Top Answers to R Interview Questions

**1. Compare R & Python**

|  |  |
| --- | --- |
| **R  programming Language** | **Python programming language** |
| Model Building is similar to  Python | Model Building is similar to R. |
| Model Interpretability is  good | Model Interpretability is  not good |
| Production is not better than Python. | Production is good |
| R has good community support over Python. | Community Support  is not better than R |
| Data Science Libraries are same as Python. | Data Science Libraries are same as R. |
| R has good data visualizations libraries and tools | Data visualization is not better than R |
| R has a steep learning curve. | Learning Curve in Python is easier than learning R. |

**2. Explain the data import in R language.**

R provides to import data in R language. To begin with the R commander GUI, user should type the commands in the command Rcmdr into the console. Data can be imported in R language in 3 ways such as:

* Select the data set in the dialog box or enter the name of the data set as required.
* Data is entered directly using the editor of R Commander via Data->New Data Set. This works good only when the data set is not too large.
* Data can also be imported from a URL or from plain text file (ASCII), or from any statistical package or from the clipboard.

**3. Explain how to communicate the outputs of data analysis using R language.**

Combine the data, code and analysis results in a single document using knitr for Reproducible research done. Helps to verify the findings, add to them and engage in conversations. Reproducible research makes it easy to redo the experiments by inserting new data values and applying it to different various problems.

**4. Difference between library () and require () functions in R language.**

|  |  |
| --- | --- |
| **library()** | **require()** |
| Library () function gives an error message display, if the desired package cannot be loaded. | Require () function is used inside function and throws a warning messages whenever a particular package is not Found |
| It loads the packages whether it is already loaded or not, | It just checks that it is loaded, or loads it if it isn’t (use in functions that rely on a certain package). The documentation explicitly states that neither function will reload an already loaded package. |

Consider a related program for the above differentiation.

if(!require(package, character.only=T, quietly=T)) {

install.packages (package)

library(package, character.only=T)

}

For multiple packages you can use

for(package in c(”, ”)) {

if(!require(package, character.only=T, quietly=T)) {

install.packages (package)

library(package, character.only=T)

}

}

**5. What is R?**

R is a programming language which is used for developing statistical software and data analysis. It is being increasingly deployed for machine learning applications as well.

**6. How R commands are written?**

By using # at the starting of the line of code like #division commands are written.

**7. What is t-tests() in R?**

It is used to determine that the means of two groups are equal or not by using t.test() function.

**8. What are the disadvantages of R Programming?**

The disadvantages are:-

* Lack of standard GUI
* Not good for big data.
* Does not provide spreadsheet view of data.

**9. What is the use of With () and By () function in R?**

with() function applies an expression to a dataset.

#with(data,expression)

By() function applies a function t each level of a factors.

#by(data,factorlist,function)

**10. In R programming, how missing values are represented?**

In R missing values are represented by NA which should be in capital letters.

**11. What is the use of subset() and sample() function in R?**

Subset() is used to select the variables and observations and sample() function is used to  generate  a random sample of the size n from a dataset.

**12. Explain what is transpose.**

Transpose is used for reshaping of the data which is used for analysis. Transpose is performed by t() function.

**13. What are the advantages of R?**

* The advantages are:-
* It is used for managing and manipulating of data.
* No license restrictions
* Free and open source software.
* Graphical capabilities of R are good.
* Runs on many Operating system and different hardware and also run on 32 & 64 bit processors etc.

**14. What is the function used for adding datasets in R?**

For adding two datasets rbind() function is used but the column of two datasets must be same.

Syntax: rbind(x1,x2……) where x1,x2: vector, matrix, data frames.

**15. How you can produce co-relations and covariances?**

Cor-relations is produced by cor() and covariances is produced by cov() function.

**16. What is difference between matrix and dataframes?**

Dataframe can contain different type of data but matrix can contain only similar type of data.

**17. What is difference between lapply and sapply?**

lapply is used to show the output in the form of list whereas sapply is used to show the output in the form of vector or data frame

**18. What is the difference between seq(4) and seq\_along(4)?**

Seq(4) means vector from 1 to 4 (c(1,2,3,4)) whereas seq\_along(4) means a vector of the  length(4) or 1(c(1)).

**19. Explain how you can start the R commander GUI.**

rcmdr command is used to start the R commander GUI.

**20. What is the memory limit of R?**

In 32 bit system memory limit is 3Gb but most versions limited to 2Gb and in 64 bit system memory limit is 8Tb.

**21. How many data structures R has?**

There are 5 data structure in R i.e. vector, matrix, array which are of homogenous type and other two are list and data frame which are heterogeneous.

**22. Explain how data is aggregated in R.**

There are two methods that is collapsing data by using one or more BY variable and other is aggregate() function in which BY variable should be in list.

**23. How many sorting algorithms are available?**

There are 5 types of sorting algorithms are used which are:-

* Bubble Sort
* Selection Sort
* Merge Sort
* Quick Sort
* Bucket Sort

**24. How to create new variable in R programming?**

For creating new variable assignment operator ‘<-’ is used  
For e.g. mydata$sum <- mydata$x1 + mydata$x2

**25. What are R packages?**

Packages are the collections of data, R functions and compiled code in a well-defined format and these packages are stored in library. One of the strengths of R is the user-written function in R language.

**26. What is the workspace in R?**

Workspace is the current R working environment which includes any user defined objects like vector, lists etc.

**27. What is the function which is used for merging of data frames horizontally in R?**

Merge()function is used to merge two data frames  
Eg. Sum<-merge(data frame1,data frame 2,by=’ID’)

**28. what is the function which is used for merging of data frames vertically in R?**

rbind() function is used to merge two data frames vertically.  
Eg.

Sum<- rbind(data frame1,data frame 2)

**29. What is the power analysis?**

It is used for experimental design .It is used to determine the effect of given sample size.

**30. Which package is used for power analysis in R?**

Pwr package is used for power analysis in R.

**31. Which method is used for exporting the data in R?**

There are many ways to export the data into another formats like SPSS, SAS , Stata , Excel Spreadsheet.

**32. Which packages are used for exporting of data?**

For excel xlsReadWrite package is used and for sas,spss ,stata foreign package is implemented.

**33. How impossible values are represented in R?**

In R NaN is used to represent impossible values.

**34. Which command is used for storing R object into a file?**

Save command is used for storing R objects into a file.  
Syntax: >save(z,file=”z.Rdata”)

**35. Which command is used for restoring R object from a file?**

load command is used for storing R objects from a file.  
Syntax: >load(”z.Rdata”)

**36. What is the use of coin package in R?**

Coin package is used to achieve the re randomization or permutation based statistical tests.

**37. Which function is used for sorting in R?**

order() function is used to perform the sorting.

**38. What is the use of tapply?**

IOS-6.1.3

**39. What happens when the application object does not handle an event?**

The event will be dispatched to your delegate for processing.

**40. Explain app specific objects which store the app contents.**

The app specific objects are Data model objects that store app’s contents.

**41. Explain the purpose of using UIWindow object?**

UIWindow object coordinates the one or more views presenting on the screen.

**42. Tell me the super class of all view controller objects.**

UIView Controller class.

**43. How to create axes in the graph?**

Using axes() function custom axes are created.

**44. What is the use of abline() function?**

abline() function is add the reference line to a graph.  
Syntax:-

abline(h=yvalues, v=xvalues)

**45. Why vcd package is used?**

vcd package provides different methods for visualizing multivariate categorical data.

**46. What is GGobi?**

GGobi is an open source program for visualization for exploring high dimensional typed data.

**47. What is iPlots?**

It is a package which provide bar plots, mosaic plots, box plots, parallel plots, scatter plots and histograms.

**48. What is the use of lattice package?**

lattice package is to improve on base R graphics by giving better defaults and it have the ability to easily display multivariate relationships.

**49. What is fitdistr() function?**

It is used to provide the maximum likelihood fitting of univariate distributions. It is defined under the MASS package.

**50. Which data structures are used to perform statistical analysis and create graphs.**

Data structures are vectors, arrays, data frames and matrices.

**51. What is the use of sink() function?**

It defines the direction of output.

**52. Why library() function is used?**

This function is used to show the packages which are installed.

**53. Why search() function is used?**

By this function we see that which packages are currently loaded.

**54. On which type of data binary operators are worked?**

Binary operators are worked on matrices, vectors and scalars.

**55. What is the use of doBY package?**

It is used to define the desired table using function and model formula.

**56. Which function is used to create frequency table?**

Frequency table is created by table() function.

**57. Define loglm() function.**

Loglm() function is used to create log-linear models.

**58. What is the use of corrgram() function?**

corrgram() function is used to plot correlograms.

**59. How to create scatterplot matrices?**

Pair() or splom() function is used for create scatterplot matrices.

**60. What is npmc?**

It is a package which gives nonparametric multiple comparisons.

**61. What is the use of diagnostic plots?**

It is used to check the normality, heteroscedasticity and influential observations.

**62. Define anova() function.**

anova() is used to compare the nested models.

**63. What is cv.lm() function?**

It is defined under the DAAG package which is used for k-fold validation.

**64. Define stepAIC() function.**

It is define under the MASS package which performs stepwise model selection under exact AIC.

**65. Define leaps().**

It is used to perform the all-subsets regression and it is defined under the leaps package.

**66. Define relaimpo package.**

It is used to measure the relative importance of each of the predictor in the model.

**67. Why car package is used?**

It  provide a variety of regression including scatter plots, variable plots and it also enhanced diagnostic.

**68. Define robust package.**

It provides a library of robust methods including regression.

**69. What is robustbase?**

It is a package which provides basic robust statistics including model selection methods.

**70. Define plotmeans().**

It is define under gplots package which includes confidence intervals and it produces mean plot for single factors.

**71. What is the full form of MANOVA?**

MANOVA stands for multivariate analysis of variance.

**72. What is the use of MANOVA?**

By using MANOVA we can test more than one dependent variable simultaneously.

**73. Define mshapiro.test( ).**

It is a function which defines in mvnormtest package. It produces the Shapiro-wilk test for multivariate normality.

**74. Define barlett.test().**

Barlett.test() is used to provide a parametric k-sample test of the equality of variances.

**75. What is fligner.test()?**

It is a function which provides a non-parametric k sample test of the equality of variances.

**76. Define hovplot().**

It is define in HH package which provides a graphic test of homogeneity of variance based on brown forsyth.

**77. Which variables are represented by lower case letters?**

Numerical variables are represented by lower case letters.

**78. Which variables are represented by upper case letters?**

Categorical factors are represented by upper case letters.

**79. What is logistic regression?**

Logistic regression is used to predict the binary outcome from the given set of continuous predictor variables.

**80. Define Poison regression.**

It is used to predict the outcome variable which represents counts from the given set of continuous predictor variable.

**81. Define Survival analysis.**

It includes number of techniques which is used for modeling the time to an event.

**82. What is the use survfit() function?**

It estimates a survival distribution one or more groups.

**83. Define survdiff().**

It determines the differences in survival distribution between two or more groups.

**84. What is coxph()?**

It is a function which is used to model the hazard function on the set of predictor variable.

**85. In which package survival analysis is defined?**

Survival analysis is defined under the survival package.

**86. What is the use of MASS package?**

MASS functions include those functions which performs linear and quadratic discriminant function analysis.

**87. Define qda().**

qda() prints a quadratic discriminant function.

**88. Define lda().**

lda() is used to print the discriminant functions which is based on centered variable.

**89. What is the use of forecast package?**

It provides the functions which are used for automatic selection of ARIMA and exponential models.

**90. Define auto.arima().**

It is used to handle the seasonal as well as non-seasonal ARIMA models.

**91. What is principal() function?**

It is define in psych package which is used to rotate and extract the principal components.

**92. What is FactoMineR?**

It is a package which includes quantitative and qualitative variables. It also includes supplementary variables and observations.

**93. What is the full form of CFA?**

CFA stands for Confirmatory Factor Analysis.

**94. What is the use of boot.sem() function?**

It is used to bootstrap the structural equation model.

**95. What is the full form of SEM?**

SEM stands for Structural Equation Modeling.

**96. Which function performs classical multidimensional scaling?**

cmdscale() function is used to perform classical multidimensional scaling.

**97. Define isoMDS().**

This function is defined under the MASS package which performs nonmetric multidimensional scaling.

**98. Which function perform individual difference scaling?**

It is done by indscal() function.

**99. What is pvclust() function ?**

It comes under the pvclust package which provides p-values for hierarchical clustering.

**100. Define cluster.stats() ?**

It is define in fpc package which provide a method for comparing the similarity of two clusters solution using different validation criteria.

**101. What we use party package?**

It is used to provide a non-parametric regression for ordinal, nominal, censored and multivariate responses.

**102. Which package provide the bootstrapping?**

boot package is used which provide bootstrapping.

**103. Define MATLAB package.**

Matlab package includes those wrapper functions and variable which are used to replicate matlab function calls.

**104. What is the of use Matrix package?**

Matrix package includes those function which support sparse and dense matrices like Lapack, BLAS etc.