Instructions: This is a closed-book and closed-notes exam. ONE two-sized notes sheet of letter size, calculator (cell-phones may NOT be used as a calculator), Minitab and R except their tutorial/help files are allowed; no internet access is allowed. Except multiple choice questions, give concise but detailed answers for full credit.

Instructions for graphing: Every axis on your graphs must be labeled by its name and in consistent scale; all slices or bars must be labeled by their frequencies or percentages.

1. [4 marks] A national polling organization wishes to estimate the percentage of all teenagers who believe social security will 'be there' for them. The organization surveys a random sample of 1500 teenagers and 37% of this sample says that they believe social security will 'be there' for them. In this survey, what is the population of interest?
   A. The 1500 teenagers who were surveyed.
   B. Teenagers who believe social security will 'be there' for them.
   C. All teenagers.
   D. The people in the sample who believe social security will 'be there' for them.
   Answer: C

Questions 2 to 5: Olivia wants to learn a foreign language. To get an idea of how satisfied other students were after taking a foreign language course, she decides to take a random sample of 20 students.

2. [4 marks] If Olivia randomly selects 5 students from French, 5 from German, 5 from Spanish, and 5 from Chinese, the sampling method is a
   A. simple random sample.
   B. stratified random sample.
   C. cluster sample.
   D. systematic sample.
   Answer: B

3. [4 marks] If Olivia randomly selects one class among all the foreign language classes taught that year, and then interviews all students in that class, the sampling method is a
   A. simple random sample.
   B. stratified random sample.
   C. cluster sample.
   D. systematic sample.
   Answer: C

4. [4 marks] If Olivia randomly selects 20 students among all students taking a foreign language class that year, the sampling method is a
   A. simple random sample.
   B. stratified random sample.
   C. cluster sample.
   D. systematic sample.
   Answer: A

5. [4 marks] If Olivia selects every 100th student from a list of the 2000 students who took a foreign language that year, the sampling method is a
   A. simple random sample.
   B. stratified random sample.
   C. cluster sample.
   D. systematic sample.
   Answer: D
Questions 6 to 8: A government agency randomly sampled 1000 voters who have party membership. The table below summarizes observed data on the gender and party membership of 1000 individuals:

<table>
<thead>
<tr>
<th>Party Membership</th>
<th>Gender</th>
<th>Democrat</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

6. [4 marks] Which one of the following statements about the relationship between gender and party is suggested by the data in the table?
   A. There is a relationship between gender and party membership.
   B. There is no relationship between gender and party membership.
   C. There is a relationship between gender and being a Democrat but not between gender and being a Republican.
   D. There is a relationship between gender and party membership for males but not for females.
   Answer: B

7. [2 marks] Is this study an observational study or control study (randomized experiment)? No explanation is needed.
   Answer: observational study

8. [8 marks] Draw ONE graph to show the association between gender and party. Explain why pie graph is not a choice in this case.
   Answer: A “relative frequency” cluster bar chart; x axes is part membership with gender clusters. Make a relative frequency table first:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Party Membership</th>
<th>Democrat</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>50%</td>
<td>50%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50%</td>
<td>50%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Note that you can switch the roles of gender and party membership.

Pie graph is only good to display percentages of which the sum is 100%.
Questions 9 and 10: In a survey of 1000 adults, respondents were asked about the expense of a college education and the relative necessity of financial assistance. The correspondents were classified as to whether they currently had a child in college or not (college status), and whether they thought the loan obligation for most college students was too high, about right, or too little (loan obligation opinion). The table below summarizes some of the survey results. Use these results to answer the following questions.

<table>
<thead>
<tr>
<th>Loan Obligation Opinion</th>
<th>College Status</th>
<th>Too High</th>
<th>About Right</th>
<th>Too Little</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Child in College</td>
<td>350</td>
<td>80</td>
<td>10</td>
<td>440</td>
</tr>
<tr>
<td></td>
<td>No Child in College</td>
<td>250</td>
<td>200</td>
<td>110</td>
<td>560</td>
</tr>
</tbody>
</table>

9. [4 marks] Which type(s) of variable is Loan Obligation Opinion? Could be more than one answer.
   a. Categorical
   b. Quantitative
   c. Discrete
   d. Ordinal

Answer: A, D

10. [10 marks] Draw ONE table to show the association between college status and loan obligation opinion. Describe what you see from the table about the association.

Answer:

<table>
<thead>
<tr>
<th>Loan Obligation Opinion</th>
<th>College Status</th>
<th>Too High</th>
<th>About Right</th>
<th>Too Little</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Child in College</td>
<td>79.5%</td>
<td>18.2%</td>
<td>2.3%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>No Child in College</td>
<td>44.6%</td>
<td>35.7%</td>
<td>19.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Based on these percentages, describe the relationship between loan obligation opinion and college status. There does appear to be some association between loan obligation opinion and college status. The parents with a child in college are much more likely to say the loan obligations are too high as compared to parents with no child in college (79.5% to 44.6%). On the other extreme, the parents with no child in college are much more likely to say the loan obligations are too little as compared to parents with a child in college (19.6% to 2.3%).
Questions 11 and 15: The following histogram is for the weights (lbs) of 119 female college students.

11. [4 marks] What is the best approximate shape of the distribution?
   a. Nearly symmetric.
   b. Slightly skewed to the left.
   c. Slightly skewed to the right.
   d. Bimodal.
   Answer: C

12. [8 marks] List down the 10 class intervals and their corresponding frequencies (approximately; make sure the total frequency is 119).
   Answer:

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency (f)</th>
<th>Cumulative freq.</th>
<th>Midpoint (m)</th>
<th>f x m</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-100</td>
<td>2</td>
<td>2</td>
<td>95</td>
<td>190</td>
</tr>
<tr>
<td>100-110</td>
<td>7</td>
<td>9</td>
<td>105</td>
<td>735</td>
</tr>
<tr>
<td>110-120</td>
<td>19</td>
<td>28</td>
<td>115</td>
<td>2185</td>
</tr>
<tr>
<td>120-130</td>
<td>22</td>
<td>50</td>
<td>125</td>
<td>2750</td>
</tr>
<tr>
<td>130-140</td>
<td>30</td>
<td>80</td>
<td>135</td>
<td>4050</td>
</tr>
<tr>
<td>140-150</td>
<td>16</td>
<td>96</td>
<td>145</td>
<td>2320</td>
</tr>
<tr>
<td>150-160</td>
<td>9</td>
<td>105</td>
<td>155</td>
<td>1395</td>
</tr>
<tr>
<td>160-170</td>
<td>5</td>
<td>110</td>
<td>165</td>
<td>825</td>
</tr>
<tr>
<td>170-180</td>
<td>5</td>
<td>115</td>
<td>175</td>
<td>875</td>
</tr>
<tr>
<td>180-190</td>
<td>4</td>
<td>119</td>
<td>185</td>
<td>740</td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>N.A.</td>
<td>N.A.</td>
<td>16065</td>
</tr>
</tbody>
</table>
13. [4 marks] The best choice for the median weight for the 119 women based on the histogram is approximately
   a. 130 pounds.
   b. 133 pounds.
   c. 135 pounds.
   d. 160 pounds.
Answer: B

14. [4 marks] The best choice for the mean weight for the 119 women based on the histogram is approximately
   a. 130 pounds.
   b. 133 pounds.
   c. 135 pounds.
   d. 160 pounds.
Answer: C

15. [4 marks per part] Assume that the standard deviation of the 119 observations is 15 and answer the following questions:
   a. According to the Empirical Rule, what interval might include about 81 of the 119 observations?
   b. Use pencil/pen to put darker shading in the area of the histogram bars that corresponds to observations of the interval of part A.
   c. Actually, how many observations does this interval contain approximately?
   d. Briefly, in this case, how would you explain the (poor or good) agreement of actual results with the Empirical Rule?
Answer:
   a. $81/119 = 68\%$, so the interval should be Mean +/- 1 St. Dev. = 135 +/- 15 = (120, 150)
   b. Shading the three bars corresponding to 120-130, 130-140, 140-150
   c. From the table in question 12, the total frequency of (120, 150) is 22+30+16= 68
   d. Out of 119 subjects, 68 is quite off from 81. The difference 13 (81-68) is 11% (13/119) of the whole sample. This is expected as Empirical Rule is only for symmetric distributions, while this histogram shows right-skewed.
Questions 16 to 19: An experiment involves tossing a pair of dice, 1 green and 1 red, and recording the numbers that come up. Let \( x \) be the outcome on the green die and \( y \) the outcome on the red die. Event A is that the sum is greater than 8 and event B is that a 2 occurs on either die.

16. [4 marks] List all outcomes corresponding to the event A and find the probability that A occurs.
17. [4 marks] List all outcomes corresponding to the event B and find the probability that B occurs.
18. [4 marks] List all outcomes corresponding to the event \( A \cap B \) and find the probability that A and B both occur.
19. [4 marks] Are event A and B independent? Why yes or why not?

Answer:

16. \( A = \{(x,y)=(3,6), (4,5), (5,4), (6,3); (4,6), (5,5), (6,4); (5,6), (6,5), (6,6)\} \Rightarrow P(A)=\frac{10}{36}=27.78\% \)

17. \( B = \{(x,y)=(1,2), (2,2), (3,2), (4,2), (5,2), (6,2); (2,1), (2,3), (2,4), (2,5), (2,6)\} \Rightarrow P(B)=\frac{11}{36}=30.56\% \)

18. Hence, \( A \cap B = \) empty set and \( P(A \cap B) \) is zero.

19. Event A and B are NOT independent because \( P(A \cap B) \) and \( P(A) \cdot P(B) \) are not equal.