100+ Commonly Asked Data Science Interview Questions

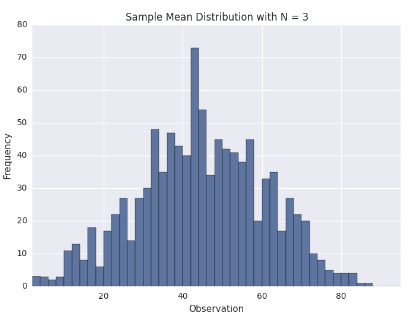
* Posted by [Vincent Granville](https://www.datasciencecentral.com/profile/VincentGranville) on October 15, 2017 at 4:30pm
* [View Blog](https://www.datasciencecentral.com/profiles/blog/list?user=3v6n5b6g08kgn)

Here is a new set of easy questions recently published, covering

* Statistics
* Programming (General, Big Data, Python, R, SQL)
* Modeling
* Behavioral
* Culture Fit
* Problem-Solving

**Statistics:**

* Prove that a random variable with a distribution on [0,1] (that is, the density function is equal to 0 outside [0, 1]) has an expectation always between 0 and 1. Prove that its variance is maximum and equal to 1/12 if and only if the distribution is uniform on [0, 1]. - [Click here](http://www.analyticbridge.datasciencecentral.com/profiles/blogs/interesting-probability-problem-for-serious-geeks) for more on this topic.
* What is the Central Limit Theorem and why is it important? - [Answer](https://spin.atomicobject.com/2015/02/12/central-limit-theorem-intro/) (or [click here](https://www.datasciencecentral.com/profiles/blogs/the-fundamental-statistics-theorem-revisited) for more advanced material on this topic)
* What is sampling? How many sampling methods do you know?
* What is the difference between Type I vs Type II error? [Answer](http://www.analyticbridge.datasciencecentral.com/profiles/blogs/type-i-and-type-ii-errors-in-one-picture)
* What is linear regression? What do the terms P-value, coefficient, R-Squared value mean? What is the significance of each of these components? - [Answer](http://www.statisticssolutions.com/assumptions-of-linear-regression/), [Answer](http://blog.minitab.com/blog/adventures-in-statistics/how-to-interpret-regression-analysis-results-p-values-and-coefficients)
* What are the assumptions required for linear regression? - There are four major assumptions: 1. There is a linear relationship between the dependent variables and the regressors, meaning the model you are creating actually fits the data, 2. The errors or residuals of the data are normally distributed and independent from each other, 3. There is minimal multicollinearity between explanatory variables, and 4. Homoscedasticity. This means the variance around the regression line is the same for all values of the predictor variable.
* What is a statistical interaction? - [Answer](http://icbseverywhere.com/blog/mini-lessons-tutorials-and-support-pages/statistical-interactions/)
* What is selection bias?
* What is an example of a dataset with a non-Gaussian distribution? - [Example](https://www.datasciencecentral.com/profiles/blogs/zipf-s-distribution-example-of-a-great-application)
* What is the Binomial Probability Formula?

[](https://api.ning.com/files/Z7b4QxQRk0AFZ1kMdDLQW4UNC5AzJwMMgnWozZ8-pbNpM7d7mFxvFx5RxW8ed56jAXFO1n1x5mhKv8ekKqb1iXAIz7YM9EJU/Capture.PNG)

**R programming**

* What are the different types of sorting algorithms available in R language?  -There are insertion, bubble, and selection sorting algorithms.
* What are the different data objects in R?
* What packages are you most familiar with? What do you like or dislike about them?
* How do you access the element in the 2nd column and 4th row of a matrix named M?
* What is the command used to store R objects in a file?
* What is the best way to use Hadoop and R together for analysis?
* How do you split a continuous variable into different groups/ranks in R?
* Write a function in R language to replace the missing value in a vector with the mean of that vector.

**SQL**

* What is the purpose of the group functions in SQL? Give some examples of group functions.
* Group functions are necessary to get summary statistics of a dataset. COUNT, MAX, MIN, AVG, SUM, and DISTINCT are all group functions
* Tell me the difference between an inner join, left join/right join, and union. - [Answer](https://www.quora.com/SQL-What-is-the-difference-between-inner-join-left-join-right-join-and-full-join)
* What does UNION do? What is the difference between UNION and UNION ALL?
* What is the difference between SQL and MySQL or SQL Server?
* If a table contains duplicate rows, does a query result display the duplicate values by default? How can you eliminate duplicate rows from a query result?

The full list is available [here](https://www.springboard.com/blog/data-science-interview-questions/).  Many lists of interview questions have been published over the last few years. Here is a selection, including some from me:

* [50 Questions to Test True Data Science Knowledge](https://www.datasciencecentral.com/profiles/blogs/25-questions-to-detect-fake-data-scientists)
* [66 job interview questions for data scientists](https://www.datasciencecentral.com/profiles/blogs/66-job-interview-questions-for-data-scientists)
* [20 Job Interview Questions for IoT Professionals](https://www.iotcentral.io/blog/20-job-interview-questions-for-iot-professionals)
* [Answers to dozens of data science job interview questions](https://www.datasciencecentral.com/profiles/blogs/answers-to-dozens-of-data-science-job-interview-questions)
* [46 SQL Job Interview Questions for Data Scientists](https://www.datasciencecentral.com/profiles/blogs/46-questions-on-sql-to-test-a-data-science-professional-skilltest)
* [R Programming: 35 Job Interview Questions and Answers](https://www.datasciencecentral.com/profiles/blogs/r-programming-job-interview-questions-and-answers)
* [Top Hadoop Big Data Interview Questions and Answers](http://www.hadoop360.datasciencecentral.com/blog/top-hadoop-big-data-interview-questions-and-answers)
* [70 MongoDB Interview Questions and Answers](http://www.hadoop360.datasciencecentral.com/blog/70-mongodb-interview-questions-and-answers)
* [100 Data Science Interview Questions and Answers](https://www.datasciencecentral.com/profiles/blogs/100-data-science-interview-questions-and-answers)
* [40 Interview Questions asked at Startups in Machine Learning](https://www.datasciencecentral.com/xn/detail/6448529%3ABlogPost%3A470301)
* [19 Worst Mistakes at Data Science Job Interviews](https://www.datasciencecentral.com/profiles/blogs/17-worst-mistakes-at-data-science-job-interviews)

**DSC Resources**

* Services: [Hire a Data Scientist](http://careers.analytictalent.com/jobs/products) | [Search DSC](https://www.datasciencecentral.com/page/search?q=Python) | [Classifieds](http://classifieds.datasciencecentral.com/) | [Find a Job](http://www.analytictalent.com/)
* Contributors: [Post a Blog](https://www.datasciencecentral.com/profiles/blog/new) | [Ask a Question](https://www.datasciencecentral.com/forum/topic/new)
* Follow us: [@DataScienceCtrl](https://www.twitter.com/datasciencectrl) | [@AnalyticBridge](https://www.twitter.com/analyticbridge)

Popular Articles

* [Difference between Machine Learning, Data Science, AI, Deep Learnin...](https://www.datasciencecentral.com/profiles/blogs/difference-between-machine-learning-data-science-ai-deep-learning)
* [What is Data Science? 24 Fundamental Articles Answering This Question](https://www.datasciencecentral.com/profiles/blogs/20-articles-about-core-data-science)
* [Hitchhiker's Guide to Data Science, Machine Learning, R, Python](https://www.datasciencecentral.com/profiles/blogs/hitchhiker-s-guide-to-data-science-machine-learning-r-python)
* [Advanced Machine Learning with Basic Excel](https://www.datasciencecentral.com/profiles/blogs/advanced-machine-learning-with-basic-excel)