

Exam completed on: April 6, 2023

Score: 77% · 37/48

Exam questions

 Show all Show wrong only

✓ Question 1 of 48

Which two methods are common algorithms for classifying new cases into existing categories?

prior probabilities and posterior probabilities

hierarchical clustering and connectionist models

✓ k-means and k-nearest neighbors

exploratory and inferential statistics

✓ Question 2 of 48

Data science methods can contribute to business intelligence by which tasks?

modeling outcomes

collecting and cleaning data

finding trends and anomalies in the data

✓ all these answers

✓ Question 3 of 48

According to Drew Conway data science Venn diagram, what elements make up data science? 

cluster computing, machine learning, and algorithmic development

✓ hacking/programming, math/statistics, and domain expertise

business acumen and social marketing skills

the ability to conduct advanced scientific research in a domain such as biology

✓ Question 4 of 48

People frequently ask whether they should learn Python or R to do data science. What is the answer that this video gives to this question? 

A professional data scientist should rely exclusively on R because it is specialized for data work and is designed for vectorized operations.

A professional data scientist should not use Python or R, but should instead focus on languages like C or Java.

- ✓ A professional data scientist should be comfortable with both Python and R, as well as several other languages used in data science.

A professional data scientists should rely exclusively on Python because it is a general purpose programming language with better memory management than R.

✗ Question 5 of 48

While anomaly detection is normally associated with negative outcomes like fraud or machine failure, it is more flexible than that. Which of the following is a positive outcome in anomaly detection?

imputing values for missing data

inferring from sample data to unknown population parameters

- ✓ identifying new markets with potential value

- ✗ identifying associations between variables

✗ Question 6 of 48

What is application programming interface (API)?

 API translates data and distributes it.

API collects data and analyzes it.

 API routes the data, translates it, and prepares it for use.

API encrypts data for software.

 Question 7 of 48

What are artificial neural networks modeled after?

other artificial intelligence programs

 neurons that are inside a biological brain

expert systems

other data science programs

 Question 8 of 48

Like general "software-as-a-service" or "SaaS," what is one of the major attractions of "machine-learning-as-a-service" or "MLaaS"?

Cloud providers are always cheaper than local hardware.

- ✓ You don't need to own or maintain the hardware, which can be flexibly allocated to your projects as needed.

Cloud processing is always faster than local machines.

Data is inherently more secure on a cloud server than on a local machine.

✗ Question 9 of 48

Which is a major goal of dimensionality reduction?

- ✓ reducing noise in the data

finding natural groupings of cases in the data

classifying new cases into existing categories

- ✗ reducing homogeneity in the data

✗ Question 10 of 48

What is an advantage of in-house data for data science projects?

✓ Certain restrictions on data usage may not apply.

In-house data tends to be well documented throughout the organization.

In-house data is usually well maintained throughout the years.

✗ In-house data is always available for use.

✓ Question 11 of 48

Which is a common application of cluster analysis?

setting the terms of a loan

imputing values for missing data

medical diagnosis

✓ audience segmentation

✓ Question 12 of 48

How do decision trees provide rules for decision-making?

Decision trees are not an effective decision-making tool because the method by which they reach decisions is too difficult to understand.

Decision trees find the observations that best serve as examples for decision-making.

Decision trees uses coefficients to combine multiple input variables into a single output variable; these coefficients can then be used for decision-making.

- Decision trees explicitly define binary decisions at each step in the data to reach the outcome; these decisions can then be used in other contexts.



Question 13 of 48

Do human-unaccessible decisions help data science contribute to the interplay of human and artificial intelligence? 

- No, because human-inaccessible decisions are not a way in which data science contributes to the interplay of human and artificial intelligence.

No, because human-unaccessible allow machines to talk to other machines.

No, because human-unaccessible decisions is where advanced algorithms can make and even implement their own decisions.

✗ No, because human-inaccessible decisions allow the algorithm to process your data and makes a recommendation.

✓ Question 14 of 48

How does linear regression provide rules for decision-making?

Linear regression explicitly involves binary decisions at each step in the data to reach the outcome; these decisions can then be used in other contexts.

✓ Linear regression uses coefficients to combine multiple input variables into a single output variable; these coefficients can then be used for decision making.

Linear regression finds the observations that best serve as examples for decision making.

Linear regression is not an effective decision-making tool because the method by which it reaches its decision is too difficult to understand.

✓ Question 15 of 48

The technique of separating time-series data into an overall trend, a seasonal or cyclical trend, and random variations or noise is known by which term?

machine learning

partitioning of variance

✓ decomposition

analytical modeling

✓ Question 16 of 48

Why have expert systems not progressed as much as machine learning in the development of decision-making systems? 

Expert systems frequently got stuck in infinite loops.

The time needed to get expert feedback on systems was prohibitive.

✓ Expert systems quickly encounter the "combinatorial explosion" in which there are simply too many possibilities to enumerate them all.

Expert systems required too much computing power.

✓ Question 17 of 48

According to many estimates, what percentage of time in a data science project is spent preparing the data? 

0.5

✓ 0.8

0.2

0.05

✓ Question 18 of 48

What is one way of characterizing the relationship between data science and artificial intelligence?



Data science and artificial intelligence are different names for the same thing; they overlap completely.

✓ AI nearly always involves data science, but only some data science projects involve AI.

Data science and AI are separate, distinct fields with no meaningful overlap.

Data science always involves AI, but only some AI projects involve data science.



Question 19 of 48



Which is considered an example of data clustering?

a Social Security number

an e-mail address

✓ marketing segments

prescription glasses



Question 20 of 48



What does it mean for a data science algorithm to be "human-in-the-loop"?

It means that the data for the algorithm was generated by humans.

✓ It means that the algorithm includes humans in the decision-making and implementation such that humans can take over the actions.

It means that the humans can be held legally liable for the results of the algorithm.

It means that humans are directly affected by the algorithm.

✓ Question 21 of 48

What does it mean that algorithms like neural networks develop "implicit" rules?



- ✓ It means that the rules algorithms use cannot be easily described to humans.

It means that the rules can only involve a restricted set of permissible operations.

It means that the rules function only as approximations.

It means that the rules are based on suggestions from the programmers.

✓ Question 22 of 48

In-house data is a type of data that ____ .



always exists

is well maintained

- ✓ is already available in your organization

is well documented

✓ Question 23 of 48

Statistical applications like SPSS or jamovi are useful to data projects in what way? ▼

They are useful for data cleaning and validation.

They serve as flexible interfaces to databases, allowing you to select and import cases quickly.

They are useful for interactive data exploration.

✓ Their point-and-click interfaces make common analyses easier for non-specialists to conduct.

✓ Question 24 of 48

What does it mean when a machine learning algorithm is referred to as a "black box"? ▼

✓ A black box means a process that is hidden from view or difficult to understand.

A black box is a private, proprietary mechanism that is not viewable by the public but that is viewable to its owners.

A black box is a process that paradoxically produces accurate output even though the process is known to be flawed or incomplete.

A black box is a decoy process that secretly channels processing to other, standard algorithms.

✓ Question 25 of 48

Which measure of a distribution's center is like a physical balance point on a number line?



none of these answers

the median

the mode

✓ the mean

✗ Question 26 of 48

What is a method for self-generating data?



the falsification of data in published reports

scraping data from social media posts

 deep learning neural networks

 generative adversarial networks

 Question 27 of 48

Which general step within descriptive statistics allows you to look at one variable at a time? 

measure the association between the variables in your data

 visualize your data

 compute univariate descriptive statistics

collect and analyze your data

 Question 28 of 48

Once you have created and validated a model, what do you need to consider? 

What was the period when you acquired the results and insights from your analysis?

 Who uses the results and the insights that you receive from your analysis?

Where did you get the results from your analysis?

✗ How did you get the results and the insights from your analysis?

✓ Question 29 of 48

Which critical step in predictive models is often neglected?

Model the outcome using any of many possible choices.

Take the model and you apply it to new data.

Find used and relevant past data.

✓ Validate your model by testing it against new data.

✓ Question 30 of 48

What is meant in the video by the phrase "data is for doing"?

✓ It means that data analysis is usually designed with the goal of answering specific questions and guiding actions.

it means that data can only be used to answer operational questions, not theoretical ones.

It means that data must be analyzed.

It means that data analysis is a time-consuming and cognitively-demanding task.



Question 31 of 48

Which is not considered a type of reinforcement learning?



generative adversarial networks

external reinforcement learning

internal reinforcement learning

✓ degenerative adversarial networks



Question 32 of 48

How do data engineers contribute to data science?



Data engineers are people who have extensive work in computer science and in mathematics.

Data engineers are the ones who have the intimate understanding of the algorithms.

Data engineers are the people who work in artificial intelligence.

- ✓ Data engineers are the people who focus on the hardware and the software that make data science possible.

✓ Question 33 of 48

Data science companies should keep in the three major forms of fairness when developing their products and services. What are these forms of fairness?

distribution, procedures, and interactions

care, justice, and retribution

- ✓ equity, equality, and need

sharing, ranking, and pricing

✗ Question 34 of 48

How can you make raw data points useable?

Raw data points can be compared one by one to look for any insight they might give.

- ✓ Raw data points can be combined to create new features in your dataset.

Raw data is unusable.

- ✗ Raw data points can be broken down even further to see how the data was acquired.

✗ Question 35 of 48

Why is algebra important in data science?

is not useful in data science

- ✓ provides scalability and the ability to generalize

- ✗ provides the ability to apply specific parameters to each problem

provides the amount of data that is required to develop projects

✓ Question 36 of 48

What are the elements that constitute "big data"?

variability, vision, and volubility of data

big data refers to any high value dataset collection by a multinational corporation

✓ volume, velocity, and variety of data

coding/programming, math/statistics, and domain expertise

✓ Question 37 of 48

Which are methods used for validating models?

k-means and k-nearest neighbors

lasso regression and ridge regression

✓ holdout testing data and cross-validation testing data

principal component analysis and factor analysis

✓ Question 38 of 48

Calculus is used to answer questions related to which topics?

location and variation

validity and reliability

✓ maximization and minimization

scaling and generalization

✓ Question 39 of 48

APIs, or Application Programming Interfaces, generally serve what function in a data science project?

APIs are the programming environment where you can work with languages like Python or R.

APIs are primarily a means of acquiring data that was not structured for sharing, such as scraping data from images of heat maps.

APIs allow you to use GUIs (graphical user interfaces) in languages that rely on lines of code.

✓ APIs allow you to access data and include it in your data science programming.

✓ Question 40 of 48

One common method of gathering new data is A/B testing. What does this refer to?

A/B testing refers to a form of responsive web design, where a web site's content is reformatted by the kind of device a person is using

to access it.

- ✓ A/B testing is a randomized experiment in which people visiting a website see one of two different versions. The response rates are then compared to choose the most effective version of the website.

A/B testing refers to the inclusion of alternative text for images on a website.

A/B testing refers to the creation of alternative websites that are entirely text-based, for use on poor internet connections.



Question 41 of 48

What is the purpose of aggregating the predictions of multiple models in data science? ▼

By aggregating models, more possibilities are available for



[Download certificate](#)

[Retake exam](#)

[Return to course](#)

By aggregating models, researchers are able to compensate for bias in sampling.

- ✓ By aggregating models, the combined predictions tend to be more accurate and more stable than the individual predictions.

By aggregating models, researchers are able to impute values for missing data.

✓ Question 42 of 48

Data scraping is characterized by all the following, EXCEPT?

You can use programming languages such as Python and R.

Data that is not necessarily created with that integration in mind.

You can use specialized data scraping applications.

✓ Data that is well structured and ready for use.

✓ Question 43 of 48

Why is it productive to aggregate models?

No one research is able to produce a model that is accurate in all situations; models from multiple researchers must be used.

Aggregation overcomes any potential social or cultural bias in training data.

Multiple computers and multiple programming languages must be used to create the different models.

- ✓ Different models tend to overestimate and underestimate their predictions, so the differences frequently cancel out.

✓ Question 44 of 48

What is another name for optimization formulas?

ROI (return-on-investment)

quantitative scenarios

- ✓ mathematical programming

cost-benefit analysis

✓ Question 45 of 48

What rare qualities do data scientists possess that make them so valuable?

the ability to coordinate marketing interactions

the ability to generalize from sample data to population values

the ability to combine multiple statistical models

- ✓ the ability to find order, meaning, and value in unstructured data

✓ Question 46 of 48

Passive collection of data refers to which practice?



The use of surveys in pop-up windows on websites.

- ✓ Gathering data without having to personally create it, as with photo classifications that people provide by tagging their own photos on social media.

Passive collection refers to data collection without identifiers.

The collection of data from digital kiosks at conferences and other public places.

✗ Question 47 of 48

What are some of practical methods for identifying cause-and-effect relationships in your data?



- ✗ all these answers

- ✓ quasi-experiments

the use of data with high volume, velocity, and variety

the use of black-box machine learning algorithms

✓ Question 48 of 48

What does Bayes' theorem allow you to do?

✓ It allows you to incorporate prior probabilities into your analysis.

It does not allow you to incorporate an element of probability in your data analysis.

It allows you route data, translate it, and get it ready for use.

It does not deal with data science, instead it deals with ethical dilemmas.

[Return to course](#)

English (English) ▾

About

Become an Instructor

Buy for my team

Help

Privacy & Terms ▾

Accessibility

Apps ▾

