BCS Level 4 Certificate in Data Analysis Tools
QAN 603/0824/2

Sample Paper A

Record your surname / last / family name and initials on the answer sheet.

Sample paper only 40 multiple-choice questions – 1 mark awarded to each question. Mark only one answer to each question. There are no trick questions.

A number of possible answers are given for each question, indicated by either **A. B. C. or D**. Your answers should be clearly indicated on the answer sheet.

Pass mark is 26/40.

Copying of this paper is expressly forbidden without the direct approval of BCS, The Chartered Institute for IT.

This qualification is regulated by Ofqual (in England).

1   During data profiling, what type of analysis is used to determine how often a given attribute is populated, versus blank or null?

A   Uniqueness.
B   Range.
C   Pattern.
D   Completeness.

2   If a query is nested inside a SELECT, INSERT, UPDATE, or DELETE statement, what is this called?

A   Prequery.
B   Subquery.
C   Inner query.
D   Outer query.

3   In a distribution that is positively skewed, which of the following is **TYPICALLY TRUE**?

A   The standard deviation is greater than the range.
B   The median is greater than the mean.
C   The median and mean are equal.
D   The mean is greater than the median.

4   How is data organised in an Excel spreadsheet?

A   Lines and spaces.
B   Layers and planes.
C   Height and width.
D   Rows and columns.

5   In R programming, how is a missing value represented?

A   NULL
B   #
C   %
D   NA

6   In SQL, if a column of a dataset lacks any value, which function can be used to set the result to a specific value?

A   Round
B   Max
C   Convert
D   ISNULL

7   R can easily be extended via what?

A   Models.
B   Packages.
C   Slides.
D   Pockets.

8   The following are examples of which data type?

    • Photos and graphic images.
    • Videos.
    • Streaming instrument data.
    • Webpages.
    • PDF files.
    • PowerPoint presentations.

A   Semi-structured.
B   Unstructured.
C   Structured.
D   Multi-structured.

9   The process of establishing and maintaining consistency among systems and subsequent continuous updates, is known as what?

A   Data integration.
B   Data synchronisation.
C   Data validation.
D   Data optimisation.

10   What do functional requirements describe?

A   The speed at which a system must work.
B   What the system must do.
C   If the system will limit access to authorised users.
D   If the system will accept numeric data entry.

11   What is the median in the following set of data?

       1, 2, 2, 3, 5, 6, 9

A   2.
B   3.
C   4.
D   5.

12   What measure will always change if a single value in a dataset changes?

A   Median.
B   Mode.
C   Mean.
D   All of these.

13   What type of join is used to combine the result-set of two or more SELECT statements?

A   Full.
B   Outer.
C   Union.
D   Inner.

14   What will be reduced if multiple disparate datasets are integrated?

A   Data availability.
B   Data collaboration.
C   Data complexity.
D   Data integrity.

15 When implementing data integration techniques, what does ETL stand for?

A Extract, Transform, Load.
B Export, Transfer, Loop.
C Extract, Transfer, Load.
D Export, Transform, Loop.

16 When preparing data for analysis to improve accuracy, quality and usefulness, which of the following are **NOT** suitable ways to respond to an error?

    a) Accept.
    b) Correct error.
    c) Ignore.
    d) Create null value.

A a and b only.
B b and c only.
C a and d only.
D c and d only.

17 Which of the following describes managing an organisation's critical data, providing a single point of reference?

A Important Data Management.
B Master Data Management.
C Vital Data Management.
D Essential Data Management.

18 Which command is used to select only **one** copy of each set of duplicate rows in SQL?

A SELECT DIFFERENT
B SELECT NONE
C SELECT DISTINCT
D SELECT ALL

19 Which of the following are data quality dimensions?

    a) Uniqueness.
    b) Timeliness.
    c) Consistency.
    d) Cleanliness.

A   a and b only.
B   c and d only.
C   a, b and c only.
D   b, c and d only.

20 Which of the following are user interfaces?

    a) Data dashboard.
    b) Performance scorecard.
    c) Data transparency.
    d) Dynamic charts.

A   a and b only.
B   c and d only.
C   b, c and d only.
D   a, b and d only.

21 Which of the following describes User Acceptance Testing (UAT)?

A   Testing by users of software against real-world scenarios.
B   Testing by users of hardware components of a network.
C   Testing by users of peripheral devices (e.g. a mouse or keyboard).
D   Testing by users of the effect difficult conditions have on software.

22 Which of the following is a monthly report, which uses dynamic charts and tables to summarise a company's sales turnover?

A   Dashboard.
B   Graph.
C   Screen.
D   Data.

23 Dates in a column of a spreadsheet are an example of which type of data?

A Unstructured.
B Semi-structured.
C Structured.
D Multi-structured.

24 Which of the following is **NOT** a core data type in Python?

A Class.
B Lists.
C Dictionary.
D Tuples.

25 Which of the following is **NOT** a reason for developing appropriate testing strategies?

To ensure that unified data sets are…

A Correct.
B Complete.
C Legal.
D Up to date.

26 Which of the following is **NOT** a type of join?

A Left.
B Right.
C Select into.
D Insert into.

27 Which of the following is **NOT** an option to accomplish data migration?

A Merge two systems into a brand new one.
B Merge one system into another.
C Remove one of two merged systems from another, without further action.
D Create a common view on top of two separate systems.

28   Which SQL statement is often used with aggregate functions to assemble the result-set by one or more columns?

A   Count
B   Order by
C   Group by
D   Round

29   Which technique can be used to unveil the content and structure of a data set?

A   Data modelling.
B   Data profiling.
C   Data processing.
D   Data describing.

30   Which testing process is used to validate the resource usage, scalability, and reliability of a product?

A   Owner acceptance testing.
B   Technical acceptance testing.
C   User acceptance testing.
D   Performance stress testing.

31   Which of the following is the **MOST LIKELY** reason for integrating data from separate sources?

A   To allow analysis only possible after integration.
B   To provide access to data for GDPR legislation.
C   To reduce the amount of storage space required.
D   To allow for the normalisation of complex data.

32   When integrating data which SQL statement is typically used to extract data?

A   Insert.
B   Select.
C   Delete.
D   Update.

33   What is the **MOST LIKELY** purpose of a data integration dashboard?

A   Display metrics representing the progress of data amalgamation.
B   Provide a common interface to manage the processing.
C   Display key metrics about the contents of the database.
D   Provide common tools used to integrate data sources.

34   How would the following SQL statement assist with data integration?

SELECT [Name], [City], [Country]
FROM [AcmeUK-Customers]
UNION
SELECT [Name], [City], [Country]
FROM [AcmeEU-Customers]

A   Display data common to AcmeUK-Customers and AcmeEU-Customers.
B   Display data in AcmeUK-Customers but not in AcmeEU-Customers.
C   Display data in AcmeEU-Customers but not in *AcmeUK-Customers.*
D   Display data in AcmeEU-Customers and in AcmeUK-Customers.

35   How is the validity of data checked?

A   Check that data conforms to the correct format, type and range.
B   Check all data that should have been collated is stored.
C   Check that data is not recorded more than once.
D   Check that data is stored in a standard way for all data elements.

36   Which of the following is a VALID explanation of the challenge of processing large volumes of data from social media?

A   Large volumes of structured data often require the use of specialised 'big data' tools.
B   Large volumes of unstructured data often require the use of specialised 'big data' tools.
C   Large volumes of structured data are best processed using a large on-premise SQL Databases.
D   Large volumes of unstructured data are best processed using a large on-premise SQL Database.

37   Why is it important that testing data sets are complete, current and accurate?

A   To ensure that it is not possible to crash a system with invalid data.
B   So that any data capable of crashing a system is used for testing.
C   To ensure that tests performed meet current legislative requirements.
D   So that tests performed using test data accurately reflects the real world.

38   Which statistical programming language with built in functions would be **BEST** suited to calculate standard deviation, variance and t-tests.

A   R.
B   Python.
C   SQL.
D   Graph.

39   Which tool would be best suited to quickly sort very large volumes of structured data into alphabetical or
A   R.
B   Python.
C   SQL.
D   UIMA.

40   How is statistical significance used within data analysis?

A   Provides a measure of certainty that a result is not a random chance.
B   Provides a measure of the effect size of changing dependent variables.
C   Provides a result that is very important to the business.
D   Provides a result that is very important to a customer.

-End of Paper-