Studybeesofficial
EDUCATIONAL PLATFORM
www.studybeesofficial.com

HSSC-I NATIONAL BOOK FOUNDATION FBISE

COMPUTER TEST - CHAPTER 4: DATA AND ANALYSIS

Name:		Roll No.		Section:

Total Time: 2:30 Hour
Time Allowed: 150 Minutes

Total Marks: 150

PAPER OBJECTIVE
SECTION-A (MARKS 150)

NOTE: Section-A is compulsory. All parts of this section are to be answered on these pages and handed over to the center superintendent. Deleting / overwriting is not allowed. Do not use lead pencil.

Q.1 Fill the relevant bubble for each part. All parts carry one mark.

Questions	A	B	C	D	A B C D
1. Data analysis is a process where data is cleaned, explored and:	Printed	Encrypted	Deleted	Summarized	○○○●
2. Data analysis helps to reduce the risks in:	Decision-making	Typing	Printing	Booting	●○○○
3. Data analysis results are often presented as charts, images, tables and:	Viruses	Cables	Graphs	Passwords	○○●○
4. Statistical modeling uses mathematical models and statistical assumptions to:	Repair hardware	Draw pictures only	Store passwords	Generate sample data and make predictions	○○○●
5. A statistical model is a collection of:	Programming languages	Network protocols	Probability distributions of all possible outcomes	Hardware parts	○○●○
6. In the relationship y = 5x + 2, variable y is the:	Dependent variable	Operator	Constant	Independent variable	●○○○
7. In y = 5x + 2, variable x is the:	Independent variable	Dependent variable	Result	Output	●○○○
8. Statistical modeling is described as the mathematical relationship between:	Two networks	Two printers	Two files	Random and non-random variables	○○○●
9. A use case is a methodology used in:	System analysis	Typing	Hardware repair	Printing	●○○○
10. A data science use case is a concrete real-world task solved using:	Available data	Guesses only	Paper only	Hardware only	●○○○
11. KPI stands for:	Key Performance Indicators	Keyboard Power Index	Known Program Instructions	Key Process Input	●○○○
12. In use case planning, KPIs are defined as a measure of:	Failure	Cost only	Speed only	Success	○○○●

StudyBees OFFICIAL PAK

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio

Questions	A	B	C	D	A B C D
13. Which is a common approach to solve data science use cases?	Soldering	Forecasting	Formatting	Typing	○●○○
14. Which of the following is a common approach in data science use cases?	Screen cleaning	Pattern and anomaly detection	Cable crimping	Disk defragmentation	○●○○
15. Customer churn rate prediction is an example of a typical:	Typing task	Printing task	Hardware task	Data science use case	○○○●
16. Fraud detection and price optimization are:	Typical data science use cases	Network protocols	Hardware parts	Programming languages	●○○○
17. The first step of the data science roadmap is:	Data wrangling	Data collection	Formulating the right question	Result communication	○○●○
18. The second step of the data science roadmap is:	Data wrangling	Data analysis and modeling	Result communication	Data collection	○○○●
19. The third step of the data science roadmap is:	Formulating the question	Data collection	Result communication	Data wrangling	○○○●
20. In data wrangling, raw data is:	Deleted	Encrypted	Printed	Cleaned, preprocessed and transformed	○○○●
21. The last step of the data science roadmap is:	Result communication	Formulating the question	Data collection	Data wrangling	●○○○
22. Sharing findings and conclusions with management and shareholders is:	Result communication	Modeling	Data wrangling	Data collection	●○○○
23. Weather forecasting is helpful for aviation, agriculture, mountaineering and:	Businesses	Printing	Soldering	Typing	●○○○
24. Predictive analysis of stock market data helps:	Only cables	Only printers	Nobody	Investors and consumers	○○○●
25. Analysis of population data helps governments in:	Typing	Printing money	Cabling	Planning and distribution of resources	○○○●
26. Natural disaster prediction helps people take:	Typing steps	Coding steps	Printing steps	Preventive measures	○○○●
27. In the weather case study, which tool measures rainfall?	Compass	Barometer	Rain gauge	Thermometer	○○●○
28. In the weather case study, which tool measures barometric pressure?	Compass	Barometer	Rain gauge	Thermometer	○●○○
29. In the weather case study, which tool is used for wind direction?	Barometer	Compass	Rain gauge	Thermometer	○●○○

StudyBees OFFICIAL PAK

Questions	A	B	C	D	A B C D
30. During data cleaning, duplicate values are removed and missing values are:	Printed	Doubled	Hidden	Imputed	○○○●
31. In supervised learning, the algorithm uses a:	Empty dataset	Unlabeled dataset only	Random dataset	Labeled dataset	○○○●
32. Supervised learning techniques in statistical modeling include:	Printing and typing	Regression and classification	Encoding and decoding	Clustering and association	○●○○
33. If the outcome of a model is a continuous value, it is called:	Clustering	Association	Regression	Classification	○○●○
34. The most common regression model is the:	Random model	Linear model	Circular model	Tree model	○●○○
35. In simple linear regression, the equation of the line is:	y = x / b	y = mb	y = mx + b	y = m + b + x	○○●○
36. In y = mx + b, m is the:	Constant output	Intercept	Slope	Dependent variable	○○●○
37. In y = mx + b, b is the:	y-intercept	Independent variable	Slope	Run	●○○○
38. The y-intercept is the point where the graph line meets the:	y-axis	Origin only	Top of page	x-axis	●○○○
39. Slope is calculated as:	Rise / run	Run / rise	Rise + run	Rise x run	●○○○
40. In a regression model, the variable used for prediction is the:	Dependent variable (y)	Independent variable (x)	Label	Constant	○●○○
41. If the outcome of a model is a discrete value, it is called:	Regression	Slope	Intercept	Classification	○○○●
42. Predicting whether an employee will get a raise or not is an example of:	Clustering	Classification	Regression	Association	○●○○
43. Predicting how much salary raise an employee will get is an example of:	Classification	Regression	Association	Clustering	○●○○
44. In unsupervised learning, the algorithm is given:	No data	Unlabeled data	An answer key	Labeled data	○●○○
45. Which of the following is an example of unsupervised learning?	Sorting by hand	Clustering	Regression	Classification with labels	○●○○
46. Clustering means grouping data items with:	Maximum similarities	Maximum differences	Random labels	Equal sizes	●○○○

StudyBees OFFICIAL PAK

Questions	A	B	C	D	A B C D
47. Association in unsupervised learning means how likely a second action is when the:	First action is done	User is absent	Data is deleted	Computer is off	●○○○
48. Finding that customers who buy bread are likely to buy milk is an example of:	Slope	Classification	Association	Regression	○○●○
49. K-means clustering combines a specified number of data points into groups based on:	Differences	Similarities	Colours	Sizes	○●○○
50. Which of the following can be used to build statistical models?	Only calculators	MS-Excel, Weka, R Studio and Python	Only paper	Only printers	○●○○
51. Popular platforms to download datasets include:	Notepad and Paint	Kaggle and GitHub	Word and Excel	Winzip and WinRAR	○●○○
52. The online Python environment used in the book is:	Google Colab	Notepad	Calculator	Paint	●○○○
53. In the Python example, a linear relationship like y = 3x + 4 can be used to generate:	A password	A virus	A random dataset	A network	○○●○
54. Which library is used in the example to draw plots?	turtle	matplotlib	tkinter	pygame	○●○○
55. Experimental design is about precise planning to prevent:	Hardware failure	Power failure	Erroneous conclusions	Typing mistakes	○○●○
56. Experimental design is used in engineering, psychology, agriculture and:	Printing only	Medicine	Sports only	Cooking only	○●○○
57. An experiment produces precise and trustworthy results if you select the right:	Font	Printer	Keyboard	Sample size	○○○●
58. In experimental design, the first step is to define the:	Research question	Report	Result	Conclusion	●○○○
59. Hypotheses clarify the differences and correlations between:	Files	Printers	Networks	Variables	○○○●
60. In experimental design, fixed variables are called:	Confounding variables	Random variables	Independent variables	Dependent variables	○○●○
61. In 'salary depends on years of experience', the independent variable is:	Age	Company name	Salary	Years of experience	○○○●
62. In 'salary depends on years of experience', the dependent variable is:	Salary	Years of experience	Skill type	Gender	●○○○

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio

Questions	A	B	C	D	A B C D
63. Factors like skill type that may also affect the result are called:	Confounding factors	Constants	Labels	Independent variables	●○○○
64. Factorial and randomized block designs are examples of:	Hardware devices	Programming languages	Data types	Experimental designs	○○○●
65. Random assignment and selection helps to ensure there is no:	Data	Result	Bias	Variable	○○●○
66. ANOVA stands for:	Analysis of vectors	Analysis of values	Analysis of variance	Average of variance	○○●○
67. ANOVA is a statistical test for detecting differences in:	File sizes	Group means	Spellings	Hardware	○●○○
68. The 'Discuss and Report' step ensures reproducibility and:	Transparency	Speed	Secrecy	Confusion	●○○○
69. The principle that divides people into treatment groups at random is:	Blocking	Randomization	Local control	Replication	○●○○
70. A control group is a group that has:	Received the new treatment	Received all treatments	Not received any treatment	Been deleted	○○●○
71. The principle that uses a control group for comparison is:	Replication	Randomization	Blocking	Local control	○○○●
72. Grouping participants by a trait that could affect results is called:	Control	Randomization	Replication	Blocking	○○○●
73. In the blocking example, 56 women are divided into two groups of:	14 each	56 each	20 each	28 each	○○○●
74. In the blocking example, 44 men are divided into two groups of:	22 each	11 each	20 each	44 each	●○○○
75. Repeating an experiment again and again is called:	Randomization	Blocking	Replication	Control	○○●○
76. Replication ensures that results are:	Hidden	Random	Biased	Reliable	○○○●
77. Correlation describes an association between:	Printers	Programs	Variables	Files	○○●○
78. Causation means that a change in one variable:	Doubles the other	Has no effect on the other	Brings about a change in the other	Deletes the other	○○●○
79. Which statement is correct?	Correlation always implies causation	Causation never implies correlation	Correlation does not imply causation	Correlation is the same as causation	○○●○
80. In the smartphone lag example, the cause of lag was:	Lack of RAM	The battery colour	The SIM card	Sending a text	●○○○

StudyBees OFFICIAL PAK

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio

Questions	A	B	C	D	A B C D
81. A population is a:	Complete set of items of interest	Single item	Subset of items	Random number	●○○○
82. A random sample is a subset of the population in which each member has:	No chance of being selected	A higher chance if older	A fixed chance of 50%	An equal chance of being selected	○○○●
83. A parameter describes an aspect of the:	Computer	Entire population	Sample only	Network	○●○○
84. A statistic describes an aspect of a:	Hypothesis	Population	Sample	Parameter	○○●○
85. A parameter is usually:	Known and variable	Fixed and unknown	Always zero	Always one	○●○○
86. A statistic is a:	Known number that depends on the sample	Network	Fixed unknown number	Program	●○○○
87. The mean weight of 100 Persian cats is an example of a:	Population	Hypothesis	Statistic	Parameter	○○●○
88. The average income of all government employees in Pakistan is an example of a:	Variable	Sample	Statistic	Parameter	○○○●
89. The mean, median and mode are different types of:	Graphs	Tables	Files	Averages	○○○●
90. Data collection is the process of collecting, measuring and _____ information.	Analyzing	Hiding	Printing	Deleting	●○○○
91. Primary data is collected:	Firsthand from the main source	From libraries only	From the Internet only	By someone else earlier	●○○○
92. Primary data refers to data that has:	Been used many times	Been deleted	Never been used in the past	Been encrypted	○○●○
93. Which is a primary data collection method?	Interviews	Government archives	Internet search	Libraries	●○○○
94. Which is another primary data collection method?	Internet search	Government archives	Focus groups	Libraries	○○●○
95. A direct method of data collection where the interviewer asks questions is:	Observation	Interview	Library	Archive	○●○○
96. In observation, researchers observe a situation and:	Delete data	Record the findings	Print data	Encrypt data	○●○○
97. Surveys and questionnaires provide a broad perspective from:	Large groups of people	No one	Machines only	A single person	●○○○

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio

Questions	A	B	C	D	A B C D
98. A focus group is conducted with people who all have something:	Unknown	In common	Different	Hidden	○●○○
99. A drawback of focus groups is:	No participants	Low cost	Too much privacy	Lack of privacy and domination by one or two participants	○○○●
100. Oral histories collect data linked to a:	Hardware device	Large database	Single phenomenon	Printer	○○●○
101. Secondary data refers to data that has been:	Already collected by someone else	Deleted	Collected firsthand	Never collected	●○○○
102. Compared with primary data, secondary data is:	Always original	Cheaper and easier to collect	Harder to collect	More expensive	○●○○
103. Which is a secondary data collection method?	Focus groups	Government archives	Interviews	Observations	○●○○
104. Which is another secondary data collection method?	Libraries	Interviews	Surveys	Oral histories	●○○○
105. The advantage of government archives is that the data is:	Always free	Always easy to access	Authentic and verifiable	Always up to date	○○●○
106. While using the Internet for data, you should source only from:	Unknown sites	Authentic sites	Any site	Social media only	○●○○
107. A/B testing is also known as:	Block testing	Mirror testing	Binary testing	Split testing	○○○●
108. A/B testing compares how many versions of something?	Four	Two	Three	One	○●○○
109. Which companies use A/B testing, according to the book?	Facebook, Airbnb and YouTube	Only Apple	Only Microsoft	Only Google Maps	●○○○
110. The first step in analyzing a dataset is:	Data locking	Data printing	Data deletion	Data exploration	○○○●
111. Summary statistics are calculated to understand tendency and:	Size	Speed	Dispersion	Colour	○○●○
112. Summary statistics include mean, median, mode, count and:	Font	Password	Colour	Frequency	○○○●
113. A data product is a tool that uses data to help businesses improve their:	Hardware	Fonts	Decisions and processes	Cables	○○●○
114. Data visualization is the graphical representation of:	Information and data	Passwords	Hardware	Software licences	●○○○

StudyBees OFFICIAL PAK

Questions	A	B	C	D	A B C D
115. Data visualization helps to see trends, outliers and:	Viruses	Cables	Passwords	Patterns	○○○●
116. The Python library most important for data analysis is:	turtle	pandas	random	datetime only	○●○○
117. The command to install pandas is:	install pandas	pip install pandas	get pandas	python pandas	○●○○
118. The function used to read a CSV file in pandas is:	pd.load_txt()	pd.open()	pd.read_csv()	pd.get()	○○●○
119. The statement data.head(10) displays:	The top 10 rows	10 columns	The data types	The last 10 rows	●○○○
120. matplotlib is a data _____ library.	Compression	Networking	Visualization	Encryption	○○●○
121. Scatter plots are used to observe:	Only names	Only dates	Relationships between variables	Only totals	○○●○
122. The method used to draw a scatter plot is:	bar()	plot()	hist()	scatter()	○○○●
123. The function used to draw a line chart is:	scatter()	pie()	plot()	bar()	○○●○
124. The method used to create a bar chart is:	hist()	pie()	bar()	plot()	○○●○
125. A histogram is a type of bar plot in which the X-axis represents:	Colours	Bin ranges	Names	Frequency	○●○○
126. In a histogram, the Y-axis gives information about:	Bin ranges	Frequency	Names	Colours	○●○○
127. The function used to create a histogram is:	hist()	plot()	bar()	pie()	●○○○
128. A boxplot is also known as a:	Line plot	Scatter plot	Pie plot	Box-and-whisker plot	○○○●
129. A boxplot helps to identify central tendency, spread and:	Outliers	Colours	Axes	Titles	●○○○
130. The function used to compute a boxplot is:	hist()	pie()	scatter()	boxplot()	○○○●
131. A pie chart is a circular chart divided into:	Dots	Bars	Lines	Sectors	○○○●

StudyBees OFFICIAL PAK

Questions	A	B	C	D	A B C D
132. The function used to draw a pie chart is:	bar()	plot()	pie()	hist()	○○●○
133. In a pie chart, slice areas are calculated as each value divided by the:	Sum of values	Mean	Smallest value	Largest value	●○○○
134. Which function gives a plot a title in matplotlib?	plt.name()	plt.label()	plt.title()	plt.head()	○○●○
135. Which function displays the plot in matplotlib?	plt.print()	plt.show()	plt.open()	plt.draw_all()	○●○○
136. Reinforcement learning involves training the algorithm to iterate over:	Two attempts only	No attempts	One attempt	Many attempts	○○○●
137. Which of the following is a classification model?	Association rule	Linear regression	K-means	Decision tree	○○○●
138. Which of the following is a common classification model?	Pie chart	Naive Bayes	Slope	Boxplot	○●○○
139. Linear regression allows us to predict a response for a given:	Colour	Printer	Predictor value	Cable	○○●○
140. A regression model is designed to analyze the relationship between:	Two files	Two printers	Independent and dependent variables	Two screens	○○●○
141. The most common step after data exploration and before visualization is:	Data deletion	Data cleaning	Data hiding	Data locking	○●○○
142. Experimental design is a tool to organize, conduct and clarify the results of an:	Printer	Experiment	Program	Network	○●○○
143. The 'Find the variables' step decides which variables will be:	Fixed and changed	Printed	Copied	Deleted	●○○○
144. Calculating the needed sample size helps to generate appropriate:	Cables	Colours	Fonts	Statistical results	○○○●
145. The 'Carry out the experiment' step requires following the defined:	Font	Colour	Procedure	Cable	○○●○
146. Interpreting and drawing conclusions happens after:	Sample selection	Data analysis	Data collection	Problem definition	○●○○
147. Statistical modeling methods use data gathered from the cloud, spreadsheets, databases and:	Only books	Only radios	Other sources	Only paper	○○●○
148. Data gathering is the foundation of:	Printing	Statistical modeling	Typing	Soldering	○●○○

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio

Questions	A	B	C	D	A B C D
149. The principle of blocking makes sure results aren't affected by:	Computers	Colours	Printers	Group differences	○○○●
150. Education data analysis helps in planning:	Educational resources for students and teachers	Cables	Passwords	Printer drivers	●○○○

Follow Studybeesofficial on Instagram, YouTube Channel & WhatsApp Channel — Link in Bio