

Программа дисциплины «Анализ данных для лингвистов»

Утверждена

Академическим советом ОП

Протокол № 15 от «28» июня 2018 г.

Автор	Г. А. Мороз
Число кредитов	3
Контактная работа (час.)	36
Самостоятельная работа (час.)	78
Курс	3, 4
Формат изучения дисциплины	без использования онлайн-курса

I. ЦЕЛЬ И РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Задачей курса «Анализ данных для лингвистов» является продолжение знакомства с различными методами анализа данных. Курс разделен на несколько тематических блоков: первый связан с применением байесовских статистических методов (байесовский апдейт, байесовский доверительный интервал, байесовский фактор, байесовская эмпирическая оценка), второй связан с методами уменьшения размерности (PCA, LDA, CA, MCA), третий блок связан с методами кластеризации (k-means, иерархическая кластеризация, смешанные модели) и последний блок будет посвящен проблемам применения регрессионного анализа (регрессия со смешанными эффектами, обобщенная аддитивная модель).

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

1. работа с разными распределениями
2. байесовские статистические методы
3. кластеризация и смешанные модели
4. уменьшение размерности
5. проблемы применения регрессионного анализа

III. ОЦЕНИВАНИЕ

Итоговая оценка складывается из следующих компонентов:

- текущий контроль домашнего чтения с весом 40%,
- экзамен с весом 60%, в том числе дополнительная задача на оценку выше 8.

Текущий контроль складывается из оценок за контрольные работы и зачета в конце четвертого модуля. Результирующая оценка выставляется по следующей формуле:

$$O_{\text{итоговый}} = 0.6 \times O_{\text{экзамен}} + 0.4 \times O_{\text{текущий контроль}}$$

Способ округления накопленной оценки промежуточного (итогового) контроля в форме зачета: арифметический.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

- Посчитайте значение правдоподобия распределения $\mathcal{N}(\mu = 22, \sigma^2 = 6)$ для двух наблюдений 57 и 43.
- Проведите байесовский апдейт данных $Beta(11, 34)$, используя априорное распределение $Beta(23, 45)$, и посчитайте симметричный 95% байесовский доверительный интервал и 95% интервал максимальной плотности.

V. РЕСУРСЫ

- 5.1 Основная литература**
- 5.2 Дополнительная литература**
- 5.3 Программное обеспечение**
- 5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**
- 5.5 Материально-техническое обеспечение дисциплины**