

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ:

Прото ректора

М.В. Румянцев

« _____ » _____ 2020 г.

**ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ**

«Анализ данных в R»

Красноярск 2020

I. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Аннотация программы

Умение работать с данными определяет глубину конечного результата при изучении того или иного феномена, с которым имеет дело исследователь. Программа R позволяет проводить как базовый, так и углубленный анализ данных, не привлекая сложных предварительных вычислений. Эта особенность является преимуществом для исследователей нематематических специальностей, которые оперируют данными из своей области, но для их обработки не обладают нужными навыками. Представленный курс даёт слушателям возможность получить опыт анализа с помощью простых команд, встроенных в пакеты R, тем самым получив от своей работы большую отдачу.

1.2. Цель программы

Программа «Анализ данных в R» направлена на формирование базовых навыков и формирование базовых компетенций работы с научными данными различного происхождения. Курс начинается с краткого введения в систему R и основ статистики. Основное время в обучении отведено на методы анализа на примерах различных типов данных.

1.3. Компетенции (трудовые функции) в соответствии с Профессиональным стандартом (формирование новых или совершенствование имеющихся)

Профессиональный стандарт «Педагог профессионального обучения, профессионального образования и дополнительного профессионального образования» Приказ Минтруда России N 608н от 08.09.2015:

- Н/02.6 Организация научно-исследовательской, проектной, учебно-профессиональной и иной деятельности обучающихся по программам бакалавриата и(или) ДПП под руководством специалиста более высокой квалификации.
- I/03.7 Руководство научно-исследовательской, проектной, учебно-профессиональной и иной деятельностью обучающихся по программам бакалавриата, специалитета, магистратуры и(или) ДПП.
- J/03.8 Руководство подготовкой аспирантов (адъюнктов) по индивидуальному учебному плану.

1.4. Планируемые результаты обучения

РО 1. Базовое понимание научного метода в исследовании
РО 2. Владение базовыми навыками работы в системе R
РО 3. Умение предварительной обработки данных: РО 3.1 Анализировать эмпирические и теоретические распределения. РО 3.2 Выбирать трансформацию сырых данных
РО 4. Умение анализировать статистические гипотезы
РО 5. Умение выявлять зависимости в данных: РО 5.1 Проводить корреляционный анализ. РО 5.2 Выполнять регрессионный анализ. РО 5.3 Выявлять значимые для результата факторы. РО 5.4 Применять простые математические модели. РО 5.5 Применять интервальные оценки
РО 6. Владение базовым навыком классификационного анализа

1.5. Категория слушателей: преподаватели высших учебных заведений, аспиранты, научные работники

1.6. Требования к уровню подготовки поступающего на обучение: к обучению допускаются любые слушатели независимо от их базового образования и профессиональной деятельности.

1.7. Продолжительность обучения: 36 часов.

1.8. Форма обучения: дистанционная.

1.9. Требования к материально-техническому обеспечению, необходимое для реализации дополнительной профессиональной программы повышения квалификации (требования к аудитории, компьютерному классу, программному обеспечению).

Для реализации программы необходим компьютерный класс с установленными пакетами R и R studio (свободное программное обеспечение).

1.10. Документ об образовании: удостоверение о повышении квалификации установленного образца

II. ОСНОВНОЕ СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Учебно-тематический план

№ п/п	Наименование и содержание разделов и тем программы	Всего часов	В том числе:		Использование средств ЭО и ДОТ	Результаты обучения
			Аудиторные	Дистанционные		
1.	Научный метод в интерпретации данных	2	-	2	Электронный курс в системе LMS Moodle e.sfu-kras.ru	РО 1
1.1	Роль статистического анализа	0,5	-	0,5		РО 1
1.2	Генеральная совокупность и выборка. Репрезентативность выборок	0,5	-	0,5		РО 1
1.3	Ошибки, возникающие при измерениях, повторяемость результатов	0,5	-	0,5		РО 1
1.4	Модели	0,5	-	0,5		РО 1
2.	Основы работы с R	4	-	4		РО 2
2.1	Объекты R, базовые действия с данными, data.frame	3	-	3		РО 2
2.2	Установка пакетов и графики в R	1	-	1		РО 2
3.	Теоретические и экспериментальные распределения	6	-	6		РО 3
3.1	Типы данных, трансформация данных	2	-	2		РО 3.2
3.2	Параметры распределения данных, аппроксимация эмпирического распределения теоретическим	2	-	2		РО 3.1
3.3	Графический анализ распределения данных	2	-	2		РО 3.1
4.	Постановка и проверка гипотез	6	-	6		РО 4
4.1	Формулировка нулевой и альтернативной гипотез	1	-	1	РО 4	

№ п/п	Наименование и содержание разделов и тем программы	Всего часов	В том числе:		Использование средств ЭО и ДОТ	Результаты обучения
			Аудиторные	Дистанционные		
4.2	Статистические распределения и p-value	1	-	1		РО 4
4.3	Тестирование гипотез для одной выборки	2	-	2		РО 4
4.4	Тестирование гипотез при сравнении выборок	2	-	2		РО 4
5.	Выявление зависимостей	8	-	8		РО 5
5.1	Корреляция и ковариация	1	-	1		РО 5.1
5.2	Линейная и множественная линейная регрессия	2	-	2		РО 5.2
5.3	Нелинейная регрессия	1	-	1		РО 5.2
5.4	ANOVA и факторный анализ	2	-	2		РО 5.3
5.5	Типичные модели	2	-	2		РО 5.4
6.	Классификационный анализ	4	-	4		РО 5.4
6.1	Метод главных компонент и факторный анализ	1	-	1		РО 5.4
6.2	Иерархическая кластеризация	2	-	2		РО 5.4
7.	Оформление результатов в пакете ggplot	4	-	4		РО 6
8.	Итоговый контроль	2	-	2		
	ИТОГО	36	-	36		

2.2. План учебной деятельности

Результаты обучения	Учебные действия/формы текущего контроля	Используемые ресурсы/инструменты/технологии
РО 1. Базовое понимание научного метода в исследовании	Структурирование знаний	Лекция, литература, статьи, электронный курс
РО 2. Владение базовыми навыками работы в системе R	Практические действия в системе R. Задание	Лекция, электронный курс
РО 3. Умение предварительной обработки данных	Практические действия в системе R. Задание	Лекция, электронный курс
РО 3.1 Анализировать эмпирические и теоретические распределения	Практические действия в системе R	Лекция, электронный курс
РО 3.2 Выбирать трансформацию сырых данных	Практические действия в системе R	Лекция, электронный курс
РО 4. Умение анализировать статистические гипотезы	Практические действия в системе R. Задание	Лекция, электронный курс
РО 5. Умение выявлять зависимости в данных	Практические действия в системе R. Задание	Лекция, электронный курс
РО 5.1 Проводить корреляционный анализ	Практические действия в системе R	Лекция, электронный курс
РО 5.2 Выполнять регрессионный анализ	Практические действия в системе R	Лекция, электронный курс
РО 5.3 Выявлять значимые для результата факторы	Практические действия в системе R	Лекция, электронный курс

Результаты обучения	Учебные действия/формы текущего контроля	Используемые ресурсы/ инструменты/технологии
РО 5.4 Применять простые математические модели	Практические действия в системе R	Лекция, электронный курс
РО 5.5 Применять интервальные оценки	Практические действия в системе R	Лекция, электронный курс
РО 6. Владение базовым навыком классификационного анализа	Практические действия в системе R. Задание	Лекция, электронный курс

2.3. Виды и содержание самостоятельной работы

Самостоятельная работа проводится в виде вычислений, применяемых к предоставленному набору данных, который выдается индивидуально каждому слушателю. Приветствуется наличие у слушателя данных своих экспериментов. Итоговый файл с результатами прикрепляется в соответствующий раздел курса в системе LMS Moodle СФУ.

III. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

3.1. Учебно-методическое обеспечение, в т.ч. электронные ресурсы в корпоративной сети СФУ и сети Интернет

1. Bada O., Liebl D. phtt: Panel Data Analysis with Heterogeneous Time Trends // R package version. – 2012. – Т. 3. – № 2.
2. Barnett A. G., Baker P., Dobson A. Analysing seasonal data // R Journal. – 2012. – Т. 4. – № 1. – С. 5-10.
3. Benakli N. et al. Introducing computational thinking through hands-on projects using R with applications to calculus, probability and data analysis // International Journal of Mathematical Education in Science and Technology. – 2017. – Т. 48. – № 3. – С. 393–427.
4. Bessant C., Oakley D., Shadforth I. Building Bioinformatics Solutions. – OUP Oxford, 2014.
5. Bohn A. et al. Content-based social network analysis of mailing lists // The R Journal. – 2011. – Т. 3. – № 1. – С. 11–18.
6. Fay M.P. Two-sided exact tests and matching confidence intervals for discrete data // R journal. – 2010. – Т. 2. – № 1. – С. 53–58.
7. Fox J., Friendly M., Weisberg S. Hypothesis tests for multivariate linear models using the car package // The R Journal. – 2013. – Т. 5. – № 1. – С. 39–52.
8. Gesmann M., de Castillo D. Using the Google visualisation API with R // The R Journal. – 2011. – Т. 3. – № 2. – С. 40-44.
9. Peterlin P. Using R for data analysis and graphing in an introductory physics laboratory // arXiv preprint arXiv:0911.4570. – 2009.
10. Ruppert D. Statistics and data analysis for financial engineering. – New York: Springer, 2011. – Т. 13.
11. Thiem A., Dusa A. QCA: A package for qualitative comparative analysis // The R Journal. – 2013. – Т. 5. – № 1. – С. 87–97.
12. Schandry N.A practical guide to visualization and statistical analysis of R. solanacearum infection data using R // Frontiers in plant science. – 2017. – Т. 8. – С. 623.
13. Staniak M., Biecek P. The Landscape of R Packages for Automated Exploratory Data Analysis // arXiv preprint arXiv:1904.02101. – 2019.
14. Егошин В.Л. и др. Основы работы в программной среде R при анализе исследовательских данных // Экология человека. – 2018. – № 7.

3.2. Информационное обеспечение (информационные обучающие системы, системы вебинаров, сетевые ресурсы хостинга видео, изображений, файлов, презентаций, программное обеспечение и др.).

Информационная обучающая система СФУ «eКурсы» <https://e.sfu-kras.ru/>, видеохостинг Youtube, свободно распространяемое программное обеспечение R и R studio.

IV. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

4.1. Формы аттестации, оценочные материалы, методические материалы

Аттестация проходит заочно. Индивидуальное выданное задание прикрепляется в виде R-кода в соответствующий раздел eКурса. В качестве основы для аттестационного задания используется искусственно сгенерированная база данных или данные, предоставляемые самим обучающимся (достаточность базы данных для итоговой аттестации оценивается преподавателем).

4.2. Требования и содержание итоговой аттестации

Основанием для аттестации является выполнение итоговой аттестационной работы. Итоговая аттестация осуществляется преподавателями программы на основе двухбалльной («удовлетворительно» — «зачтено», «неудовлетворительно» — «не зачтено») системы оценок. Слушатель считается аттестованным, если имеет положительную оценку.

Программу составил:
к.ф.-м.н.



Шуваев А.Н.

Руководитель программы:
к.ф.-м.н.



Шуваев А.Н.