INSTRUCTIONS:
- FILL OUT your personal information above.
- BUBBLE IN SCANTRON: Name, UFID, section number, Test Form Code and Special Code.
- FORMAT: This exam contains 29 Multiple Choice questions. Each question is worth 3.5 points, for a total of 101.5 points (so there are 1.5 bonus points on the test).
- ANSWERS: 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.
- YOU MUST SUBMIT THIS TEST to the instructors together with the scantron sheet when you are finished.
- DOUBLE CHECK your personal information, test code and special code. If you miss any of those, 1.5 points will be deducted from your score.
- SCORES on the exam will be posted in Sakai within a week – please see course web page for details.
- SIGN HONOR PLEDGE:
"On my honor, I have neither given nor received unauthorized aid on this examination."

Signature: ________________________________

Questions 1–5 A logistic regression model was used to predict the probability of an adult receiving a flu shot in the most recent year depending on the level of income for the person’s household (measured in thousands of dollars, so $x = 75$ represents an income of $75,000 per year). The estimated coefficients were $a = -2.198$ and $b = 0.0143$.

1. Which of the following variables is/are assumed to have a normal distribution in the logistic regression model?
   a) whether or not a person received a flu shot  
   b) level of income for the person’s household 
   c) both a) and b) 
   d) neither a) nor b)

2. The coefficients indicate that the probability of getting a flu shot:
   a) increases with income.  
   b) increases by 1% as income increases by $1000$. 
   c) decreases with income.  
   d) decreases by 2% as income increases by $1000$.
   e) is not related to income.

3. Estimate the probability that a person from a household with an income of $68,000$ will get a flu shot.
   a) 0.294  
   b) 0.227  
   c) 0.367  
   d) 0.726  
   e) 0.773

   \[ \exp(-2.198 + 0.0143 \times 68) = -1.2254 \]
   \[ \frac{e^{-2.198}}{1 + e^{-2.198}} = 0.2936 \]
   \[ \frac{e^{0.0143 \times 68}}{1 + e^{0.0143 \times 68}} = 0.2269 \]

4. The probability that a person gets a flu shot is 50% when their annual income is:
   a) $153,700  
   b) $314,000  
   c) $21,980  
   d) $143,000  
   e) $50,000

   \[ x = \frac{-a}{b} = \frac{-(-2.198)}{0.0143} = 153,711 \]

5. The effect of income on probability of getting a flu shot:
   a) is not significant, because the coefficient is so small.
   b) is not significant, because the constant is negative.
   c) is significant, because income is measured in thousands of dollars.
   d) is significant, because the assumptions of the test are satisfied.
   e) may be significant, but it’s impossible to tell from the information given. - need p-value
Questions 6 - 14 A recent survey asked people how happy they were, and how many times per year they attended religious services. The data is shown below.

<table>
<thead>
<tr>
<th>Religious Service Attendance in a Year</th>
<th>Not too happy</th>
<th>Pretty happy</th>
<th>Very happy</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;12</td>
<td>189</td>
<td>908</td>
<td>382</td>
<td>1479</td>
</tr>
<tr>
<td>12-45</td>
<td>53</td>
<td>311</td>
<td>180</td>
<td>544</td>
</tr>
<tr>
<td>&gt;45</td>
<td>46</td>
<td>335</td>
<td>(294)</td>
<td>675</td>
</tr>
<tr>
<td>TOTAL</td>
<td>288</td>
<td>1554</td>
<td>856</td>
<td>2698</td>
</tr>
</tbody>
</table>

Match the following probabilities from the table with the correct description. The probability that someone:

6. attends religious services >45 times a year and is very happy
   a) 675/2698
   b) 294/675
   c) 294/2698
   d) 294/856
   e) 856/2698

7. who attends religious services >45 times a year is very happy
   a) 675/2698
   b) 294/675
   c) 294/2698
   d) 294/856
   e) 856/2698

8. who is very happy attends religious services >45 times a year
   a) 675/2698
   b) 294/675
   c) 294/2698
   d) 294/856
   e) 856/2698

9. Find the expected count in the “Pretty happy, <12” cell under the null hypothesis of the chi-squared test.
   a) 493.00
   b) 518.00
   c) 299.77
   d) 851.88
   e) 777.00

10. Find the contribution to the test statistic from the “Pretty happy, <12” cell.
    a) 3.70
    b) 5.61
    c) 0.07
    d) 3.15
    e) 3.47

11. Under the null hypothesis of the chi-squared test, the test statistic has a chi-squared distribution with:
    a) df numerator = 3
    b) df = 6
    c) df denominator = 4
    d) df = 4
    e) df = 9

12. In order to determine if there is a strong association between happiness and frequency of religious service attendance, look at:
    a) the p-value
    b) the alternative hypothesis
    c) the degrees of freedom
    d) the marginal probabilities
    e) the residuals

13. To help us determine the type of association we can look at:
    a) the p-value
    b) the alternative hypothesis
    c) the degrees of freedom
    d) the marginal probabilities
    e) the residuals

14. Minitab reports a p-value of 0.000 for the chi-squared test. We can say that:
    a) there is evidence of independence between happiness and frequency of religious service attendance.
    b) there is evidence of an association between happiness and frequency of religious service attendance.
    c) there is no evidence of independence between happiness and frequency of religious service attendance.
    d) there is no evidence of an association between happiness and frequency of religious service attendance.
    e) the procedure used here was not appropriate for the data.
15. In a Kruskal-Wallis Test we have evidence for:
   a) $H_0$: the sum of ranks are similar to each other.
   b) $H_a$: the sum of ranks are similar to each other.
   c) $H_0$: the average ranks are similar to each other.
   d) $H_a$: the average ranks are similar to each other.
   e) None of the above.

16. The mean ranks for Muzak and Advertisement are:
   a) 4.4 and 6.4
   b) 3.8 and 5.6
   c) 4.5 and 6.5
   d) 4 and 5.8
   e) 2.8 and 2.7

17. The nonparametric test would compare:
   a) the average amount of time people wait on hold for these two recordings in the population.
   b) the median amount of time people wait on hold for these two recordings in the population.
   c) the distribution of the amount of time people wait on hold for these two recordings in the population.
   d) both a and c are correct
   e) both b and c are correct

18. The data was analyzed in Minitab using both a Nonparametric and a Normal-based procedure. The $p$-value for the Mann-Whitney test was 0.3457, while the $p$-value for the Two-Sample T test was 0.271. What are the conclusions we should draw from these analyses?
   a) People remain on hold exactly the same amount of time for any type of recording.
   b) People are clearly willing to listen longer to an Advertisement than to Muzak.
   c) Classical music would be a better choice if the airline wants people to remain on hold.
   d) A larger study would be needed to determine if an Advertisement is better than Muzak.
   e) A smaller sample would be needed to determine if an Advertisement is better than Muzak.

19. If the person that remained on hold for 11 minutes had instead remained for 61 minutes, would it make a difference in the analyses of this data?
   a) It would make both procedures invalid.
   b) We expect it would only affect the conclusions of the Normal-based procedure.
   c) We expect it would only affect the conclusions of the Nonparametric procedure.
   d) It would probably affect the conclusions of both of these procedures.
   e) It would probably affect the conclusions of neither of these procedures.

20. If the study compared telephone holding times for these two types of recording, but used the same exact people instead, then we would have:
   a) independent samples instead of matched pairs
   b) matched pairs instead of independent samples
   c) one more treatment, but the same number of response variables
   d) one more response variable, but the same number of treatments
   e) one more experimental unit, but the same number of treatments
Questions 21–24 For each of the following stories, determine which would be the simplest type of statistical analysis that would be appropriate to use. You should use each type of analysis no more than once.

a) Two-sample t-test
b) Matched pairs t-test
c) Wilcoxon rank-sum test
d) Multiple linear regression
e) Logistic regression

21. Doctors want to find out if the person survives a heart attack is related to the average number of hours of sleep that the person gets per night.

22. Doctors want to see if patients who receive a certain blood pressure medicine by injection have a higher concentration of the medicine in their bloodstream than patients who receive it by pill. However, the researchers expect that due to a rare genetic defect, a few people in one or both groups may have extremely high concentrations of the drug compared to other patients.

23. Doctors want to see how exercise affects blood pressure by measuring the same person's blood pressure before and after a six-month exercise regimen.

24. Doctors want to predict patients’ blood pressure based on their age, height, and weight.

Questions 25–29 For each of the following stories, determine which would be the simplest type of statistical analysis that would be appropriate to use. You should use each type of analysis no more than once.

a) One way ANOVA
b) Two Way ANOVA
c) Kruskal Wallis
d) Contingency Table
e) One Proportion

25. Doctors want to find out if the probability of surviving a heart attack is different for males and females.

26. Doctors want to compare the average blood pressure changes of people that have followed three different diets for six months.

27. Doctors want to compare the median blood pressure changes of people that have followed three different diets for six months.

28. Doctors want to predict the probability of surviving a second heart attack for the population as a whole.

29. Doctors want to compare the mean blood pressure changes of males and females that have followed three different diets for six months.