

Многомерный анализ данных, 4/8, 2010/2011)

Голяндина Н.Э., примерные темы к экзамену

1. Меры зависимости. Меры зависимости двух признаков – коэффициент корреляции Пирсона, ранговые коэффициенты корреляции, корреляционное отношение. Соотношение между ними. Гипотеза о независимости, критерий хи-квадрат. Частный и множественный коэффициенты корреляции. Парная регрессия.
2. Квадратичные формы от нормального распределения, теорема Кочрена. На примере проверки гипотезы о значении математического ожидания уметь рассказывать, откуда берутся распределения, число степеней свободы, точный критерий, асимптотический критерий.
3. Сравнение двух распределений. Сравнение независимых выборок: t-критерий (в разных моделях), непараметрические критерии Манна-Уитни, Колмогорова-Смирнова. Сравнение двух зависимых выборок: t-критерий, непараметрические критерии знаков и парный тест Вилкоксона.
4. Однофакторный дисперсионный анализ, множественные сравнения. Поправка Бонферрони, студентизированный размах, LSD, HSD и пр.
5. Многомерное нормальное распределение Расстояние Махаланобиса, его распределение. Распределения Хотеллинга, Уишарта, Wilks' Lambda distribution. (это вопрос раскидан по конспекту)
6. Множественная линейная регрессия. Выражение для коэффициентов регрессии. Разбиение регрессоров. Теорема Маркова, ОМП. Значимость регрессии, доверительные интервалы для предсказания, доверительные интервалы для коэффициентов регрессии. Распределение вектора из коэффициентов регрессии. Сравнение одномерных и двумерных доверительных областей в случае двух объясняющих переменных. Регрессия с ограничениями. Проверка гипотез о выполнении линейных условий на параметры. Пошаговый регрессионный анализ. Выделяющиеся наблюдения в регрессии (анализ остатков, расстояния Махаланобиса и Кука, удаленные остатки).
7. Параметрические многомерные гипотезы. Гипотеза о значении многомерного среднего (статистика Хотеллинга). Доверительные области для параметров с использованием нормально-распределенных несмещенных оценок на основе расстояния Махаланобиса, поиск outliers. Гипотеза о равенстве многомерных средних (независимые выборки). Гипотеза о равенстве средних зависимых признаков (repeated measures), контрасты. Гипотеза о равенстве ковариационных матриц, M-статистика Бокса
8. Многомерный дисперсионный анализ (MANOVA). Критерии лямбда Уилкса, Пиллая, обобщенный Хотеллинга, критерий максимального корня Роя.
9. Дискриминантный анализ. Общий подход к решению задачи классификации. Классифицирующие функции, априорные и апостериорные вероятности. Матрица классификации, cross-validation. Нормальная модель, одинаковые и различные ковариационные матрицы. Линейный дискриминантный анализ. Значимость дискриминации, связь с MANOVA. Канонические переменные. Значимость канонических переменных. Факторные веса. Пошаговый дискриминантный анализ.
10. Канонические корреляции. Canonical roots. Извлеченная дисперсия.
11. **Разложение дисперсии (ковариационной матрицы). Связь между дискриминантным анализом, многомерной множественной регрессией, корреляционным анализом и MANOVA.**
12. Кластерный анализ. Расстояния. Меры связи. Иерархическое дерево. Метод К-средних.