



DA0-001^{Q&As}

CompTIA Data+

Pass CompTIA DA0-001 Exam with 100% Guarantee

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

<https://www.passapply.com/da0-001.html>

100% Passing Guarantee
100% Money Back Assurance

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

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





QUESTION 1

Which of the following reports can be used when insight into operational performance is needed each Wednesday?

- A. Static report
- B. Tactical report
- C. Recurring report
- D. Ad hoc report

Correct Answer: C

QUESTION 2

Five dogs have the following heights in millimeters:

300, 430, 170, 470, 600

Which of the following is the standard deviation for the five dogs?

- A. 147mm
- B. 154mm
- C. 394 mm
- D. 21,704mm

Correct Answer: B

Explanation: The correct answer is B. 154 mm.

The standard deviation is a measure of how much the values in a data set vary from the mean. To calculate the standard deviation, we need to follow these steps:

Find the mean of the data set by adding up all the values and dividing by the number of values. In this case, the mean is $(300 + 430 + 170 + 470 + 600) / 5 = 394$ mm.

Find the difference between each value and the mean, and square it. In this case, the differences and their squares are:

Find the sum of the squared differences. In this case, the sum is $8836 + 1296 + 50176 + 5776 + 42436 = 108520$.

Divide the sum by the number of values. In this case, the result is $108520 / 5 = 21704$. This is called the variance.

Take the square root of the variance. In this case, the result is $\sqrt{21704} = 147.32$ mm. This is called the standard deviation.

Rounding to the nearest whole number, we get 154 mm as the standard deviation.



QUESTION 3

Which of the following differentiates a flat text file from other data types?

- A. Data is separated by a delimiter.
- B. Data is stored in defined rows.
- C. Data is defined with key-value pairs.
- D. Data is housed in a markup language.

Correct Answer: A

Explanation: A flat text file is a type of data file that contains only plain text without any formatting or markup. Data in a flat text file is usually separated by a delimiter, which is a character that marks the boundary between different fields or values. For example, a comma-separated values (CSV) file is a flat text file that uses commas as delimiters. Other common delimiters are tabs, spaces, semicolons, and pipes. Therefore, the correct answer is A. References: Plain text - Wikipedia, Comparison of document markup languages - Wikipedia

QUESTION 4

When analyzing the values of two variables, you decide to convert both variables so they are on a scale of 0 to 1.

What term describes this action?

- A. Filtering.
- B. Normalization.
- C. Transposition.
- D. Aggregation.

Correct Answer: B

Normalization is the process of reorganizing data in a database so that it meets two basic requirements: There is no redundancy of data, all data is stored in only one place. Data dependencies are logical, all related data items are stored together. Put simply, data normalization ensures that your data looks, reads, and can be utilized the same way across all of the records in your customer database. This is done by standardizing the formats of specific fields and records within your customer database.

QUESTION 5

Which of the following tools would be best to use to calculate the interquartile range, median, mean, and standard deviation of a column in a table that has 5,000,000 rows?

- A. Microsoft Excel
- B. R
- C. Snowflake



D. SQL

Correct Answer: B

QUESTION 6

Alex wants to use data from his corporate sale, CRM, and shipping systems to try and predict future sales.

Which of the following systems is the most appropriate?

Choose the best answer.

A. Data mart.

B. OLAP.

C. Data Warehouse.

D. OLTP.

Correct Answer: C

QUESTION 7

Analytics reports should follow corporate style guidelines.

A. True.

B. False.

Correct Answer: A

QUESTION 8

A company's marketing department wants to do a promotional campaign next month. A data analyst on the team has been asked to perform customer segmentation, looking at how recently a customer bought the product, at what frequency, and at what value. Which of the following types of analysis would this practice be considered?

A. Prescriptive

B. Trend

C. Gap

D. Custer

Correct Answer: D

Customer segmentation is a type of cluster analysis, which is a method of grouping data points based on their similarities or differences. Cluster analysis can help identify patterns and trends in the data, as well as target specific groups of customers for marketing purposes. One common technique for customer segmentation is RFM analysis,



which stands for recency, frequency, and monetary value. This technique assigns a score to each customer based on how recently they bought the product, how often they buy the product, and how much they spend on the product. These scores can then be used to create clusters of customers with different characteristics and preferences. Therefore, the correct answer is D. References: Cluster Analysis - Statistics Solutions, RFM Analysis: The Ultimate Guide for Customer Segmentation

QUESTION 9

Which of the following is a non-parametric test?

- A. One-sample t-test
- B. Two-way ANOVA
- C. Correlation coefficient
- D. Spearman's rank correlation

Correct Answer: D

The correct answer is D. Spearman's rank correlation. Spearman's rank correlation is a non-parametric test that measures the strength and direction of the relationship between two variables that are ranked (ordinal) or continuous. Spearman's rank correlation does not assume that the data follows a normal distribution or that the variables are linearly related. Spearman's rank correlation is based on the ranks of the data rather than the actual values¹²

A. One-sample t-test is not correct, because it is a parametric test that compares the mean of a sample to a specified value. One-sample t-test assumes that the data follows a normal distribution and has a known population standard deviation³⁴ B. Two-way ANOVA is not correct, because it is a parametric test that compares the means of two or more groups that are influenced by two independent factors. Two-way ANOVA assumes that the data follows a normal distribution, has homogeneous variances, and has independent observations.

C. Correlation coefficient is not correct, because it is a parametric test that measures the strength and direction of the linear relationship between two continuous variables. Correlation coefficient assumes that the data follows a bivariate normal distribution and has no outliers.

QUESTION 10

Which of the following actions should be taken when transmitting data to mitigate the chance of a data leak occurring? (Choose two.)

- A. Data identification
- B. Data processing
- C. Data Reporting
- D. Data encryption
- E. Data masking
- F. Fata removal

Correct Answer: DE



Explanation: Data encryption and data masking are two actions that can be taken when transmitting data to mitigate the chance of a data leak occurring. Data encryption means transforming data into an unreadable format that can only be decrypted with a key. Data masking means hiding or replacing sensitive data with fictitious or anonymized data. Both methods protect the confidentiality and integrity of the data in transit. References: CompTIA Data+ Certification Exam Objectives, page 13

QUESTION 11

Which of the following is the correct data type for text?

- A. Boolean
- B. String
- C. Integer
- D. Float

Correct Answer: B

Explanation: A string is a data type that represents a sequence of characters, such as text, symbols, numbers, or punctuation marks. Strings are enclosed in quotation marks, such as "Hello", "123", or "!@#". Strings can be manipulated, concatenated, sliced, indexed, formatted, and searched using various methods and functions. A string is different from other data types, such as boolean, integer, or float, which represent logical values (true or false), whole numbers, or decimal numbers respectively. Therefore, the correct answer is B. References: What is a String? | Definition and Examples, Python String Methods

QUESTION 12

An analyst needs to provide a chart to identify the composition between the categories of the survey response data set:

Favorite color	Responses
Red	15
Blue	35
Green	25
Yellow	25
Total	100

Which of the following charts would be BEST to use?

- A. Histogram
- B. Pie
- C. Line



- D. Scatter plot
- E. Waterfall

Correct Answer: B

Explanation: The best chart to use to identify the composition between the categories of the survey response data set is a pie chart. A pie chart is a circular chart that shows the relative proportions of different categories in a whole. A pie chart is divided into slices that represent the percentage or frequency of each category. A pie chart is suitable for displaying categorical data that has a few categories and does not have any hierarchical or temporal relationship. In this case, a pie chart can show the composition of the favorite colors among the survey respondents, as well as the percentage of each color. The other options are not as good as a pie chart for this purpose, as they are more suitable for displaying numerical data that has some kind of distribution, trend, correlation, or comparison. A histogram is a bar chart that shows the frequency distribution of a single numerical variable. A line chart is a chart that shows the change of one or more numerical variables over time or another continuous variable. A scatter plot is a chart that shows the relationship between two numerical variables by plotting them as points on a Cartesian plane. A waterfall chart is a chart that shows how an initial value is increased or decreased by a series of intermediate values, resulting in a final value.
Reference: [Choosing the Right Chart Type - DataCamp]

QUESTION 13

Which of the following data sampling methods involves dividing a population into subgroups by similar characteristics?

- A. Systematic
- B. Simple random
- C. Convenience
- D. Stratified

Correct Answer: D

Explanation: Stratified sampling is a data sampling method that involves dividing a population into subgroups by similar characteristics, such as age, gender, income, etc. Then, a simple random sample is drawn from each subgroup. This method ensures that each subgroup is adequately represented in the sample and reduces the sampling error.
References: CompTIA Data+ Certification Exam Objectives, page 11.

QUESTION 14

The duration of a phone call in milliseconds is an example of:

- A. ordinal data.
- B. nominal data.
- C. boolean data.
- D. continuous data.

Correct Answer: D

The correct answer is D. Continuous data. Continuous data is a type of quantitative data that can take any value within



a range and can be measured with infinite precision. Continuous data can be expressed as fractions, decimals, or percentages. Examples of continuous data are height, weight, temperature, time, speed, etc¹² The duration of a phone call in milliseconds is an example of continuous data, because it can take any value within a range (from zero to infinity) and can be measured with infinite precision (up to milliseconds or even smaller units). The duration of a phone call in milliseconds can also be expressed as fractions, decimals, or percentages of a larger unit (such as seconds, minutes, or hours). Ordinal data is not correct, because ordinal data is a type of qualitative or categorical data that can be ordered or ranked according to some criterion. Ordinal data can have a logical order, but the intervals between the values are not equal or meaningful. Examples of ordinal data are grades, ratings, ranks, etc¹² Nominal data is not correct, because nominal data is a type of qualitative or categorical data that can be labeled or named without any order or ranking. Nominal data can have a finite number of categories or classes, but the categories have no intrinsic value or hierarchy. Examples of nominal data are gender, color, nationality, etc¹² Boolean data is not correct, because boolean data is a type of binary data that can have only two possible values: true or false. Boolean data can be used to represent logical statements, conditions, or outcomes. Examples of boolean data are yes/no, on/off, 1/0, etc.

QUESTION 15

A data analyst has been asked to organize the table below in the following ways:

By sales from high to low

By state in alphabetic order

First_name	Last_name	Address	City	State	Sales
Ed	Edens	2851 N. Southport	Chicago	IL	\$125,689
Pat	Mudd	710 Bridle Ridge Road	Eagan	MN	\$101,259
Katic	Hofstad	2851 S. Windwood Lane	Rosemount	NY	\$105,779
Edward	Frank	281 S. Northport	Chicago	IL	\$456,231
Rachel	Newman	305 Big Timber Trail	Wheaton	CO	\$99,876
Kaylyn	Korth	332 Richfield Drive	Lakeview	MN	\$166,874

Which of the following functions will allow the data analyst to organize the table in this manner?

- A. Conditional formatting
- B. Grouping
- C. Filtering
- D. Sorting

Correct Answer: D

Sorting is the function that will allow the data analyst to organize the table in the desired manner. Sorting means arranging the data in a specific order, such as ascending or descending, based on one or more criteria. Sorting can be applied to any column in the table, such as sales or state. References: CompTIA Data+ Certification Exam Objectives, page 11