Questions 1 – 5 A logistic regression model was used to predict the probability of voting for the Republican candidate in a presidential election, depending on the voter’s total family income (measured in thousands of dollars, so \( x = 75 \) represents an income of $75,000 per year). The estimated coefficients were \( a = -1.0 \) and \( b = 0.02 \).

1. The signs of the coefficients imply that probability of voting Republican:
   a) decreases by 1% as family income increases by $1000  
   b) increases with family income  
   c) increases by 2% as family income increases by $1000  
   d) decreases with family income  
   e) does not depend at all on family income

2. The probability that a family votes Republican is 50% when their income is ______ a year.
   a) $100,000  
   b) $50,000  
   c) $10,000  
   d) $20,000  
   e) $2,000

3. Predict the probability of voting Republican for a family whose annual income is 200 thousand dollars.
   a) 0.20  
   b) 0.82  
   c) 0.45  
   d) 0.98  
   e) 0.95

4. The reason Logistic Regression was used in this case instead of Least-Squares Regression is because:
   a) we are not sure if the data was randomly selected  
   b) we are worried about outliers in the data  
   c) the response variable is not Normal  
   d) it allowed for several predictors instead of only one  
   e) all of the above

5. The effect of income on probability of voting Republican:
   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.
Questions 6 – 14  A recent survey asked a group of people for their opinion on global warming and their position on abortion to determine if there is any association between the two questions. The data is shown below.

<table>
<thead>
<tr>
<th>GLOBAL WARMING</th>
<th>ABORTION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pro-Life</td>
</tr>
<tr>
<td>Real</td>
<td>221</td>
</tr>
<tr>
<td>Unproven</td>
<td>90</td>
</tr>
<tr>
<td>Hoax</td>
<td>64</td>
</tr>
<tr>
<td>Undecided</td>
<td>90</td>
</tr>
<tr>
<td>TOTAL</td>
<td>465</td>
</tr>
</tbody>
</table>

6. Find the expected count in the “Hoax, Pro-Life” cell under the null hypothesis of the chi-squared test.
   a) 6.30
   b) 18.22
   c) 28.73
   d) 45.78
   e) 86.33

\[ \exp = \frac{\text{row tot} \times \text{col tot}}{\text{big tot}} = \frac{102 \times 465}{1036} = 45.78 \]

7. Find the contribution to the test statistic from the “Hoax, Pro-Life” cell.
   a) 0.28
   b) 0.40
   c) 5.19
   d) 7.25
   e) 18.22

\[ \left( \frac{64 - 45.78}{45.78} \right)^2 = 7.25 \]

8. Under the null hypothesis of the chi-squared test, the test statistic has a chi-squared distribution with
   a) \( df = 11 \)
   b) \( df = 3 \)
   c) \( df = 4 \)
   d) \( df = 6 \)
   e) \( df = 8 \)

9. Using the chi-squared table, we could find the p-value by looking up the area to the right of:
   a) the test statistic
   b) the sample size
   c) the observed counts
   d) the degrees of freedom
   e) the expected counts

10. Minitab reports a p-value of 0.000 for the chi-squared test. At all usual significance levels, we should
    a) reject the null hypothesis, and we should conclude that there is evidence of an association between beliefs about global warming and abortion.
    b) reject the null hypothesis, and we should conclude that there is evidence of independence between beliefs about global warming and abortion.
    c) fail to reject the null hypothesis, and we should conclude that there is evidence of an association between beliefs about global warming and abortion.
    d) fail to reject the null hypothesis, and we should conclude that there is evidence of independence between beliefs about global warming and abortion.
    e) conclude the procedure was not appropriate for this data.

11. Which of the following is true regarding the assumptions for this test?
    a) Since the sample size is so large we don't have any worries about the assumptions of randomness.
    b) The conclusions are valid as long as the data was randomly selected.
    c) We need to check that all expected counts are at least five.
    d) Since the data is randomly selected we don't need to worry about the sample sizes.
    e) Both b and c are correct.

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

- 12. who is pro-life believes global warming is real
- 13. who believes global warming is real is pro-life
- 14. is pro-life and believes global warming is real

a) \( \frac{465}{1036} \)

b) \( \frac{221}{465} \)

c) \( \frac{221}{520} \)

d) \( \frac{221}{1036} \)

e) \( \frac{520}{1036} \)
Questions 15–17  Suppose we want to use a nonparametric test to investigate whether the average starting salaries are higher for recent UF graduates who majored in psychology compared to those who majored in sociology. The starting salaries for five randomly selected graduates with each major are shown below (in thousands of dollars).

15. Find the mean rank for psychology.
   a) 2.7
   b) 3.8
   c) 4.2
   d) 4.7
   e) 4.9

   Psychology
   34 (2.5)
   39 (8)
   35 (4)
   31 (1)
   44 (9)

   Sociology
   34 (2.5)
   37 (5.5)
   3(5.5)
   93 (15)
   38 (7)

16. Minitab reported a p-value of 0.5993 for the nonparametric test. What conclusion should we draw from this analysis?
   a) The data shows no significant difference between starting salaries for psychology and sociology majors.
   b) The data shows a significant difference between starting salaries for psychology and sociology majors
   c) We should not conclude anything from this result, because the data contains an outlier.
   d) We should not conclude anything from this result, because the sample size is too small to use this test.
   e) We should not conclude anything from this result, because Minitab does not do this test.

17. Which of the following statements about our data and analysis is true?
   a) A two-sample t test would have been a better choice since the average salary is clearly higher for sociology majors than for psychology majors.
   b) The nonparametric procedure is the better choice since the distribution of salaries is clearly higher for sociology majors than for psychology majors.
   c) A two-sample t test would have been a better choice since the assumptions for the nonparametric procedure are clearly violated.
   d) The nonparametric procedure is the better choice since the assumptions for the parametric procedure are clearly violated.
   e) Neither is appropriate because the data is quantitative.

18. In a chi-squared test:
   a) the larger the test statistic, the more evidence to reject Ho.
   b) the larger the test statistic, the more evidence to reject H0.
   c) the closer to df the test statistic, the more evidence to reject Ho.
   d) the closer to df the test statistic, the more evidence to reject H0.
   e) none of the above.

19. In a Wilcoxon Signed Rank Test we have evidence for:
   a) Ho when half the differences are positive and half negative.
   b) Ha when half the differences are positive and half negative.
   c) Ho when the average ranks are similar to each other.
   d) Ha when the average ranks are similar to each other.
   e) None of the above.

20. If we want to do statistical inference and the data is categorical and not normal we:
   a) Should use nonparametric methods.
   b) Can use either normal based or nonparametric methods.
   c) Should use normal based procedures.
   d) Should use contingency tables.
   e) Can use either contingency tables or nonparametric methods.
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. Use each type of analysis no more than once.

a) Contingency table
b) One-way ANOVA
c) Two-way ANOVA
d) Kruskal-Wallis test
e) Simple linear regression

21. A hospital administrator wants to test the effectiveness of three different sterilization procedures for surgical instruments by measuring bacteria levels on instruments undergoing each procedure. Note that inadequate sterilization may result in certain instruments’ bacteria levels being thousands of times higher than those of others.

22. A hospital administrator wants to see if the number of hours needed to perform a certain surgery is related to which one of two doctors performs the surgery and which one of three specific procedures is used.

23. A hospital administrator wants to look for a relationship between which of three operating rooms is used to perform a surgery and whether or not a patient develops complications from the surgery.

24. A hospital administrator in a cold-weather city wants to examine the relationship between the amount of money spent per month to heat the hospital and the average outdoor temperature that month.

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. Use each type of analysis no more than once.

a) Logistic Regression
b) Two sample t test
c) Matched pair t test
d) Wilcoxon Rank Sum test
e) Multiple Regression

25. A hospital administrator wants to compare the mean amount of time it takes to perform two types of surgery at her hospital.

26. A hospital administrator wants to see if the number of hours needed to perform a certain surgery is related to whether or not a patient develops complications from the surgery.

27. A hospital administrator wants to see if the median number of hours needed to perform a certain surgery is related to which one of two doctors performs the surgery.

28. A hospital administrator in a cold-weather city wants to examine the amount of money spent per month to operate the hospital compared to the amount spent each month the previous year.

29. A hospital administrator wants to predict the length of stay at the hospital based on a patient’s age, gender, and type of surgical procedure performed.