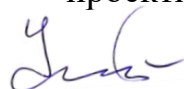


**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Утверждаю:
Зав. кафедрой компьютерных
интеллектуальных технологий
проектирования


_____ М.И. ЧИЖОВ
«21» декабря 2021 г.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

«Язык R и базовая статистика»

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Искусственный интеллект


Квалификация выпускника магистр

Нормативный период обучения 2 года / 2 года и 5 м.

Форма обучения очная / заочная

Год начала подготовки 2022

Разработчик



П.Ю. Гусев

Воронеж – 2021

Процесс изучения дисциплины «Язык R и базовая статистика» направлен на формирование следующих компетенций:

УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

ОПК-9 - Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта

ОПК-10 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований

Перечень планируемых результатов обучения и показателей оценивания сформированности компетенций на этапе промежуточной аттестации

№ п/п	Компетенция	Результаты обучения, характеризующие сформированность компетенции	Тип ОМ	Показатели оценивания
1	УК-1	Знать методы описания исходных данных для статистического анализа	Вопросы (тест) к зачету	Полнота знаний
		Уметь применять методы базовой статистики	Стандартные задания	Наличие умений
		Владеть навыками формализации задач для статистического анализа	Прикладные задания	Наличие навыков
2	ОПК-9	Знать базовые конструкции языка R	Вопросы (тест) к зачету	Полнота знаний
		Уметь реализовывать алгоритмы обработки данных на языке R	Стандартные задания	Наличие умений
		Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта	Прикладные задания	Наличие навыков
3	ОПК-10	Знать сквозные цифровые технологии искусственного интеллекта; современный опыт применения систем искусственного интеллекта	Вопросы (тест) к зачету	Полнота знаний
		Уметь проводить сравнительный анализ методов и инструментальных средств для решения задач искусственного интеллекта	Стандартные задания	Наличие умений
		Владеть навыками определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных	Прикладные задания	Наличие навыков

ОПИСАНИЕ ПОКАЗАТЕЛЕЙ, КРИТЕРИЕВ И ШКАЛ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА ЭТАПЕ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Показатели оценивания компетенций	Шкала и критерии оценки уровня сформированности компетенции			
	Неудовлетворительный	Минимально допустимый (пороговый)	Средний	Высокий
Полнота знаний	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущены не грубые ошибки.	Уровень знаний в объёме, соответствующем программе подготовки. Допущены некоторые погрешности.	Уровень знаний в объёме, соответствующем программе подготовки
Наличие умений	При выполнении стандартных заданий не продемонстрированы основные умения. Имели место грубые ошибки.	Продемонстрированы основные умения. Выполнены типовые задания с не грубыми ошибками. Выполнены все задания, но не в полном объёме (отсутствуют пояснения, неполные выводы)	Продемонстрированы все основные умения. Выполнены все основные задания с некоторыми погрешностями. Выполнены все задания в полном объёме, но некоторые с недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Задания выполнены в полном объёме без недочетов.
Наличие навыков (владение опытом)	При выполнении стандартных заданий не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для выполнения стандартных заданий с некоторыми недочетами.	Продемонстрированы базовые навыки при выполнении стандартных заданий с некоторыми недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Продемонстрирован творческий подход к решению нестандартных задач.
Характеристика сформированности компетенции	Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач. Требуется повторное обучение.	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству профессиональных задач.	Сформированность компетенций в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных профессиональных задач.	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных профессиональных задач.

ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ МАТЕРИАЛОВ

Вопросы (тестовые задания) для оценки результатов обучения, характеризующих сформированность компетенций

<i>УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий</i>	
1.	Идеология R. Структуры данных (векторы, матрицы, массивы данных, таблицы данных, факторы, списки).
2.	Загрузка данных. Пакеты. Числовые и текстовые функции. Циклы и условия.
3.	Пользовательские функции
4.	Визуализация данных. Диаграммы. Гистограммы.
5.	Визуализация данных Диаграммы плотности распределений. Диаграммы размахов. Диаграммы рассеяния
6.	Оценки центрального положения и вариабельности
7.	Исследование распределений данных
8.	Корреляция. Тест Стьюдента
<i>ОПК-9 - Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта</i>	
1.	Непараметрические тесты межгрупповых различий
2.	Визуализация групповых различий
3.	А/В-тестирование. Проверка статистических гипотез
4.	Статистическая значимость и р-значение
5.	Проверка на основе t-статистики
6.	Дисперсионный анализ
7.	Проверка на основе статистики хи-квадрат
8.	Простая линейная регрессия
<i>ОПК-10 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований</i>	
1.	Множественная линейная регрессия
2.	Предсказания на основе регрессии
3.	Оценка качества предсказания. Перекрестная проверка
4.	Нелинейная регрессия
5.	Наивный байесовский алгоритм
6.	Логистическая регрессия. Предсказание значений в логистической регрессии. Оценка качества предсказания
7.	Алгоритм k ближайших соседей
8.	Древовидные алгоритмы (дерево решений, случайный лес)
9.	Алгоритм бустинга
10.	Переобучение. Подбор гиперпараметров алгоритма

Практические задания для оценки результатов обучения, характеризующих сформированность компетенций

<i>УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий</i>	
1.	Какая из следующих функций используется для просмотра набора данных в формате электронной таблицы? А) disp()

	<p>+B) View() C) seq() D) все ответы верны</p>
2.	<p>Какая из следующих команд разделит окно графика на окна 4 X 3 и графики входят в столбец окна. A) par(split=c(4,3)) B) par(mfcol=c(4,3)) C) par(mfrow=c(4,3)) D) par(col=c(4,3))</p>
3.	<p>Какая из следующих команд поможет нам переименовать второй столбец в фрейме данных с именем «table» с 'alpha' на 'beta'? A) colnames(table)[2]='beta' B) colnames(table)[which(colnames=='alpha')]='beta' C) setnames(table,'alpha','beta') D) Все ответы верны</p>
4.	<p>Какая из следующих команд удалит объект R / переменную с именем «santa» из рабочей области? A) remove(santa) B) rm(santa) +C) a) и b) D) Нет верного ответа</p>
5.	<p>«dplyr» — один из самых популярных пакетов, используемых в R для управления данными, и он содержит 5 основных функций для обработки данных. Что из следующего не является одной из основных функций пакета dplyr? A) select() B) filter() C) arrange() +D) summary()</p>
6.	<p>Каковы будут выходные данные следующих команд? A<-paste("alpha","beta","gamma",sep=" ") B<-paste("phi","theta","zeta",sep="") parts<-strsplit(c(A,B),split=" ") A) alpha +B) beta C) gamma D) phi E) theta F) zeta</p>
7.	<p>Что будет выведено следующей командой? grepl("neeraj",c("dheeraj","Neeraj","neeraj","is","NEERAJ")) A) [FALSE TRUE TRUE FALSE TRUE] B) [FALSE TRUE TRUE FALSE FALSE] +C) [FALSE FALSE TRUE FALSE FALSE] D) Все ответы не верны</p>
8.	<p>Что будет выведено следующей командой? A<-c("I can use because thrice in a sentence because because is a special word.")</p>

	+A) gsub("because","since",A) B) sub("because","since",A) C) regex("because","since",A) D) Все ответы не верны
9.	Какая из следующих команд поможет нам удалить повторяющиеся строки на основе обоих столбцов? A) df[!duplicated(df),] B) unique(df) C) dplyr::distinct(df) +D) Все ответы верны
10.	Группировка является важным действием в Data Analytics и помогает нам обнаружить некоторые интересные тенденции, которые могут быть не видны в необработанных данных. Предположим, у вас есть набор данных, созданный с помощью следующих строк кода. <pre>table<-data.table(foo=c("A","B","A","A","B","A"),bar=1:6)</pre> Какая из следующих команд поможет нам вычислить среднее значение bar, сгруппированного по переменной foo? A) aggregate(bar~foo,table,mean) B) table::df[,mean(bar),by=foo] C) dplyr::table%>%group_by(foo)%>%summarize(mean=mean(bar)) +D) Все ответы верны

ОПК-9 - Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта

1.	Ниже приведен набор данных, необходимо построить гистограмму для переменной «Значение». <table border="1"> <thead> <tr> <th>Parameter</th> <th>State</th> <th>Value</th> <th>Dependents</th> </tr> </thead> <tbody> <tr><td>Alpha</td><td>Active</td><td>50</td><td>2</td></tr> <tr><td>Beta</td><td>Active</td><td>45</td><td>5</td></tr> <tr><td>Beta</td><td>Passive</td><td>25</td><td>0</td></tr> <tr><td>Alpha</td><td>Passive</td><td>21</td><td>0</td></tr> <tr><td>Alpha</td><td>Passive</td><td>26</td><td>1</td></tr> <tr><td>Beta</td><td>Active</td><td>30</td><td>2</td></tr> <tr><td>Beta</td><td>Passive</td><td>18</td><td>0</td></tr> </tbody> </table> Какая из следующих команд выполнит эту задачу? A) hist(dataframed\$Value) B) ggplot2::qplot(dataframed\$Value,geom="Histogram") C)ggplot2::ggplot(data=dataframed,aes(dataframe\$Value))+geom_histogram() D) Все ответы верны	Parameter	State	Value	Dependents	Alpha	Active	50	2	Beta	Active	45	5	Beta	Passive	25	0	Alpha	Passive	21	0	Alpha	Passive	26	1	Beta	Active	30	2	Beta	Passive	18	0
Parameter	State	Value	Dependents																														
Alpha	Active	50	2																														
Beta	Active	45	5																														
Beta	Passive	25	0																														
Alpha	Passive	21	0																														
Alpha	Passive	26	1																														
Beta	Active	30	2																														
Beta	Passive	18	0																														
2.	<table border="1"> <thead> <tr> <th>Parameter</th> <th>State</th> <th>Value</th> <th>Usage</th> </tr> </thead> <tbody> <tr><td>Alpha</td><td>Active</td><td>50</td><td>0</td></tr> <tr><td>Beta</td><td>Active</td><td>45</td><td>1</td></tr> <tr><td>Beta</td><td>Passive</td><td>25</td><td>0</td></tr> <tr><td>Alpha</td><td>Passive</td><td>21</td><td>0</td></tr> <tr><td>Alpha</td><td>Passive</td><td>26</td><td>1</td></tr> <tr><td>Beta</td><td>Active</td><td>30</td><td>1</td></tr> <tr><td>Beta</td><td>Passive</td><td>18</td><td>0</td></tr> </tbody> </table>	Parameter	State	Value	Usage	Alpha	Active	50	0	Beta	Active	45	1	Beta	Passive	25	0	Alpha	Passive	21	0	Alpha	Passive	26	1	Beta	Active	30	1	Beta	Passive	18	0
Parameter	State	Value	Usage																														
Alpha	Active	50	0																														
Beta	Active	45	1																														
Beta	Passive	25	0																														
Alpha	Passive	21	0																														
Alpha	Passive	26	1																														
Beta	Active	30	1																														
Beta	Passive	18	0																														

Некоторые алгоритмы, такие как XGBOOST, работают только с числовыми данными. В этом случае категориальные переменные, присутствующие в наборе данных, сначала преобразуются в переменные DUMMY, которые представляют наличие или отсутствие уровня категориальной переменной в наборе данных. Например, после создания фиктивной переменной для функции «Параметр» набор данных выглядит так, как показано ниже.

Parameter_Alpha	Parameter_Beta	State	Value	Usage
1	0	Active	50	0
0	1	Active	45	1
0	1	Passive	25	0
1	0	Passive	21	0
1	0	Passive	26	1
0	1	Active	30	1
0	1	Passive	18	0

Какая из следующих команд позволяет достичь этого?

A) `dummies::dummy.data.frame(dataframe, names=c('Parameter'))`

B) `dataframe$Parameter_Alpha=0`

`dataframe$Gende_Beta=0`

`dataframe$Parameter_Alpha[which(dataframe$Parameter=='Alpha')]=1`

`dataframe$Parameter_Beta[which(dataframe$Parameter=='Alpha')]=0`

`dataframe$Parameter_Alpha[which(dataframe$Parameter=='Beta')]=0`

`dataframe$Parameter_Beta[which(dataframe$Parameter=='Beta')]=1`

C) `contrasts(dataframe$Parameter)`

+D) A) и B)

3.

Column1	Column2	Column3	Column4	Column5	Column6	
Name1	Alpha	12	24	54	0	Alpha
Name2	Beta	16	32	51	1	Beta
Name3	Alpha	52	104	32	0	Gamma
Name4	Beta	36	72	84	1	Delta
Name5	Beta	45	90	32	0	Phi
Name6	Alpha	12	24	12	0	Zeta
Name7	Beta	32	64	64	1	Sigma
Name8	Alpha	42	84	54	0	Mu
Name9	Alpha	56	112	31	1	Eta

Необходимо рассчитать корреляцию между «Столбцом2» и «Столбцом3» в наборе данных. Какой код из приведенных ниже достигнет цели?

A) `corr(dataframe$column2, dataframe$column3)`

B) $(\text{cov}(\text{dataframe}\$column2, \text{dataframe}\$column3)) / (\text{var}(\text{dataframe}\$column2) * \text{sd}(\text{dataframe}\$column3))$

C)

$(\text{sum}(\text{dataframe}\$Column2 * \text{dataframe}\$Column3) - (\text{sum}(\text{dataframe}\$Column2) * \text{sum}(\text{dataframe}\$Column3) / \text{nrow}(\text{dataframe}))) / (\text{sqrt}((\text{sum}(\text{dataframe}\$Column2 * \text{dataframe}\$Column2) - (\text{sum}(\text{dataframe}\$Column2)^2 / \text{nrow}(\text{dataframe}))) * (\text{sum}(\text{dataframe}\$Column3 * \text{dataframe}\$Column3) - (\text{sum}(\text{dataframe}\$Column3)^2 / \text{nrow}(\text{dataframe}))))$

D) Ни один ответ не верный

4.

Parameter	State	Value	Dependents
Alpha	Active	50	2

	<table> <tbody> <tr> <td>Beta</td> <td>Active</td> <td>45</td> <td>5</td> </tr> <tr> <td>Beta</td> <td>Passive</td> <td>25</td> <td>0</td> </tr> <tr> <td>Alpha</td> <td>Passive</td> <td>21</td> <td>0</td> </tr> <tr> <td>Alpha</td> <td>Passive</td> <td>26</td> <td>1</td> </tr> <tr> <td>Beta</td> <td>Active</td> <td>30</td> <td>2</td> </tr> </tbody> </table> <p>В переменной с именем «dataframe» ассоциирован набор данных, где первая строка представляет имя столбца. Какой набор инструкций будет выбирать только те строки, для которых параметр Alpha?</p> <p>A) subset(dataframe, Parameter='Alpha')</p> <p>B) subset(dataframe, Parameter=='Alpha')</p> <p>C) filter(dataframe,Parameter=='Alpha')</p> <p>+D) Только b или c</p> <p>E) Все ответы верны</p>	Beta	Active	45	5	Beta	Passive	25	0	Alpha	Passive	21	0	Alpha	Passive	26	1	Beta	Active	30	2
Beta	Active	45	5																		
Beta	Passive	25	0																		
Alpha	Passive	21	0																		
Alpha	Passive	26	1																		
Beta	Active	30	2																		
5.	<p>Приведенный ниже набор данных хранится в переменной с именем data.</p> <table> <tbody> <tr> <td>A</td> <td>B</td> </tr> <tr> <td>1</td> <td>Right</td> </tr> <tr> <td>2</td> <td>Wrong</td> </tr> <tr> <td>3</td> <td>Wrong</td> </tr> <tr> <td>4</td> <td>Right</td> </tr> <tr> <td>5</td> <td>Right</td> </tr> <tr> <td>6</td> <td>Wrong</td> </tr> <tr> <td>7</td> <td>Wrong</td> </tr> <tr> <td>8</td> <td>Right</td> </tr> </tbody> </table> <p>Предположим, что B — категориальная переменная, и необходимо нарисовать блок-диаграмму для каждого уровня категориального уровня. Какая из приведенных ниже команд выполнит это?</p> <p>A) boxplot(A,B,data=data)</p> <p>+B) boxplot(A~B,data=data)</p> <p>C) boxplot(A B,data=data)</p> <p>D) Все ответы не верны</p>	A	B	1	Right	2	Wrong	3	Wrong	4	Right	5	Right	6	Wrong	7	Wrong	8	Right		
A	B																				
1	Right																				
2	Wrong																				
3	Wrong																				
4	Right																				
5	Right																				
6	Wrong																				
7	Wrong																				
8	Right																				
6.	<p>Набор данных «df» содержит следующие данные:</p> <pre>Dates 2017-02-28 2017-02-27 2017-02-26 2017-02-25 2017-02-24 2017-02-23 2017-02-22 2017-02-21</pre> <p>После прочтения приведенных выше данных необходимо получить следующий вывод данных</p> <pre>Dates 28 Tuesday Feb 17 27 Monday Feb 17 26 Sunday Feb 17 25 Saturday Feb 17 24 Friday Feb 17 23 Thursday Feb 17 22 Wednesday Feb 17 21 Tuesday Feb 17</pre>																				

	<p>Какая из следующих команд даст желаемый результат?</p> <p>A) <code>format(df,"%d %A %b %y")</code> B) <code>format(df,"%D %A %b %y")</code> C) <code>format(df,"%D %a %B %Y")</code> +D) Все ответы не верны</p>																																																																																						
7.	<p>Во время выбора функций с использованием следующего набора данных (именованной таблицы) «Столбец1» и «Столбец2» оказались несущественными. Следовательно, мы не хотели бы использовать эти два критерия в нашей прогностической модели.</p> <table border="1"> <thead> <tr> <th></th> <th>Column1</th> <th>Column2</th> <th>Column3</th> <th>Column4</th> <th>Column5</th> <th>Column6</th> <th></th> </tr> </thead> <tbody> <tr> <td>Name1</td> <td>Alpha</td> <td>12</td> <td>24</td> <td>54</td> <td>0</td> <td></td> <td>Alpha</td> </tr> <tr> <td>Name2</td> <td>Beta</td> <td>16</td> <td>32</td> <td>51</td> <td>1</td> <td></td> <td>Beta</td> </tr> <tr> <td>Name3</td> <td>Alpha</td> <td>52</td> <td>104</td> <td>32</td> <td>0</td> <td></td> <td>Gamma</td> </tr> <tr> <td>Name4</td> <td>Beta</td> <td>36</td> <td>72</td> <td>84</td> <td>1</td> <td></td> <td>Delta</td> </tr> <tr> <td>Name5</td> <td>Beta</td> <td>45</td> <td>90</td> <td>32</td> <td>0</td> <td></td> <td>Phi</td> </tr> <tr> <td>Name6</td> <td>Alpha</td> <td>12</td> <td>24</td> <td>12</td> <td>0</td> <td></td> <td>Zeta</td> </tr> <tr> <td>Name7</td> <td>Beta</td> <td>32</td> <td>64</td> <td>64</td> <td>1</td> <td></td> <td>Sigma</td> </tr> <tr> <td>Name8</td> <td>Alpha</td> <td>42</td> <td>84</td> <td>54</td> <td>0</td> <td></td> <td>Mu</td> </tr> <tr> <td>Name9</td> <td>Alpha</td> <td>56</td> <td>112</td> <td>31</td> <td>1</td> <td></td> <td>Eta</td> </tr> </tbody> </table> <p>Какая из следующих команд выберет все строки от столбца 3 до столбца 6 для приведенного ниже фрейма данных с именем таблицы?</p> <p>A) <code>dplyr::select(table,Column3:Column6)</code> B) <code>table[,3:6]</code> C) <code>subset(table,select=c('Column3','Column4','Column5','Column6'))</code> +D) Все ответы верны</p>								Column1	Column2	Column3	Column4	Column5	Column6		Name1	Alpha	12	24	54	0		Alpha	Name2	Beta	16	32	51	1		Beta	Name3	Alpha	52	104	32	0		Gamma	Name4	Beta	36	72	84	1		Delta	Name5	Beta	45	90	32	0		Phi	Name6	Alpha	12	24	12	0		Zeta	Name7	Beta	32	64	64	1		Sigma	Name8	Alpha	42	84	54	0		Mu	Name9	Alpha	56	112	31	1		Eta
	Column1	Column2	Column3	Column4	Column5	Column6																																																																																	
Name1	Alpha	12	24	54	0		Alpha																																																																																
Name2	Beta	16	32	51	1		Beta																																																																																
Name3	Alpha	52	104	32	0		Gamma																																																																																
Name4	Beta	36	72	84	1		Delta																																																																																
Name5	Beta	45	90	32	0		Phi																																																																																
Name6	Alpha	12	24	12	0		Zeta																																																																																
Name7	Beta	32	64	64	1		Sigma																																																																																
Name8	Alpha	42	84	54	0		Mu																																																																																
Name9	Alpha	56	112	31	1		Eta																																																																																
8.	<table border="1"> <thead> <tr> <th></th> <th>Column1</th> <th>Column2</th> <th>Column3</th> <th>Column4</th> <th>Column5</th> <th>Column6</th> <th></th> </tr> </thead> <tbody> <tr> <td>Name1</td> <td>Alpha</td> <td>12</td> <td>24</td> <td>54</td> <td>0</td> <td></td> <td>Alpha</td> </tr> <tr> <td>Name2</td> <td>Beta</td> <td>16</td> <td>32</td> <td>51</td> <td>1</td> <td></td> <td>Beta</td> </tr> <tr> <td>Name3</td> <td>Alpha</td> <td>52</td> <td>104</td> <td>32</td> <td>0</td> <td></td> <td>Gamma</td> </tr> <tr> <td>Name4</td> <td>Beta</td> <td>36</td> <td>72</td> <td>84</td> <td>1</td> <td></td> <td>Delta</td> </tr> <tr> <td>Name5</td> <td>Beta</td> <td>45</td> <td>90</td> <td>32</td> <td>0</td> <td></td> <td>Phi</td> </tr> <tr> <td>Name6</td> <td>Alpha</td> <td>12</td> <td>24</td> <td>12</td> <td>0</td> <td></td> <td>Zeta</td> </tr> <tr> <td>Name7</td> <td>Beta</td> <td>32</td> <td>64</td> <td>64</td> <td>1</td> <td></td> <td>Sigma</td> </tr> <tr> <td>Name8</td> <td>Alpha</td> <td>42</td> <td>84</td> <td>54</td> <td>0</td> <td></td> <td>Mu</td> </tr> <tr> <td>Name9</td> <td>Alpha</td> <td>56</td> <td>112</td> <td>31</td> <td>1</td> <td></td> <td>Eta</td> </tr> </tbody> </table> <p>Какая из следующих команд выберет строки, имеющие значения «Альфа» в «Столбце 1» и значение меньше 50 в «Столбце 4»? Фрейм данных хранится в переменной с именем table.</p> <p>A) <code>dplyr::filter(table,Column1=='Alpha', Column4<50)</code> B) <code>dplyr::filter(table,Column1=='Alpha' & Column4<50)</code> +C) a) и b) D) Все ответы не верны</p>								Column1	Column2	Column3	Column4	Column5	Column6		Name1	Alpha	12	24	54	0		Alpha	Name2	Beta	16	32	51	1		Beta	Name3	Alpha	52	104	32	0		Gamma	Name4	Beta	36	72	84	1		Delta	Name5	Beta	45	90	32	0		Phi	Name6	Alpha	12	24	12	0		Zeta	Name7	Beta	32	64	64	1		Sigma	Name8	Alpha	42	84	54	0		Mu	Name9	Alpha	56	112	31	1		Eta
	Column1	Column2	Column3	Column4	Column5	Column6																																																																																	
Name1	Alpha	12	24	54	0		Alpha																																																																																
Name2	Beta	16	32	51	1		Beta																																																																																
Name3	Alpha	52	104	32	0		Gamma																																																																																
Name4	Beta	36	72	84	1		Delta																																																																																
Name5	Beta	45	90	32	0		Phi																																																																																
Name6	Alpha	12	24	12	0		Zeta																																																																																
Name7	Beta	32	64	64	1		Sigma																																																																																
Name8	Alpha	42	84	54	0		Mu																																																																																
Name9	Alpha	56	112	31	1		Eta																																																																																
9.	<table border="1"> <thead> <tr> <th></th> <th>Column1</th> <th>Column2</th> <th>Column3</th> <th>Column4</th> <th>Column5</th> <th>Column6</th> <th></th> </tr> </thead> <tbody> <tr> <td>Name1</td> <td>Alpha</td> <td>12</td> <td>24</td> <td>54</td> <td>0</td> <td></td> <td>Alpha</td> </tr> <tr> <td>Name2</td> <td>Beta</td> <td>16</td> <td>32</td> <td>51</td> <td>1</td> <td></td> <td>Beta</td> </tr> <tr> <td>Name3</td> <td>Alpha</td> <td>52</td> <td>104</td> <td>32</td> <td>0</td> <td></td> <td>Gamma</td> </tr> <tr> <td>Name4</td> <td>Beta</td> <td>36</td> <td>72</td> <td>84</td> <td>1</td> <td></td> <td>Delta</td> </tr> <tr> <td>Name5</td> <td>Beta</td> <td>45</td> <td>90</td> <td>32</td> <td>0</td> <td></td> <td>Phi</td> </tr> </tbody> </table>								Column1	Column2	Column3	Column4	Column5	Column6		Name1	Alpha	12	24	54	0		Alpha	Name2	Beta	16	32	51	1		Beta	Name3	Alpha	52	104	32	0		Gamma	Name4	Beta	36	72	84	1		Delta	Name5	Beta	45	90	32	0		Phi																																
	Column1	Column2	Column3	Column4	Column5	Column6																																																																																	
Name1	Alpha	12	24	54	0		Alpha																																																																																
Name2	Beta	16	32	51	1		Beta																																																																																
Name3	Alpha	52	104	32	0		Gamma																																																																																
Name4	Beta	36	72	84	1		Delta																																																																																
Name5	Beta	45	90	32	0		Phi																																																																																

	<p>Name6 Alpha 12 24 12 0 Zeta</p> <p>Name7 Beta 32 64 64 1 Sigma</p> <p>Name8 Alpha 42 84 54 0 Mu</p> <p>Name9 Alpha 56 112 31 1 Eta</p> <p>Какая из следующих инструкций будет сортировать фрейм данных на основе «Столбец2» в порядке возрастания и «Столбец3» в порядке убывания?</p> <p>A) <code>dplyr::arrange(table,desc(Column3),Column2)</code> B) <code>table[order(-Column3,Column2),]</code> +C) a) и b) D) Все ответы не верны</p>																																	
10	<p>Какая из следующих команд преобразует следующий набор данных с именем <code>maverick</code> в показанный внизу?</p> <p>Входящий Dataframe – “maverick”</p> <table border="1"> <thead> <tr> <th>Grade</th> <th>Male</th> <th>Female</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>10</td> <td>15</td> </tr> <tr> <td>B</td> <td>20</td> <td>15</td> </tr> <tr> <td>A</td> <td>30</td> <td>35</td> </tr> </tbody> </table> <p>Обработанный dataframe</p> <table border="1"> <thead> <tr> <th>Grade</th> <th>Sex</th> <th>Count</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Male</td> <td>10</td> </tr> <tr> <td>A</td> <td>Female</td> <td>15</td> </tr> <tr> <td>B</td> <td>Male</td> <td>30</td> </tr> <tr> <td>B</td> <td>Female</td> <td>15</td> </tr> <tr> <td>A</td> <td>Male</td> <td>30</td> </tr> <tr> <td>A</td> <td>Female</td> <td>35</td> </tr> </tbody> </table> <p>+A) <code>tidyr::Gather(maverick, Sex,Count,-Grade)</code> B) <code>tidyr::spread(maverick, Sex,Count,-Grade)</code> C) <code>tidyr::collect(maverick, Sex,Count,-Grade)</code> D) Все ответы не верны</p>	Grade	Male	Female	A	10	15	B	20	15	A	30	35	Grade	Sex	Count	A	Male	10	A	Female	15	B	Male	30	B	Female	15	A	Male	30	A	Female	35
Grade	Male	Female																																
A	10	15																																
B	20	15																																
A	30	35																																
Grade	Sex	Count																																
A	Male	10																																
A	Female	15																																
B	Male	30																																
B	Female	15																																
A	Male	30																																
A	Female	35																																
<p><i>ОПК-10 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований</i></p>																																		
1.	<p>Дана следующая функция</p> <pre>f <- function(x) { g <- function(y) { y + z } z <- 4 x + g(x) }</pre> <p>Если мы выполним следующие команды (написанные ниже), что будет на выходе?</p> <pre>z <- 10 f(4)</pre> <p>+A) 12 B) 7 C) 4 D) 16</p>																																	

2.	<p>В наборе данных ирисов есть разные виды цветов, такие как Setosa, Versicolor и Virginica, с их длиной чашелистика. Необходимо понять распределение длины чашелистиков у всех видов цветов. Один из способов сделать это — визуализировать это отношение с помощью графика, показанного ниже.</p> <p>Какую функцию можно использовать для построения графика, показанного выше?</p> <p>A) xuplot() +B) stripplot() C) barchart() D) bwplot()</p>
3.	<p>Формат файла Excel является одним из наиболее распространенных форматов, используемых для хранения наборов данных. Ниже представлен файл excel, в котором были введены данные на третьем листе.</p> <pre>Alpha 125.5 0 Beta 235.6 1 Beta 212.03 0 Beta 211.30 0 Alpha 265.46 1 Dataframe.xlsx</pre> <p>4) Какой из следующих кодов будет считывать вышеуказанные данные на третьем листе в кадр данных в R?</p> <p>A) <code>Openxlsx::read.xlsx("Dataframe.xlsx",sheet=3,colNames=FALSE)</code> B) <code>Xlsx::read.xlsx("Dataframe.xlsx",sheetIndex=3,header=FALSE)</code> C) <code>XLCConnect::readWorksheetFromFile("Dataframe.xlsx",sheet=3,header=FALSE)</code> +D) Все ответы верны</p>
4.	<pre>Alpha 125.5 0 Beta 235.6 1 Beta 212.03 0 Beta 211.30 0 Alpha 265.46 1 Dataframe.csv</pre> <p>Какая из следующих команд правильно прочитает приведенный выше CSV-файл с 5 строками в наборе данных?</p> <p>A) <code>csv('Dataframe.csv')</code> B) <code>csv('Dataframe.csv',header=TRUE)</code> C) <code>dataframe('Dataframe.csv')</code> +D) <code>csv2('Dataframe.csv',header=FALSE,sep=',')</code></p>
5.	<p>Отсутствующие значения в этом CSV-файле представлены восклицательным знаком («!») и вопросительным знаком («?»). Какой из приведенных ниже кодов правильно прочитает приведенный выше CSV-файл в R?</p> <p>A) <code>csv('Dataframe.csv')</code> B) <code>csv('Dataframe.csv',header=FALSE, sep=',',na.strings=c('?'))</code> +C) <code>csv2('Dataframe.csv',header=FALSE,sep=',',na.strings=c('?', '!'))</code> D) <code>dataframe('Dataframe.csv')</code></p> <pre>6. Column 1 Column 2 Column 3 Row 1 15.5 14.12 69.5 Row 2 18.6 56.23 52.4 Row 3 21.4 47.02 63.21 Row 4 36.1 56.63 36.12 Dataframe.csv</pre>

6.	<p>Приведенный выше CSV-файл имеет название строк, а также название столбцов. Какой из следующих кодов правильно прочитает приведенный выше CSV-файл в R?</p> <p>A) <code>delim('Train.csv',header=T,sep=',',row.names=TRUE)</code> B) <code>csv2('Train.csv',header=TRUE, row.names=TRUE)</code> C) <code>dataframe('Train.csv',header=TRUE,sep=',')</code> +D) <code>csv('Train.csv',header=TRUE,sep=',')</code></p>																																																																														
7.	<table border="1" data-bbox="271 369 766 560"> <thead> <tr> <th></th> <th>Column 1</th> <th>Column 2</th> <th>Column 3</th> </tr> </thead> <tbody> <tr> <td>Row 1</td> <td>15.5</td> <td>14.12</td> <td>69.5</td> </tr> <tr> <td>Row 2</td> <td>18.6</td> <td>56.23</td> <td>52.4</td> </tr> <tr> <td>Row 3</td> <td>21.4</td> <td>47.02</td> <td>63.21</td> </tr> <tr> <td>Row 4</td> <td>36.1</td> <td>56.63</td> <td>36.12</td> </tr> </tbody> </table> <p>Dataframe.csv</p> <p>Какой из следующих кодов будет читать только первые две строки CSV-файла?</p> <p>+A) <code>csv('Dataframe.csv',header=TRUE,row.names=1,sep=',',nrows=2)</code> B) <code>csv2('Dataframe.csv',row.names=1,nrows=2)</code> C) <code>delim2('Dataframe.csv',header=T,row.names=1,sep=',',nrows=2)</code> D) <code>dataframe('Dataframe.csv',header=TRUE,row.names=1,sep=',',skip.last=2)</code></p>		Column 1	Column 2	Column 3	Row 1	15.5	14.12	69.5	Row 2	18.6	56.23	52.4	Row 3	21.4	47.02	63.21	Row 4	36.1	56.63	36.12																																																										
	Column 1	Column 2	Column 3																																																																												
Row 1	15.5	14.12	69.5																																																																												
Row 2	18.6	56.23	52.4																																																																												
Row 3	21.4	47.02	63.21																																																																												
Row 4	36.1	56.63	36.12																																																																												
8.	<table border="1" data-bbox="159 851 1436 1097"> <thead> <tr> <th colspan="4">Dataframe1</th> <th colspan="3">Dataframe2</th> </tr> <tr> <th>Feature1</th> <th>Feature2</th> <th>Feature3</th> <th>Feature4</th> <th>Feature1</th> <th>Feature2</th> <th>Feature3</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>1000</td> <td>25.5</td> <td>10</td> <td>E</td> <td>5000</td> <td>65.5</td> </tr> <tr> <td>B</td> <td>2000</td> <td>35.5</td> <td>34</td> <td>F</td> <td>6000</td> <td>75.5</td> </tr> <tr> <td>C</td> <td>3000</td> <td>45.5</td> <td>78</td> <td>G</td> <td>7000</td> <td>85.5</td> </tr> <tr> <td>D</td> <td>4000</td> <td>55.5</td> <td>3</td> <td>H</td> <td>8000</td> <td>95.5</td> </tr> </tbody> </table> <p>Здесь хранятся два кадра данных Dataframe1 и Dataframe2, показанные выше. Какой из следующих кодов выдаст вывод, показанный ниже?</p> <table border="1" data-bbox="255 1164 1165 1534"> <thead> <tr> <th></th> <th>Feature1</th> <th>Feature2</th> <th>Feature3</th> </tr> </thead> <tbody> <tr> <td>A</td> <td></td> <td>1000</td> <td>25.5</td> </tr> <tr> <td>B</td> <td></td> <td>2000</td> <td>35.5</td> </tr> <tr> <td>C</td> <td></td> <td>3000</td> <td>45.5</td> </tr> <tr> <td>D</td> <td></td> <td>4000</td> <td>55.5</td> </tr> <tr> <td>E</td> <td></td> <td>5000</td> <td>65.5</td> </tr> <tr> <td>F</td> <td></td> <td>6000</td> <td>75.5</td> </tr> <tr> <td>G</td> <td></td> <td>7000</td> <td>85.5</td> </tr> <tr> <td>H</td> <td></td> <td>8000</td> <td>95.5</td> </tr> </tbody> </table> <p>A) <code>merge(dataframe[,1:3],dataframe2)</code> B) <code>merge(dataframe1,dataframe2)[,1:3]</code> C) <code>merge(dataframe1,dataframe2,all=TRUE)</code> +D) Both 1 and 2 E) All of the above</p>	Dataframe1				Dataframe2			Feature1	Feature2	Feature3	Feature4	Feature1	Feature2	Feature3	A	1000	25.5	10	E	5000	65.5	B	2000	35.5	34	F	6000	75.5	C	3000	45.5	78	G	7000	85.5	D	4000	55.5	3	H	8000	95.5		Feature1	Feature2	Feature3	A		1000	25.5	B		2000	35.5	C		3000	45.5	D		4000	55.5	E		5000	65.5	F		6000	75.5	G		7000	85.5	H		8000	95.5
Dataframe1				Dataframe2																																																																											
Feature1	Feature2	Feature3	Feature4	Feature1	Feature2	Feature3																																																																									
A	1000	25.5	10	E	5000	65.5																																																																									
B	2000	35.5	34	F	6000	75.5																																																																									
C	3000	45.5	78	G	7000	85.5																																																																									
D	4000	55.5	3	H	8000	95.5																																																																									
	Feature1	Feature2	Feature3																																																																												
A		1000	25.5																																																																												
B		2000	35.5																																																																												
C		3000	45.5																																																																												
D		4000	55.5																																																																												
E		5000	65.5																																																																												
F		6000	75.5																																																																												
G		7000	85.5																																																																												
H		8000	95.5																																																																												
9.	<table border="1" data-bbox="271 1758 526 2049"> <thead> <tr> <th></th> <th>V1</th> <th>V2</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>121.5</td> <td>461</td> </tr> <tr> <td>2</td> <td>516</td> <td>1351</td> </tr> <tr> <td>3</td> <td>451</td> <td>6918</td> </tr> <tr> <td>4</td> <td>613</td> <td>112</td> </tr> <tr> <td>5</td> <td>112.36</td> <td>230</td> </tr> <tr> <td>6</td> <td>25.23</td> <td>1456</td> </tr> <tr> <td>7</td> <td>12</td> <td>457</td> </tr> </tbody> </table> <p>Dataframe</p>		V1	V2	1	121.5	461	2	516	1351	3	451	6918	4	613	112	5	112.36	230	6	25.23	1456	7	12	457																																																						
	V1	V2																																																																													
1	121.5	461																																																																													
2	516	1351																																																																													
3	451	6918																																																																													
4	613	112																																																																													
5	112.36	230																																																																													
6	25.23	1456																																																																													
7	12	457																																																																													

	<p>Набор данных был прочитан в R и сохранен в переменной «dataframe». Какой из приведенных ниже кодов выдаст сводку (среднее, моду, медиану) всего набора данных в одной строке кода?</p> <p>A) summary(dataframe) B) stats(dataframe) C) summarize(dataframe) D) summarise(dataframe) +E) Нет верного ответа</p>															
10	<p>Набор данных был прочитан в R и сохранен в переменной «dataframe». Отсутствующие значения были прочитаны как NA.</p> <table data-bbox="271 515 526 705"> <tr><td>A</td><td>10</td><td>Sam</td></tr> <tr><td>B</td><td>NA</td><td>Peter</td></tr> <tr><td>C</td><td>30</td><td>Harry</td></tr> <tr><td>D</td><td>40</td><td>NA</td></tr> <tr><td>E</td><td>50</td><td>Mark</td></tr> </table> <p>Dataframe</p> <p>Какой из следующих кодов не даст количество пропущенных значений в каждом столбце?</p> <p>A) colSums(is.na(dataframe)) B) apply(is.na(dataframe),2,sum) C) sapply(dataframe,function(x) sum(is.na(x)) +D) table(is.na(dataframe))</p>	A	10	Sam	B	NA	Peter	C	30	Harry	D	40	NA	E	50	Mark
A	10	Sam														
B	NA	Peter														
C	30	Harry														
D	40	NA														
E	50	Mark														