



# 70-773<sup>Q&As</sup>

Analyzing Big Data with Microsoft R (beta)

## Pass Microsoft 70-773 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.lead4pass.com/70-773.html>

100% Passing Guarantee  
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft  
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





### QUESTION 1

You plan to read data from an Oracle database table and to store the data in the file system for later processing by dplyrXdf. The size of the data is larger than the memory on the server to be used for modelling.

You need to ensure that the data can be processed by dplyrXdf in the least amount of time possible.

How should you transfer the data from the Oracle database?

- A. Define a data source to the Oracle database server by using RxOdbcData. Use rxImport to save the data to a comma-separated values (CSV) file.
- B. Use the RODBC library, connect to the Oracle database server by using odbcConnect, and then use rxDataStep to export the data to a comma-separated values (CSV) file.
- C. Define a data source to the Oracle database server by using RxOdbcData, and then use rxImport to save the data to an XDF file.
- D. Use the RODBC library, connect to the Oracle database server by using odbcConnect, and then use rxSplit to save the data to multiple comma-separated values (CSV) file.

Correct Answer: C

---

### QUESTION 2

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Microsoft SQL Server instance that has R Services (In-Database) installed.

You need to monitor the R jobs that are sent to SQL Server.

Solution: You register an Extended Events package.

Does this meet the goal?

- A. Yes
- B. No

Correct Answer: A

References: <https://docs.microsoft.com/en-us/sql/advanced-analytics/r/extended-events-for-sql-server-r-services>

---

### QUESTION 3



You are using rxPredict for a logistic regression model.

You need to obtain prediction standard errors and confidence intervals.

Which R code segment should you use? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Values	Answer Area
<input type="text" value="confidence"/>	<pre>model &lt;- <input type="text" value="Value"/> (default ~ year + creditScore + yearsEmploy + ccDebt, data = trainingDataFileName, blocksPerRead = 2, verbose = 1, reportProgress=2, covCoef= <input type="text" value="Value"/> ) rxPredict(model, data = targetDataFileName, outData = targetDataFileName, computeStdErr = <input type="text" value="Value"/>, interval = "<input type="text" value="Value"/>", overwrite=TRUE)</pre>
<input type="text" value="FALSE"/>	
<input type="text" value="glm"/>	
<input type="text" value="none"/>	
<input type="text" value="rxLinMod"/>	
<input type="text" value="rxLogit"/>	
<input type="text" value="TRUE"/>	

Correct Answer:

Values	Answer Area
<input type="text" value="confidence"/>	<pre>model &lt;- <input type="text" value="rxLogit"/> (default ~ year + creditScore + yearsEmploy + ccDebt, data = trainingDataFileName, blocksPerRead = 2, verbose = 1, reportProgress=2, covCoef= <input type="text" value="TRUE"/> ) rxPredict(model, data = targetDataFileName, outData = targetDataFileName, computeStdErr = <input type="text" value="TRUE"/>, interval = "<input type="text" value="confidence"/>", overwrite=TRUE)</pre>
<input type="text" value="FALSE"/>	
<input type="text" value="glm"/>	
<input type="text" value="none"/>	
<input type="text" value="rxLinMod"/>	
<input type="text" value="rxLogit"/>	
<input type="text" value="TRUE"/>	

#### QUESTION 4

You have the following regression forest.



```
rxDForest(formula = stack.loss ~ Air.Flow + Water.Temp +  
Acid.Conc., data = stackloss, maxDepth = 3, nTree = 200, mTry =  
2)
```

```
Type of decision forest: anova  
Number of trees: 200  
No. of variables tried at each split: 2  
Mean of squared residuals: 44.54992 % Var explained: 65
```



Which variable contributes the most to the dependent variable?

- A. stack.loss
- B. Water.Temp
- C. Air.Flow
- D. Acid.Conc

Correct Answer: D

### QUESTION 5

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.

You are performing feature engineering and data preparation for the datasets.

The following is a sample of the dataset.

```
rxGetInfo(df)  
head(df)
```

```
   age incwage perwt  wkswork1  state  
1  50   9000    30      48  Indiana  
2  41  35000    20      48  Indiana  
3  55  40400    21      52  Indiana  
4  56  45000    30      52  Indiana  
5  46  17200    60      52  Indiana  
6  49  35000    21      52  Indiana
```



End of repeated scenario.

You plan to score some data to create data features to address empty rows.

You have the following R code.

```

xdPath <- file.path(rxGetOption("[sampleInData], "), "inputfile.xdf")
xdfLagged <- [sampleOutDataincludingFeatures](fileext = ".xdf")
rxSort(inData = xdfPath,
       outFile = xdfLagged,
       sortByVars = "Date")
rxDataStep(inData = xdfLagged,
           outFile = xdfLagged,
           transformObjects = list(
             varToLag = "Open",
             newName = "previousOpen"),
           transformFunc = lagVar,
           append = "cols",
           overwrite = TRUE)
rxDataStep(xdfLagged,
           varsToKeep = c("Date", "Open", "previousOpen"),
           numRows = 10)

```

You need to transform the data and overwrite the current dataset.

Which R code segment should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

**Answer Area**

(inData = [sampleInData], outFile=[sampleOutDataincludingFeatures],  

rxCube
rxDataStep
rxExec
transform

transformFunc = computeNonLagFeatures,

<input type="text" value=""/>
overwrite=FALSE
overwrite=TRUE
varsToDrop=All
varsToDrop=NONE

)

Correct Answer:



Answer Area

(inData = [sampleInData], outFile=[sampleOutDataincludingFeatures],  
rxCube  
rxDataStep  
rxExec  
transform  
transformFunc = computeNonLagFeatures,  )  
overwrite=FALSE  
overwrite=TRUE  
varsToDrop=All  
varsToDrop=NONE

QUESTION 6

You have a Microsoft SQL Server instance that has R Services (In-Database) installed. The server has a comma-separated values (CSV) file stored in the local file system.

For analytic purposes, you need to read the CSV file into a database table in the SQL Server instance.

You connect to the SQL Server instance by using SQL Server Management Studio.

What should you use from `sp_execute_external_script`?

- A. `RxSqlServerData` and specify the CSV file path in the connection string
- B. `rxDataStep` and specify the CSV file path as the `inFile` argument
- C. `rxImportToXdf` and specify specify the CSV file as the input
- D. `read.csv` and specify the CSV file path as the parameter

Correct Answer: D

QUESTION 7

You have a dataset that has a character variable. You need to create a bag of counts of n-grams. Which function should you use?

- A. `featurizeText()`
- B. `categoricalHash()`
- C. `concat()`
- D. `selectFeatures()`
- E. `categorical()`

Correct Answer: A



References: <https://docs.microsoft.com/en-us/machine-learning-server/python-reference/microsoftml/featurize-text>

---

### QUESTION 8

You need to use the ScaleR distributed processing in an Apache Hadoop environment. Which data source should you use?

- A. Microsoft SQL Server database
- B. XDF data files
- C. ODBC data
- D. Teradata database

Correct Answer: B

References: <https://docs.microsoft.com/en-us/machine-learning-server/r/how-to-revoscaler-hadoop>

---

### QUESTION 9

You have one-class support vector machines (SVMs).

You have a large dataset, but you do not have enough training time to fully test the model.

What is an alternative method to validate the model?

- A. Use Principal Components Analysis (PCA)-Based Anomaly Detection.
- B. Replace the SVMs with two-class SVMs.
- C. Perform feature selection.
- D. Use outlier detection.

Correct Answer: A

---

### QUESTION 10

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You need to estimate a model where the outcome variable is continuous, is in the range of  $[0, \text{inf}]$ , and has a substantial mass at an exact value of 0.

Which function should you use?

- A. rxPredict



- B. rxLogit
- C. summary
- D. rxLinMod
- E. rxTweedie
- F. stepAic
- G. rxTransform
- H. rxDataStep

Correct Answer: F

---

#### QUESTION 11

You need to build a model that looks at the probability of an outcome. You must regulate between L1 and L2. Which classification method should you use?

- A. Two-Class Neural Network
- B. Two-Class Support Vector Machine
- C. Two-Class Decision Forest
- D. Two-Class Logistic Regression

Correct Answer: D

References: <https://msdn.microsoft.com/en-us/library/azure/dn905994.aspx>

---

#### QUESTION 12

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You need to calculate a measure of central tendency and variability for the variables in a dataset that is grouped by using another categorical variable.

What should you use?

- A. the Describe package
- B. the rxHistogram function
- C. the rxSummary function
- D. the rxQuantile function





- E. the rxCube function
- F. the summary function
- G. the rxCrossTabs function
- H. the ggplot2 package

Correct Answer: C

---

### QUESTION 13

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You build a model that uses xyz regression.

You need to estimate a model that predicts a binary variable.

Which function should you use?

- A. rxPredict
- B. rxLogit
- C. summary
- D. rxLinMod
- E. rxTweedie
- F. stepAic
- G. rxTransform
- H. rxDataStep

Correct Answer: B

References: <https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-logistic-regression>

---

### QUESTION 14

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.



You use `dplyr::df`, and you discover that after you exit the session, the output files that were created were deleted.

You need to prevent the files from being deleted.

Solution: You remove all instances of the `file.remove` method.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

### QUESTION 15

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.

You are performing feature engineering and data preparation for the datasets.

The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)
```

	age	incwage	perwt	wkswork1	state
1	50	9000	30	48	Indiana
2	41	35000	20	48	Indiana
3	55	40400	21	52	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	52	Indiana
6	49	35000	21	52	Indiana

End of repeated scenario.

You need to sort the data from the dataset sample and to remove duplicates by using `wkswork1`.

Which R code segment should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),

,  = "wkswork1")

removeDupKeys = FALSE
removeDupKeys = TRUE
rxMerge = FALSE
rxMerge = TRUE

dupFreqVar
varsToDrop
VarsToKeep

Correct Answer:

Answer Area

rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),

,  = "wkswork1")

removeDupKeys = FALSE
removeDupKeys = TRUE
rxMerge = FALSE
rxMerge = TRUE

dupFreqVar
varsToDrop
VarsToKeep

[70-773 PDF Dumps](#)

[70-773 Exam Questions](#)

[70-773 Braindumps](#)



To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

## Try our product !

100% Guaranteed Success

100% Money Back Guarantee

365 Days Free Update

Instant Download After Purchase

24x7 Customer Support

Average 99.9% Success Rate

More than 800,000 Satisfied Customers Worldwide

Multi-Platform capabilities - [Windows](#), [Mac](#), [Android](#), [iPhone](#), [iPod](#), [iPad](#), [Kindle](#)

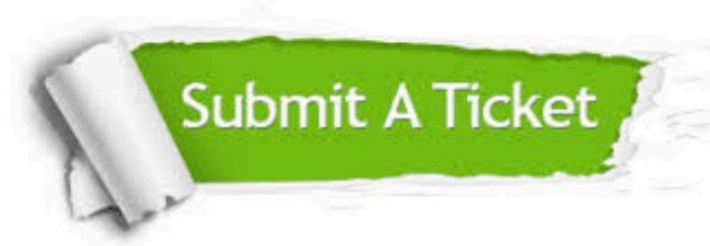
We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<https://www.lead4pass.com/allproducts>

## Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 <p><b>One Year Free Update</b> Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p><b>Money Back Guarantee</b> To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p><b>Security &amp; Privacy</b> We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information &amp; peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.

Copyright © lead4pass, All Rights Reserved.