Instructions:
This exam contains 34 Multiple Choice questions. Each question is worth 3 points, for a total of 102 points (there are TWO bonus points). Please select the best answer among the alternatives given.

You may write whatever you want on this test, but only the answers bubbled in the scantron sheet will be graded. This test MUST BE SUBMITTED to the instructors together with the scantron sheet for you to receive a grade on the exam.

Honor pledge: "On my honor, I have neither given nor received unauthorized aid on this examination."

Signature: ______________________

Questions 1 – 4
Is there a difference in dexterity between three-year-old boys and girls? A sample of 13 children was obtained, 5 boys and 8 girls. They were asked to place a set of blocks into a specified pattern. The time (in seconds) required by each child to arrange the blocks was recorded below, with the ranks for each child given in parenthesis.

<table>
<thead>
<tr>
<th>Boys</th>
<th>23 (3)</th>
<th>18 (1)</th>
<th>29 (6)</th>
<th>42 (10)</th>
<th>21 (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>37 (8)</td>
<td>56 (12)</td>
<td>39 (9)</td>
<td>34 (7)</td>
<td>26 (5)</td>
</tr>
<tr>
<td></td>
<td>104 (13)</td>
<td>48 (11)</td>
<td>25 (4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. When performing the Wilcoxon Rank Sum test, we summarize the data by computing:
   a) the average of the observations in each group
   b) the average of the ranks of the observations in each group
   c) the observed counts for each group
   d) the expected counts for each group

2. Suppose the p-value for the two-sided test was 0.06. Then we have evidence that the dexterity of 3-year-old boys is:
   a) less than that of girls, for children studied in the sample.
   b) different than that of girls, for children studied in the sample.
   c) less than that of girls, for children in the population.
   d) different than that of girls, for children in the population.

3. In Minitab, that p-value would be found by using the:
   a) Wilcoxon Rank-Sum Test
   b) Wilcoxon Sign Rank Test
   c) Kruskal-Wallis Test
   d) Mann-Whitney Test

4. It would not be appropriate to use a Normal based procedure on this data because:
   a) the assumption of Normality is violated.
   b) the sample sizes are different.
   c) the variances are unequal.
   d) the data is not quantitative.
Questions 5 – 9 For each of the following stories, determine which would be the simplest type of statistical analysis that would be appropriate to use. Use each type of analysis only once.

a) Paired t-test
b) Two sample t-test
c) ANOVA
d) Kruskal-Wallis
e) Wilcoxon Rank-Sum Test

5. Compare the average number of hours per week spent on Facebook for Freshmen, Sophomore, Juniors and Seniors at UF, based on a random sample of 100 students.

6. Compare the distribution of the number of hours per week spent on Facebook for Freshmen, Sophomore, Juniors and Seniors at UF, based on a random sample of 10 students. The standard deviation of the groups were quite different.

7. Compare the average number of hours per week spent on Facebook during the first week in April and the first week in May (finals week) for students at UF, based on a random sample of 100 students.

8. Compare the distribution of the number of hours per week spent on Facebook for male and female students at UF, based on a random sample of 10 students. There was an outlier in one of the groups.

9. Compare the average number of hours per week spent on Facebook for male and female students at UF, based on a random sample of 100 students.

Questions 10 – 14 For each of the following stories, determine which would be the simplest type of statistical analysis that would be appropriate to use. Use each type of analysis only once.

a) Confidence Interval for One Proportion
b) Contingency Table
c) Simple Linear Regression
d) Multiple Regression
e) Logistic Regression

10. Predict the average number of hours per week UF students spend on Facebook, based on the student’s age and gender.

11. Estimate the fraction of UF students who have Facebook accounts.

12. Determine if the fraction of UF students who have Facebook accounts is different for Males and Females.

13. Determine how the probability that a UF student has a Facebook account changes with the student’s age.

14. Predict the average number of hours per week UF students spend on Facebook, based on the student’s age.
Questions 15–21 Can we predict marijuana use (yes/no) on cigarette use (yes/no) and alcohol use (yes/no)? Data from a survey conducted on high school seniors on a rural area near Dayton, Ohio was used to perform the logistic regression analysis whose results appear below.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-5.30904</td>
<td>0.475190</td>
<td>-11.17</td>
<td>0.000</td>
</tr>
<tr>
<td>alcohol</td>
<td>2.98601</td>
<td>0.166471</td>
<td>6.43</td>
<td>0.000</td>
</tr>
<tr>
<td>cigarettes</td>
<td>2.84789</td>
<td>0.163839</td>
<td>17.38</td>
<td>0.000</td>
</tr>
</tbody>
</table>

15. Which of the following variables in this model are Binomial?
   a) response only  ☑️ b) response and two predictors
c) response and one predictor  d) response and three predictors  e) none of them

16. The chance of using marijuana in this population:
   a) increase with the use of alcohol. ☑️
   b) increase with the use of cigarettes. ☑️
   c) decrease with the constant.
   d) all of the above.
   e) both a and b are correct.

17. Interpret the coefficient of alcohol. The probability of marijuana use for alcohol drinkers:
   a) increases by 2.98601 with every drink
   b) is 2.98601 times the probability for non-alcohol drinkers
   c) increases by 19.8106 with every drink
   d) are 19.8106 times the probability for non-alcohol drinkers
   e) none of the above

18. Is the effect of the predictors significant?
   a) only for one predictor
   b) yes, for all three predictors
   c) yes, for both predictors
   d) only for two of the predictors
   e) none of the above

19. If the relative risk of using marijuana for a particular subgroup of students in this population are 1.70, that means we expect to see about:
   a) 17 students in this subgroup who use marijuana for every 10 who don’t.
   b) 1.7 students in this subgroup who use marijuana for every 10 who don’t.
   c) 0.17 students in this subgroup who use marijuana for every one who doesn’t.
   d) 170 students in this subgroup who use marijuana for every one who doesn’t.

20. Find the probability of marijuana use for a student in this population who drinks but does NOT smoke.
   a) -2.32303
   b) 0.09798
   c) 8.29505
   d) 0.08923

21. Find the probability of marijuana use for a student in this population who smokes but does NOT drink.
   a) -2.46115
   b) 0.085337
   c) 0.459459
   d) 0.07834
Questions 22 – 30 The University of Florida Alcohol and Drug Survey conducts yearly surveys on the attitudes, knowledge and behaviors of students regarding drug and alcohol use. The table below presents some of the results of the 2002 survey. It compares students who belong to a fraternity or sorority (Greek students) to those who don’t (non-Greek students) with respect to their frequency of binge drinking. The question asked was: “Think back over the last two weeks, how many times have you had 5 or more drinks at a setting?”

<table>
<thead>
<tr>
<th>Binge-Drinking</th>
<th>Greek</th>
<th>Non-Greek</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>65</td>
<td>686</td>
<td>751</td>
</tr>
<tr>
<td>Once/Twice</td>
<td>64</td>
<td>290</td>
<td>354</td>
</tr>
<tr>
<td>3 to 5 times</td>
<td>45</td>
<td>136</td>
<td>181</td>
</tr>
<tr>
<td>6 or more times</td>
<td>24</td>
<td>29</td>
<td>53</td>
</tr>
<tr>
<td>Total</td>
<td>198</td>
<td>1141</td>
<td>1339</td>
</tr>
</tbody>
</table>

Match the following probabilities from the table with the correct description. The probability that a student selected at random from the population:

22. is a Greek who never binge drinks = \( \frac{65}{751} \)
   a) 198/1339

23. of students who never binge drink is Greek = \( \frac{65}{751} \)
   b) 65/751

24. of Greeks never binge drinks = \( \frac{65}{198} \)
   c) 65/198

25. The hypotheses being tested here have to do with the association between:
   a) Greeks and non-Greeks.
   b) those who binge drink and those who don’t.
   c) levels of binge drinking
   d) Greek affiliation and frequency of binge drinking.

26. What is the distribution of the test statistic under the null hypothesis?
   a) \( \chi^2_1 \)
   b) \( \chi^2_8 \)
   c) \( \chi^2_3 \)
   d) \( F_{2,4} \)
   e) \( F_{1,3} \)
   
27. How many Greeks are expected to answer "None" under the null hypothesis?
   a) 36.45
   b) 111.1
   c) 33.30
   d) 225.2
   e) 49.5

28. What is the contribution to the test statistic of the non-Greek/6-or-more-times category?
   a) 5.78
   b) 45.16
   c) 9.00
   d) 0.357
   e) 44.52

29. The p-value for this test was reported by Minitab as 0.000. We can then:
   a) reject the null hypothesis and conclude there is an association
   b) reject the null hypothesis and conclude there is NO association
   c) fail to reject the null hypothesis and conclude there is an association
   d) fail to reject the null hypothesis and conclude there is NO association

30. Which of the following best describes the relationship in this table?
    a) More Greek students binge drink than non-Greek students.
    b) Less Greek students binge drink than non-Greek students.
    c) Greek students are more likely to binge drink than non-Greek students.
    d) Greek students are less likely to binge drink than non-Greek students.

31. The mode of the Chi-Squared distribution is approximately:
    a) zero
    b) the same as its degrees of freedom
    c) one
    d) the same as its mean
Questions 32 – 34 How fast does fabric decay when buried underground? 20 strips of polyester fabric were buried, and taken out at different times to determine their breaking strength (the lower the breaking strength, the weaker and more decayed the fabric). Minitab output for this data appears below.

32. Regarding the breaking strengths of the fabric at these time periods, at $\alpha=0.10$,
   a) Kruskal-Wallis finds significant differences, but ANOVA does not.
   b) ANOVA finds significant differences, but Kruskal-Wallis does not.
   c) Both methods find significant differences.
   d) Neither method finds significant differences.

33. Which is more appropriate to use for this data set?
   a) Kruskal-Wallis, since the assumption of equal variances is violated.
   b) Kruskal-Wallis, since the assumption of randomness is violated.
   c) ANOVA, since the assumption of continuity is violated.
   d) ANOVA, since none of the assumptions are violated.

34. This data shows that ANOVA is:
   a) more powerful than Kruskal-Wallis, and with less restrictive assumptions.
   b) more powerful than Kruskal-Wallis, but not always preferred.
   c) less powerful than Kruskal-Wallis, but easier to use.
   d) less powerful than Kruskal-Wallis but more widely accepted.

Descriptive Statistics: brk_str
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SEMean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>16wks</td>
<td>5</td>
<td>116.40</td>
<td>7.19</td>
<td>16.09</td>
</tr>
<tr>
<td>2wks</td>
<td>5</td>
<td>123.80</td>
<td>2.06</td>
<td>4.60</td>
</tr>
<tr>
<td>4wks</td>
<td>5</td>
<td>123.60</td>
<td>2.93</td>
<td>6.54</td>
</tr>
<tr>
<td>8wks</td>
<td>5</td>
<td>134.40</td>
<td>4.26</td>
<td>9.53</td>
</tr>
</tbody>
</table>

Kruskal-Wallis Test on brk_str
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Median</th>
<th>Ave Rank</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>16wks</td>
<td>5</td>
<td>120.0</td>
<td>6.7</td>
<td>-1.66</td>
</tr>
<tr>
<td>2wks</td>
<td>5</td>
<td>126.0</td>
<td>9.7</td>
<td>-0.35</td>
</tr>
<tr>
<td>4wks</td>
<td>5</td>
<td>126.0</td>
<td>10.2</td>
<td>-0.13</td>
</tr>
<tr>
<td>8wks</td>
<td>5</td>
<td>136.0</td>
<td>15.4</td>
<td>2.14</td>
</tr>
<tr>
<td>Overall</td>
<td>20</td>
<td></td>
<td>10.5</td>
<td></td>
</tr>
</tbody>
</table>

H = 5.60  DF = 3  P = 0.133
H = 5.63  DF = 3  P = 0.131  (adjusted for ties)

One-way ANOVA: 2wks, 4wks, 8wks, 16wks
<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>3</td>
<td>825</td>
<td>275</td>
<td>2.66</td>
<td>0.084</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>1654</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>2479</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rej Ho at $\alpha=.10$