<table>
<tr><th colspan="2">UNIT 2 STATISTICS</th></tr>
<tr><td>1.</td><td>The marks obtained by five students in a mathematics test are:
45, 50, 55, 60, and 65.
Find the mean (average) marks.</td></tr>
<tr><td>2.</td><td>Problem:
Find the median of the following numbers:
12, 18, 25, 30, and 40.</td></tr>
<tr><td>3.</td><td>Find the mode of the data set:
3, 7, 7, 9, 12, 7, 15, 18, 9.</td></tr>
<tr><td>4.</td><td>The ages of five students are: 12, 14, 10, 16, 13.
Find the range of ages.</td></tr>
<tr><td>5.</td><td>The marks of five students in a test are: 8, 10, 12, 14, 16.
Find the variance and standard deviation.</td></tr>
<tr><td>6.</td><td>A company wants to test if customer satisfaction depends on the type of service received. A random sample of 100 customers is surveyed, and the results are:
<table>
<tr><th></th><th>Satisfied</th><th>Not Satisfied</th><th>Total</th></tr>
<tr><td>Online Service</td><td>40</td><td>20</td><td>60</td></tr>
<tr><td>In-Person Service</td><td>25</td><td>15</td><td>40</td></tr>
<tr><td>Total</td><td>65</td><td>35</td><td>100</td></tr>
</table>
Does customer satisfaction depend on the type of service at a 5% significance level?
(critical chi-square value = 3.84)</td></tr>
<tr><td>7.</td><td>A researcher wants to determine the correlation between study hours and exam scores of five students. The data is:
<table>
<tr><th>Student</th><th>Study Hours (X)</th><th>Exam Score (Y)</th></tr>
<tr><td>1</td><td>2</td><td>50</td></tr>
<tr><td>2</td><td>3</td><td>60</td></tr>
<tr><td>3</td><td>5</td><td>80</td></tr>
<tr><td>4</td><td>7</td><td>90</td></tr>
<tr><td>5</td><td>8</td><td>95</td></tr>
</table>
Calculate the Pearson correlation coefficient (r).</td></tr>
<tr><td>8.</td><td>One sample T-test
A university claims that the average IQ of its students is 110. A researcher collects a sample of 10 students and records their IQ scores:
105,108,112,115,107,109,111,113,106,110105, 108, 112, 115, 107, 109, 111, 113, 106, 110105,108,112,115,107,109,111,113,106,110
Test at a 5% significance level whether the university’s claim is valid.</td></tr>
</table>

<table>
<tr><td>9.</td><td>Independent T-test
A company wants to test if there is a difference in productivity between employees working remotely and those working in-office. They collect data on weekly tasks completed for 8 employees in each group:
<table><tr><th>Remote Workers</th><th>In-Office Workers</th></tr><tr><td>15</td><td>12</td></tr><tr><td>18</td><td>14</td></tr><tr><td>16</td><td>13</td></tr><tr><td>20</td><td>15</td></tr><tr><td>22</td><td>17</td></tr><tr><td>19</td><td>14</td></tr><tr><td>21</td><td>16</td></tr><tr><td>17</td><td>13</td></tr></table>At a 5% significance level, determine if there is a significant difference in productivity.</td></tr>
<tr><td>10.</td><td>Paired t-Test
A fitness coach tests whether a new workout plan improves the number of push-ups a person can do in 1 minute. He records the performance of 6 clients before and after training:
<table><tr><th>Client</th><th>Before</th><th>After</th></tr><tr><td>1</td><td>20</td><td>25</td></tr><tr><td>2</td><td>18</td><td>22</td></tr><tr><td>3</td><td>15</td><td>19</td></tr><tr><td>4</td><td>22</td><td>27</td></tr><tr><td>5</td><td>19</td><td>23</td></tr><tr><td>6</td><td>16</td><td>21</td></tr></table>Does the workout plan significantly improve push-ups at a 5% significance level?</td></tr>
<tr><td>11.</td><td>Define Type I and Type II errors in hypothesis testing.</td></tr>
<tr><td>12.</td><td>How does sample size affect Type I and Type II errors?</td></tr>
<tr><td>13.</td><td>What is the main difference between a one-tailed and a two-tailed test?</td></tr>
<tr><td>14.</td><td>What is the need of statistics in data science</td></tr>
<tr><td>15.</td><td>Bayes Theorem
A certain disease affects 1% of a population. A test for the disease is 90% accurate for sick people (true positive) and 95% accurate for healthy people (true negative). If a person tests positive, what is the probability that they actually have the disease?</td></tr>
<tr><td>16.</td><td>A spam filter is 80% effective in identifying spam emails and 99% effective in</td></tr>
</table>

	identifying non-spam emails. Suppose **30%** of all received emails are spam. If an email is classified as spam, what is the probability that it actually is spam?
17.	What is hypothesis? Explain types of hypotheses.
18.	Explain five point number summary of box plot
19.	Define: Mean deviation, interquartile range, lower fence (bound), upper fence (bound)
20.	What is skewness? Explain the types of skewness of data.