

What is the difference between a Statistician and a Data Scientist?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the difference between a Statistician and a Data Scientist?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155152>

A statistician is a professional who primarily deals with gathering, analyzing, and interpreting data to draw conclusions and make predictions. They use mathematical and statistical techniques to organize and interpret data.

On the other hand, a data scientist is a professional who combines technical skills, domain knowledge, and programming expertise to extract meaningful insights and knowledge from large, complex datasets. They use a variety of tools and techniques, including statistical methods, machine learning, and data mining, to uncover patterns and trends in data.

While both statisticians and data scientists work with data, the key difference lies in their approach. Statisticians focus on the analysis and interpretation of data to make informed decisions, while data scientists use a more broad and interdisciplinary approach to extract valuable insights from data.

In summary, statisticians and data scientists have overlapping skills and abilities, but their roles and perspectives differ in terms of their approach to data analysis and decision-making.

Statistician vs. Data Scientist: What's the Difference?

Statisticians and data scientists both work heavily with data, but there are some key differences between the two professions:

Difference #1 (Types of Data) - Data scientists tend to spend more time gathering and cleaning imperfect data while statisticians are usually provided with tidy data.

Difference #2 (End Goals) - Data scientists tend to focus on creating models that predict outcomes while statisticians tend to focus on building models that accurately describe the relationship between variables.

Difference #3 (Production) - Data scientists tend to build models that are put into production at companies while statisticians tend to build models that can provide insights or explanations of phenomenon.

Keep reading for an in-depth explanation of these differences.

Difference #1: Types of Data

In general, data scientists often work with data that is messier, harder to extract, and much larger than the type of data used by statisticians.

For example, a data scientist that works at a real estate firm might have to extract datasets that contain millions of rows from several different external servers that are all in different formats.

She would need extensive knowledge of SQL and at least one programming language (like or) in order to extract the data and wrangle it into a format that is suitable for modeling.

By contrast, statisticians tend to work with smaller datasets that are already in a neat format.

For example, a statistician that works for a biomedical company may be given an Excel file with 50 rows that contains information about blood pressure, heart rate, and cholesterol levels for 50 different patients.

Rather than spending their time extracting and cleaning data, they would likely spend more time deciding on a suitable or model to fit to the data and checking that the of their chosen test or statistical model are met.

Difference #2: End Goals

In many cases, a data scientist's end goal is to create some type of model that can accurately predict some outcome.

For example, a data scientist who works for a financial company might attempt to build a that can accurately predict whether certain individuals will default on a loan.

They will fit a variety of models using different combinations of predictor variables and try to find the model that produces the most accurate predictions.

Their end goal is to create an accurate model rather

than quantifying exactly how each predictor variable is related to the .

For example, a statistician that works at a university might recruit 30 students to participate in a study that quantifies exactly how different studying habits affect exam scores.

In this scenario, the statistician would be more concerned with interpreting the coefficients of the regression model and analyzing their corresponding to understand whether they have a statistically significant relationship with the response variable.

Difference #3: Production

In general, data scientists tend to build statistical models that are put into production at companies far more often than statisticians.

For example, a data scientist that works at a large grocery chain might build a model that is able to accurately forecast sales of various products.

His end goal would be to work with developers at the company who can help him place his model into a

server that runs on a nightly basis that can forecast the sales of products for each new day.

By contrast, statisticians rarely create models that are put into any type of production.

For example, a statistician that works at a healthcare company might build a model that describes the relationship between various lifestyle factors (smoking, exercise, diet, etc.) but their end goal is to simply *quantify* the relationship between these factors and some response variable like lifespan.

Their end goal is to build a model that provides them with some insights rather than a model that is put into a production environment.

Conclusion

Statisticians and data scientists both work with data in their everyday roles, but they do so in different ways.

Data scientists tend to work with a wider variety of data that is often messy and needs to be wrangled while statisticians often work with smaller and more tidy datasets.

Data scientists also tend to focus more on creating models that can accurately predict outcomes while statisticians tend to build models that can accurately explain the relationship between variables.

Lastly, data scientists tend to put models into production at companies while statisticians often summarize and report their findings to provide insights into real-world phenomena.

The following articles explain the importance of statistics in various fields:

ARABPSYCHOLOGY.COM