{X}_2) \beta, \end{aligned}$$

$$\tilde{X}_1 = \begin{pmatrix} (x_{111} - \bar{X}_{11}) \dots (x_{11k} - \bar{X}_{1k}) \\ (x_{121} - \bar{X}_{11}) \dots (x_{12k} - \bar{X}_{1k}) \\ \dots \dots \dots \\ (x_{1n_1 1} - \bar{X}_{11}) \dots (x_{1n_1 k} - \bar{X}_{1k}) \end{pmatrix}, \quad \tilde{X}_2 = \begin{pmatrix} (x_{211} - \bar{X}_{21}) \dots (x_{21k} - \bar{X}_{2k}) \\ (x_{221} - \bar{X}_{21}) \dots (x_{22k} - \bar{X}_{2k}) \\ \dots \dots \dots \\ (x_{2n_2 1} - \bar{X}_{21}) \dots (x_{2n_2 k} - \bar{X}_{2k}) \end{pmatrix},$$

где

Объединенной ковариационной матрицей S называют

$$S = \frac{1}{n_1 + n_2 - 2} (\tilde{X}_1^T \cdot \tilde{X}_1 + \tilde{X}_2^T \cdot \tilde{X}_2)$$

Межгрупповую вариацию можно записать в виде:

$$(\bar{Y}_1 - \bar{Y}_2)^2 = (\beta^T \cdot \tilde{X}_1^T - \beta^T \cdot \tilde{X}_2^T)^2 = \beta^T (\tilde{X}_1^T - \tilde{X}_2^T) (\tilde{X}_1^T - \tilde{X}_2^T) \beta$$

Один из методов поиска оптимальной дискриминантной функции заключается в

максимизации отношения:

$$\lambda = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\sum_{i=1}^2 \sum_{t=1}^{n_i} (Y_{it} - \bar{Y}_i)^2} \quad \text{или} \quad l = \frac{\beta^T (\bar{X}_1 - \bar{X}_2) \cdot (\bar{X}_1 - \bar{X}_2)^T \beta}{\beta^T S \beta}, \quad l$$

отличается от λ только числовым коэффициентом. Максимизация введенных величин означает максимизацию отношения межгрупповой вариации к внутригрупповой вариации.

Будем считать, что параметры $\beta_1, \beta_2, \dots, \beta_k$ выбираются из условия максимума отношения λ или l . Вычислив производные по неизвестным параметрам и приравняв их к нулю, получим равенство, которому должны удовлетворять неизвестные параметры:

$$\hat{\beta} = c \cdot S^{-1} (\bar{X}_1 - \bar{X}_2)$$

Подставив $\hat{\beta}$ в l , получим: $\hat{l} = (\bar{X}_1 - \bar{X}_2)^T S^{-1} (\bar{X}_1 - \bar{X}_2) = D^2$. Величину D^2 называют обобщенным расстоянием или расстоянием Махаланобиса.

Константу c можно выбрать произвольным образом, так как она входит в качестве множителя в дискриминантную функцию. Значение коэффициента c не влияет на возможность дискриминации между группами. Коэффициент часто задают равным единице или выбирают так, чтобы $\hat{\beta}_1 = 1$.

Для задания правила классификации необходимо выбрать пороговое значение Y^* , если цена ошибочной классификации одинакова для обеих групп и если вероятность принадлежности наблюдения к каждой из групп одинакова, то

$$Y^* = \frac{1}{2} (\bar{Y}_1 + \bar{Y}_2) = \frac{1}{2} \hat{\beta}^T (\bar{X}_1 + \bar{X}_2)$$

Правило классификации:

отнести наблюдение $x = (x_1, x_2, \dots, x_k)^T$ к группе I, если $y(x) = \hat{\beta}^T x \geq Y^*$

отнести наблюдение $x = (x_1, x_2, \dots, x_k)^T$ к группе II, если $y(x) = \hat{\beta}^T x < Y^*$

Замечание. Для проверки эффективности можно применить правило классификации к исходным данным и подсчитать относительную долю ошибок классификации.

II. Задание.

1) Проведите линейный дискриминантный анализ данных из таблицы

	Потребительские		Средства производства
--	-----------------	--	-----------------------

	товары								
	x_1	x_2	x_3	x_4		x_1	x_2	x_3	x_4
Рис	72	50	8	0,5	Бензин	57	57	12,5	0,9
Чай	66,5	48	15	1	Свинец	100	54	17	0,5
Сахар	54	57	14	1	Чугун	100	32	16,5	0,7
Мука	67	60	15	0,9	Медь	96,5	65	20,5	0,9
Кофе	44	57	14	0,3	Цинк	79	51	18	0,9
Картофель	41	52	18	1,9	Олово	78	53	18	1,2
Масло	34,5	50	4	0,5	Каучук	48	40	21	1,6
Сыр	34,5	46	8,5	1	Ртуть	155	44	20,5	1,4
Говядина	24	54	3	1,2	Медный лист	84	64	13	0,8
					Железо	105	35	17	1,8

В таблице приведены данные по циклическому поведению цен средств производства и потребительских товаров в соответствии с оптовыми ценами в Англии за период 1860-1913 гг.:

x_1 – средняя длительность цикла в месяцах;

x_2 – средняя доля периода повышения цен по отношению к длительности всего цикла;

x_3 – средняя амплитуда за цикл в процентах от тренда;

x_4 – среднемесячное изменение цен в течение цикла.

2) Найдите в интернете другие примеры по теме.

3) Сделайте выводы экономического характера по пунктам 1 и 2.

4) Подготовьте отчет.

3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса

Для оценки обучающимися содержания и качества учебного процесса используется анкета-отзыв установленная локальными актами СПбГУ.

3.2. Кадровое обеспечение

3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий

К преподаванию привлекаются преподаватели, имеющие ученую степень физико-математических или технических наук, а также главные и ведущие специалисты в этой области. Допускается проведение занятий обучающимся в аспирантуре (под руководством

научного руководителя) для прохождения педагогической практики.

3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом

Не требуется.

3.3. Материально-техническое обеспечение

3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий

Компьютерный класс с количеством рабочих мест соответствующим количеству обучающихся с учетом рабочего места преподавателя, мультимедийный проектор, доска.

3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

Универсальные компьютеры, объединенные в локальную сеть, мультимедийное оборудование (проектор, экран). Системное программное обеспечение общего назначения (MS Windows любой современной версии).

3.3.3 Характеристики специализированного оборудования

Отсутствуют.

3.3.4 Характеристики специализированного программного обеспечения

Microsoft Excel, Python (2.7 или старше), программная среда вычислений R (3.4 или старше).

3.3.5 Перечень и объёмы требуемых расходных материалов

Не требуются.

3.4. Информационное обеспечение

3.4.1 Список обязательной литературы

1. Буре В. М., Парилина Е. М., Сedaков А. А. Методы прикладной статистики в R и Excel. / учеб. пособие.– Лань, 2016.– 152 с.
2. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика: Учебник.– Лань, 2013.– 416 с.
3. Буре В. М., Евсеев Е. А. Основы эконометрики: Учеб. Пособие.– СПб.: Изд-во Санкт-Петербургского университета, 2004.– 72 с.
4. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс: учебник. – М.: Дело, 2007.– 504 с.
5. Буре В.М., Парилина Е.М., Сedaков А.А., Шевкопляс Е.В. Прикладная статистика в R, STATISTICA и Excel. Описательная статистика. Оценка параметров. Статистические критерии/ учеб. пособие.– СПб., 2011.– 104 с.

3.4.2 Список дополнительной литературы

1. Евсеев Е.А., Буре В.М. Эконометрика: Учеб. Пособие.– Москва, 2-е издание, исправленное и дополненное, издательство «Юрайт», 2017.
2. Айвазян С. А. Методы эконометрики: учебник.– М.: Магистр: ИНФРА-М, 2015.– 512 с.
3. Айвазян С. А., Мхитарян В. С. Прикладная статистика в задачах и упражнениях: учебник для вузов.– М.: ЮНИТИ-ДАНА, 2001.– 270 с.
3. Многомерные статистические методы для экономики: учебное пособие / Б. Болч, К. Дж. Хуань; пер. с англ. А. Д. Плитман; ред., авт. предисл. С. А. Айвазян.– М.: Статистика,

3.4.3 Перечень иных информационных источников

Нет

Раздел 4. Разработчики программы

Буре Владимир Мансурович д.т.н. доцентпрофессор v.bure@spbu.ru