EXAM 1   Test Form Code A

Instructions:

This exam contains 33 Multiple Choice questions. Each question is worth 3 points, for a total of 99 points. One point will be given for bringing your ID to the exam as well as filling in your name, UF ID#, and test code on your scantron correctly.

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. You must submit the scantron sheet for you to receive a grade on the exam. You must show the copy of the exam to the test room proctors before turning in your scantron.

\[ \hat{y} = a + bx \]
\[ b = r \frac{s_y}{s_x} \]
\[ a = \bar{y} - b\bar{x} \]
\[ \bar{x} = \frac{\sum x_i}{n} \]
\[ \binom{n}{x} = \frac{n!}{x!(n-x)!} \]
\[ P(x) = \binom{n}{x} p^x (1-p)^{n-x} \]
\[ \mu = np \quad \sigma = \sqrt{np(1-p)} \]
\[ \mu = \sum x^* P(x) \]
\[ \frac{1}{\sqrt{n}} \]

\[ P(A \text{ and } B) = P(A) \cdot P(B) \]
\[ P( A \mid B) = \frac{P( A \text{ and } B)}{P(B)} \quad \text{res} = \text{obs} y - \text{pred} y \]

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

Signature: _____________________________

1. Determine if the following situation is an observational study or experiment.

“If a university administrator looked at the relationship between gpa and admittance to medical school”

a.) Observational study
b.) Experiment

c.) Extrapolation

2. An instructor wanted to know more about the career goals of students that attended the high school in which he taught, so he gave his honors science class a survey that asked about their career goals. What type of error would this be?

a.) Influential Outliers
b.) Simpson’s Paradox
c.) Extrapolation
d.) Undercoverage
Questions 3-4: A group of individuals was surveyed who use StatCrunch.com. They were asked what age they thought that people should consider retiring. The histogram of their responses is below.

3. What shape best describes the histogram?
   a.) Left skewed
   b.) Right skewed
   c.) Bell shaped
   d.) Uniform

4. Describe the spread of the histogram.
   a.) 30 to 80
   b.) 0 to 500
   c.) 53 to 75
   d.) 100 to 500
Questions 5 - 6 In the US, the proportion of adults who are allergic to peanuts is 0.01. Suppose that ten individuals were selected independently of each other. Let $X$ equals the number of people with a peanut allergy.

5. What is the probability that one person is allergic to peanuts?
   a.) 0.01
   b.) 0.09
   c.) almost zero
   d.) 0.10

6. What would be the standard deviation of $X$?
   a.) 0.01
   b.) 0.099
   c.) 0.315
   d.) 0.0099

7. Which of the following is NOT a characteristic of the normal distribution?
   a.) symmetric
   b.) bell shaped
   c.) discrete
   d.) centered at the mean, $\mu$

8. A branch of the CDC reports from the National Health and Nutrition Examination Survey. A data brief was published from the 2011-2012 data in September 2015. Please see the excerpt below.

   Data from the 2011-2012 National Health and Nutrition Examination Survey (NHAMES) were used for these analysis. NHANES is a cross-sectional survey designed to monitor the health and nutrition status of the civilian noninstitutionalized U.S. population. The survey consists of interviews conducted in participant’s homes and standardized physical examinations in mobile examination centers (MECs).”

What aspect of statistics is discussed in the excerpt above?
   a.) Design
   b.) Description
   c.) Inference
9. Which of the following would be a continuous quantitative variable?
   a.) food group  
   b.) method of cooking  
   c.) amount of protein consumed (g)  
   d.) number of meals purchased per day at fast food restaurant(s)

10. What is the best type of graph to explore the relationship between the amount of sodium consumed per day and the amount of calories consumed per day?
    a.) histogram  
    b.) contingency table  
    c.) scatterplot  
    d.) boxplots

11. What is a simple random sample? A sample of size n where
    a.) every nth person is selected  
    b.) every person in the population is selected  
    c.) equal number of men and women  
    d.) each possible sample of size n has the same chance of being selected.

12. The method of least squares regression finds the best line by minimizing the sum of ________ .
    a.) the squared correlations  
    b.) the squared residuals  
    c.) the squared slope  
    d.) the squared horizontal distance from the points to the line

13. Based on the outside class video by Hans Rosling, answer the following question -- Initially, the predicted child mortality rate was much higher than the actual value. What would this tell us about the residual?
    a.) It would be negative.  
    b.) It would be positive.  
    c.) It would be zero.  
    d.) Unknown
**Question 14** - 16 Below is a boxplot that compares the longevity of a company’s employees with the company’s policy on continuing education (CE).

![Boxplot of Years worked for your company?](image)

14. What is the interquartile range for the years worked for a company that does not encourage continued education?

   a.) 2  
   b.) 6  
   c.) 15  
   d.) 20  
   e.) 22

15. What is the third quartile for the companies that do encourage continued education?

   a.) 2  
   b.) 5  
   c.) 12  
   d.) 30  
   e.) 42

16. Which of the following sentences best describes the comparison between the two types of companies?

   a.) The median years spent at a company that supports continued education is higher than those who don’t support CE and the range for those that do support is higher than those who don’t.

   b.) The median years spent at a company that supports continued education is lower than those who don’t support CR and the range for those that do support is higher than those who don’t.

   c.) The median years spent at a company that supports continued education is higher than those who don’t support CE and the range for those that do support is lower than those who don’t.

   d.) The median years spent at a company that supports continued education is lower than those who don’t support CE and the range for those that do support is lower than those who don’t.
17. A researcher was interested in predicting the amount of studying that a student did per week based off the amount of exercise they completed per day. In the Fall 2015 Beginning of the Semester Survey, the following information was found. The value of correlation was 0.2496. Find the least squares regression equation.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studying</td>
<td>16.75</td>
<td>14.33</td>
</tr>
<tr>
<td>Exercise</td>
<td>45.74</td>
<td>33.20</td>
</tr>
</tbody>
</table>

a.) $\hat{y} = 43.93 + 0.108x$
b.) $\hat{y} = 11.82 + 0.108x$
c.) $\hat{y} = -9.70 + 0.578x$
d.) $\hat{y} = 45.70 + 0.578x$

Questions 18 - 20 Use the below table to answer the following questions. In the 2014, General Social Survey participants were asked if they had attended a stress management class and about their political party.

<table>
<thead>
<tr>
<th>Attended Stress Management class</th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>15</td>
<td>65</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>Did not Attend Stress Management class</td>
<td>258</td>
<td>86</td>
<td>235</td>
<td>579</td>
</tr>
<tr>
<td>Total</td>
<td>301</td>
<td>101</td>
<td>300</td>
<td>702</td>
</tr>
</tbody>
</table>

18. What is the proportion of those surveyed that were Democrats and attended a stress management class?
   a.) $43 / 301$  b.) $43 / 123$  c.) $43 / 702$  d.) $123 / 702$  e.) $301 / 702$

19. What is the proportion of Republicans that did attend a stress management class?
   a.) $65 / 123$  b.) $65 / 300$  c.) $65 / 702$  d.) $123 / 300$  e.) $300 / 702$

20. What is the proportion of those that did not attend a stress management class this year that was Independents?
   a.) $86 / 579$  b.) $86 / 101$  c.) $86 / 300$  d.) $101 / 579$  e.) $579 / 702$
21. What does it mean if an observation has a value of z equal to -1.3?
   a.) The value is 1.3 times the value of the standard deviation.
   b.) The value is 1.3 times the value of the mean.
   c.) The standard deviation is less than what we would predict it to be.
   d.) The value is 1.3 standard deviations below the mean.
   e.) You can’t have a negative value of z.

Question 22 - 23 In Gainesville, Florida, residents use roughly 71 gallons of water per day on average (according to the GRU website). Suppose that water usage is normally distributed with a standard deviation of 12.

22. What is the first quartile of water usage?
   a.) 79.04
   b.) 55.96
   c.) 58.96
   d.) 60.96
   e.) 62.96

23. What proportion of residents use less than 80 gallons a day?
   a.) 0.77
   b.) 0.75
   c.) 0.73
   d.) 0.25
   e.) 0.23

24. The correlation between the divorce rate in Maine and the per capita consumption of margarine is 0.9926. Suppose that someone suggested that we should decrease the amount of margarine sold to decrease the divorce rate. What type of error is this?
   a.) Extrapolation
   b.) Misuse of Cause and Effect
   c.) Influential Outlier
   d.) Simpson’s Paradox
Question 25 - 27 Using data from the Fall 2015 beginning of the semester survey, a least squares regression equation was computed to predict the number of friends on Facebook based on the number of parties attended per week.

25. Interpret $R^2$ (if applicable).
   a.) 23.0% of the variability of the number of parties attended is explained by linear regression on the number of friends in Facebook.
   b.) For every party attended per week, there is an increase of 209.7 friends in Facebook.
   c.) For every friend in Facebook, the student tends to go to 23.0% more parties on average.
   d.) 23.0% of the variability of the number of friends in Facebook is explained by linear regression on the number of parties attended per week.

26. Interpret $y$-intercept (if applicable).
   a.) Do not interpret.
   b.) The predicted number of friends in Facebook when the student attends 0 parties is 209.7.
   c.) The predicted number of friends in Facebook when the student attends 0 parties is 465.4.
   d.) The predicted number of parties attended when the student has no friends is 209.7.
   e.) The predicted number of parties attended increases by 23.0% when the student reports that they have 0 friends.

27. One student reported attending one party and has 703 Facebook friends. What is the residual?
   a.) 675.1     b.) 27.9     c.) -27.9     d.) 151.2     e.) -151.2
28. Which statement about the standard deviation $s$ is false?
   
a.) $s$ can never be negative.
   
b.) $s$ can never be zero.
   
c.) $s$ is nonresistant (sensitive to outliers) measure of variability.
   
d.) For bell shaped distributions, about 95% of the data fall within $\bar{x} \pm 2s$.

Questions 29 – 30 In the United States, 1% of adults living in the United States have celiac disease. One test for celiac disease is tTG-IgA Test. This test correctly identifies 98% of patients with celiac disease and correctly identifies 95% of patients without celiac disease. Let C stand for the person who has celiac disease and POS stand for the patient has a positive test result.

29. What is the correct notation for the “95%”?
   
a.) $P(\text{POS}^c)$
   
b.) $P( C | \text{POS}^c )$
   
c.) $P(\text{POS}^c | C)$
   
d.) $P(\text{POS}^c | C^c)$
   
e.) $P(\text{POS}^c \text{ and } C^c)$

30. Find the probability that someone does not have celiac disease and gets a positive test result.
   
a.) 0.99
   
b.) 0.0495
   
c.) 0.9405
   
d.) 0.049
   
e.) 0.0002

31. The Physician Health Study Stage 3 had 14,641 male US physicians who participated in a study to see if taking a multivitamin would help prevent cancer. Half of the men received a placebo and the other half took a multivitamin every morning for a decade. What are the level(s) in this study?
   
a.) 14,641 male physicians
   
b.) Placebo and multivitamin
   
c.) Type of medication
   
d.) Occurrence of cancer
32. Read the below excerpt from an article about almonds posted on the Medical News Today website.

A study published in the Journal of the American Dietetic Association in 2005 suggested that consuming almonds increases vitamin E levels in the plasma and red blood cells and also lowers cholesterol levels.

Ella Haddad, DrPH, RD, an author of the study, said:

"This study is important because it shows that eating almonds can significantly boost levels of vitamin E in the diet and bloodstream. Vitamin E is a powerful antioxidant that defends your cells against damage on a daily basis and prevents artery-clogging oxidation of cholesterol. Eating a handful of almonds a day is a great way to get the vitamin E your body needs to stay healthy."

Which is the explanatory variable?

a.) eating almonds
b.) other food eaten daily
c.) American Dietetic Association
d.) Vitamin E in the bloodstream; cholesterol levels

33. Below is a scatterplot of the relationship between violent crime and property crime per year in the state of California.

![Scatterplot](image)

Estimate the value of r.

a.) -0.71
b.) 0.23
c.) 0.97
d.) -0.31
e.) 0.71