Disclaimer: These are sample test questions of what a competitor will find in this competition. To view the official study guide, with the knowledge area and objective list, visit the guidelines on the Competitive Events page of the national website at www.fbla.org.

Probability and Statistics Foundations

1. Which measure of central tendency is most affected by extreme values?
   a) Mean
   b) Median
   c) Mode
   d) Range

2. The probability of two independent events both occurring is found by:
   a) Adding their probabilities
   b) Multiplying their probabilities
   c) Subtracting their probabilities
   d) Dividing their probabilities

3. A normal distribution is defined by its:
   a) Mean and mode
   b) Mean and standard deviation
   c) Median and range
   d) Variance and mode

4. In probability, the complement of an event A represents:
   a) The event happening twice
   b) The event not occurring
   c) The event occurring first
   d) The event occurring with another event

5. The standard deviation measures:
   a) How centered the data is around the mean
   b) The spread of data values from the mean
   c) The probability of an event occurring
   d) The relationship between two variables

6. If the probability of rain is 0.3, what is the probability it will NOT rain?
   a) 0.3
   b) 0.5
   c) 0.7
   d) 0.9

7. Which of the following is a continuous random variable?
   a) Number of books on a shelf
   b) Height of students in a class
   c) Number of defective products
   d) Count of website visits

8. In statistics, a p-value is used to:
   a) Determine if a result is statistically significant
   b) Measure the average of a data set
   c) Calculate the mean difference
   d) Determine probability of independence

Data Analysis and Statistics for AI

9. In AI, overfitting occurs when:
   a) The model is too simple
   b) The model learns noise instead of patterns
   c) The model is under-trained
   d) The dataset is too small

10. Which measure is best for comparing the spread of two datasets with different units?
    a) Standard deviation
    b) Coefficient of variation
    c) Mean
    d) Median

11. A correlation coefficient of -0.85 suggests:
    a) A strong positive relationship
    b) A weak negative relationship
    c) A strong negative relationship
    d) No relationship

12. Which is NOT a step in the data analysis process?
    a) Data collection
    b) Data cleaning
    c) Model deployment
    d) Data visualization

13. The confusion matrix is used to evaluate:
    a) Regression models
    b) Classification models
    c) Clustering algorithms
    d) Data cleaning methods

14. A histogram is most useful for:
    a) Showing frequency distribution of data
    b) Comparing two datasets
    c) Showing relationships between variables
    d) Plotting real-time changes

15. Which metric is most appropriate for measuring model accuracy in imbalanced datasets?
    a) Accuracy
    b) Precision and recall
    c) Mean absolute error
    d) Root mean square error

16. Data normalization is performed to:
    a) Remove duplicate data
    b) Scale features to a similar range
    c) Convert categorical data to numerical data
    d) Remove outliers

Tools for Data and AI

17. Which programming language is most widely used for AI and data science?
    a) Java
    b) Python
    c) C++
    d) Ruby

18. SQL is primarily used for:
    a) Managing and querying relational databases
    b) Creating machine learning models
    c) Designing websites
    d) Running operating systems

19. Which Python library is best known for data visualization?
    a) NumPy
    b) Matplotlib
    c) Pandas
    d) TensorFlow

20. TensorFlow is primarily used for:
    a) Database management
    b) Deep learning model development
    c) Data cleaning
    d) Statistical tests

21. Which tool is best suited for interactive data analysis and visualization in a browser?
    a) Jupyter Notebook
    b) Apache Hadoop
    c) GitHub
    d) Tableau Server

AI Basics

22. Artificial Intelligence is best defined as:
    a) Computers mimicking human intelligence tasks
    b) Computers storing large amounts of data
    c) Faster internet processing
    d) Automated backups

23. Which is an example of narrow AI?
    a) Self-driving car software
    b) General AI with human-like reasoning
    c) An AI that can perform all intellectual tasks
    d) Human brain emulation

24. Turing Test is designed to:
    a) Measure AI speed
    b) Evaluate if a machine exhibits human-like intelligence
    c) Test internet bandwidth
    d) Measure algorithm efficiency

25. Which branch of AI focuses on enabling machines to interpret visual information?
    a) Robotics
    b) Computer vision
    c) Natural language processing
    d) Neural networks

26. Which is NOT a goal of AI?
    a) Learning
    b) Reasoning
    c) Forgetting
    d) Problem-solving

Machine Learning

27. Supervised learning requires:
    a) Labeled training data
    b) Unlabeled data
    c) Data clustering
    d) No data

28. In machine learning, a feature is:
    a) An input variable used for prediction
    b) The predicted output
    c) A hidden layer
    d) A training algorithm

29. Which algorithm is used for classification tasks?
    a) Linear regression
    b) Decision trees
    c) K-means clustering
    d) Apriori algorithm

30. Unsupervised learning is typically used for:
    a) Predicting stock prices
    b) Grouping similar items without labels
    c) Predicting house prices
    d) Fraud detection

31. Overfitting in machine learning can be reduced by:
    a) Increasing the number of features
    b) Regularization techniques
    c) Using a smaller training dataset
    d) Ignoring validation data

Perception, Representation, and Reasoning

32. Knowledge representation in AI refers to:
    a) Storing knowledge in a form that computers can process
    b) Visualizing data
    c) Creating animations
    d) Encoding passwords

33. Which is an example of an AI perception system?
    a) Image recognition software
    b) SQL database
    c) CPU processing unit
    d) Data warehouse

34. An expert system uses:
    a) Human expertise to make decisions
    b) Predefined rules and reasoning to solve problems
    c) Only statistical models
    d) Random guesswork

35. Semantic networks are used to:
    a) Represent relationships between concepts
    b) Encrypt data
    c) Build neural networks
    d) Store images

36. Which reasoning approach draws conclusions from general principles to specific cases?
    a) Inductive reasoning
    b) Deductive reasoning
    c) Abductive reasoning
    d) Predictive reasoning

Privacy and Ethics

37. Which is an example of bias in AI?
    a) Algorithmic decisions favoring one demographic group unfairly
    b) An AI predicting stock prices
    c) An AI with low accuracy
    d) Slow training times

38. GDPR primarily focuses on:
    a) Cybersecurity hardware installation
    b) Data privacy and protection in the EU
    c) Machine learning algorithms
    d) Software development standards

39. Which principle ensures individuals know how their data will be used?
    a) Transparency
    b) Confidentiality
    c) Availability
    d) Integrity

40. Ethical AI aims to:
    a) Maximize profits
    b) Make decisions aligned with societal values
    c) Avoid legal compliance
    d) Reduce computing costs

41. Which is NOT a best practice for ethical AI?
    a) Inclusive training datasets
    b) Regular bias audits
    c) Hiding algorithm decision-making
    d) Clear accountability

Data Literacy and Foundations

42. Data literacy is best defined as:
    a) The ability to read, work with, analyze, and communicate data
    b) Writing computer programs
    c) Building neural networks
    d) Encrypting data

43. Structured data is:
    a) Organized in a fixed format, like databases
    b) Unorganized text and images
    c) Only numbers
    d) Encrypted data

44. Which is an example of qualitative data?
    a) Temperature readings
    b) Customer satisfaction ratings in words
    c) Age of participants
    d) Monthly revenue

45. A primary key in a database:
    a) Uniquely identifies a record
    b) Encrypts the database
    c) Links two databases
    d) Stores large files

46. Which is NOT a step in the data lifecycle?
    a) Data collection
    b) Data storage
    c) Data destruction
    d) Data advertising

47. Which term refers to data that arrives in real-time streams?
    a) Batch data
    b) Streaming data
    c) Historical data
    d) Stale data

48. Data cleaning aims to:
    a) Remove or correct inaccurate data
    b) Increase dataset size
    c) Hide data from users
    d) Encrypt sensitive information

49. Metadata is:
    a) Data about data
    b) A type of AI algorithm
    c) A database query
    d) An encryption standard

50. Which is an example of unstructured data?
    a) Customer emails
    b) Spreadsheet with sales figures
    c) SQL database entries
    d) Inventory list

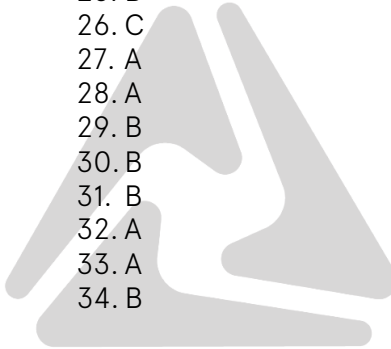**Answer Key**

1.	A	18.	A	35.	A
2.	B	19.	B	36.	B
3.	B	20.	B	37.	A
4.	B	21.	A	38.	B
5.	B	22.	A	39.	A
6.	C	23.	A	40.	B
7.	B	24.	B	41.	C
8.	A	25.	B	42.	A
9.	B	26.	C	43.	A
10.	B	27.	A	44.	B
11.	C	28.	A	45.	A
12.	C	29.	B	46.	D
13.	B	30.	B	47.	B
14.	A	31.	B	48.	A
15.	B	32.	A	49.	A
16.	B	33.	A	50.	A
17.	B	34.	B		