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. This test MUST BE SUBMITTED to the instructors together with the scantron sheet for you to receive a grade on the exam.

\[
\hat{y} = a + bx \\
b = r \frac{s_y}{s_x} \\
a = \bar{y} - bx \\
\bar{x} = \frac{\sum x_i}{n} \\
\binom{n}{r} = \frac{n!}{r!(n-r)!}
\]

\[
P(x) = \binom{n}{x} p^x (1-p)^{n-x} \\
\mu = np \\
\sigma = \sqrt{np(1-p)} \\
\mu = \sum xP(x) \quad \frac{1}{\sqrt{n}}
\]

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

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

Signature: ________________________________

1. Suppose that someone did an election poll for the upcoming presidential election. Suppose that someone randomly selected 500 potential voters and asked them if they had to choose today between Romney and Obama who would they chose. What is the margin of error?
   a) 0.000004
   b) 0.002
   c) 0.024
   d) 0.045
   e) Cannot be determined.

   \[
   \text{margin of error} = \frac{1}{\sqrt{500}} = 0.0447
   \]

   \[
   \frac{1}{\sqrt{500}} = 0.045
   \]
2. An online article on CNN health entitled “Dark Chocolate may lower blood pressure” posted by Leslie Wade, CNN Medical Producer talks about an experiment to see if dark chocolate reduces blood pressure. Below is an excerpt from that article. (article posted on 8/16/2012)

"Scientists in Melbourne, Australia, curious about the role of dark chocolate in heart health, looked at 20 studies in which adults ate dark chocolate or cocoa. More than 850 people participated in the trials that generally ran from two to eight weeks."

What is the explanatory variable for these experiments?
   a.) 850 people
   b.) Chocolate
   c.) Blood pressure levels
   d.) Genetics

3. For two events A and B, \( P(A) = 0.3 \) and \( P(B) = 0.2 \). Then \( P(A \text{ and } B) \) equals
   a.) 0.06, if A and B are independent
   b.) 0.10, if A and B are independent
   c.) 0.50, if A and B are independent
   d.) 0.10, if A and B are disjoint

4. What type of graph would be the best way to explore the relationship between income and credit card debt?
   a.) pie charts
   b.) boxplots
   c.) scatterplot
   d.) contingency table

5. Suppose that you are interested in predicting the price of a very good condition Acura MDX based on its age. You collect information about the past five model years and prices using Kelly’s Blue Book. Below is the summary statistics of your survey.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>stdev</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price ( y )</td>
<td>27930</td>
<td>5552</td>
<td>-0.986</td>
</tr>
<tr>
<td>Years ( x )</td>
<td>3</td>
<td>1.58</td>
<td></td>
</tr>
</tbody>
</table>

Compute the equation of the least squares regression line.
   a.) \( \hat{y} = -96792539.2 - 3462.54x \)
   b.) \( \hat{y} = 38324.19 - 3464.73x \)
   c.) \( \hat{y} = 10.84 - 0.00028x \)
   d.) \( \hat{y} = 27930 - 0.00028x \)
Questions 6-8: For students planning to enter a graduate program in elementary education, the students taking the old format (before June 2011) of the GRE verbal section had an average score of 440 and a standard deviation of 75. The GRE scores can be considered to be normally distributed.

6. What percentage of students in this population who score less than 400?
   a.) 0.47
   b.) 0.30
   c.) 0.53
   d.) 0.70

\[ z = \frac{400 - 440}{75} = -0.53 \]

7. What is the score of the student in this population that scores at the 90th percentile?
   a.) 526
   b.) 536
   c.) 546
   d.) 556

\[ z = 1.28 \quad \Rightarrow \quad x = z \sigma + \mu = 1.28(75) + 440 = 536 \]

8. Suppose that Orlando (a student planning on entering the elementary education graduate program) took the old format and scored a 434. Suppose that the new format of the GRE for students planning on entering a graduate program in elementary education has an average of 131 with a standard deviation of 8.75. Suppose that Nelly took the new format of the GRE and scored a 124. Who scored better on the test?
   a.) Nelly
   b.) Orlando
   c.) They performed the same.
   d.) It cannot be determined.

Nelly: \[ x = 124 \quad \Rightarrow \quad z = \frac{124 - 131}{8.75} = -0.8 \]
Orlando: \[ x = 434 \quad \Rightarrow \quad z = \frac{434 - 440}{8.75} = -0.68 \]

9. Which of the following is NOT a property of correlation, \( r \)?
   a) \( r \) is always between -1 and 1.
   b) \( r \) depends on which of the two variables is designated as the response variable.
   c) \( r \) measures the strength of the linear relationship between \( x \) and \( y \).
   d) \( r \) does not depend on the units of \( y \) or \( x \).
   e) \( r \) has the same sign as the slope of the regression equation.

10. Determine if the following statement is true or false.

"The mean is always one of the data points."

a) This is a true statement.
b) This is a false statement.

The mean does not have to be one of the data points.
Questions 11 – 13 In 2010, the General Social included a question that asked respondents if the approved or disapproved of increasing the use of nuclear power to generate electricity in the United States? Below is a table of the responses split up by political party. Answer the following questions based on the table.

<table>
<thead>
<tr>
<th></th>
<th>Strong Democrat</th>
<th>Independent</th>
<th>Strong Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly favor</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Favor</td>
<td>34</td>
<td>32</td>
<td>30</td>
<td>96</td>
</tr>
<tr>
<td>Oppose</td>
<td>17</td>
<td>27</td>
<td>3</td>
<td>47</td>
</tr>
<tr>
<td>Strongly oppose</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>75</td>
<td>41</td>
<td>183</td>
</tr>
</tbody>
</table>

11. What is the conditional proportion of strong democrats that strongly favor increasing the use of nuclear power to generate electricity in the United States?
   (a.) $\frac{7}{21}$
   (b.) $\frac{7}{183}$
   (c.) $\frac{7}{67}$
   (d.) $\frac{67}{183}$
   (e.) $\frac{21}{183}$

12. What is the proportion of those that strongly oppose increasing the use of nuclear power to generate electricity in the United States?
   (a.) $\frac{9}{67}$
   (b.) $\frac{9}{183}$
   (c.) $\frac{19}{67}$
   (d.) $\frac{67}{183}$
   (e.) $\frac{19}{183}$

13. What is the proportion of those that oppose the use of nuclear power to generate electricity in the United States given that they are a Strong Republican?
   (a.) $\frac{41}{183}$
   (b.) $\frac{47}{183}$
   (c.) $\frac{3}{47}$
   (d.) $\frac{3}{41}$
   (e.) $\frac{3}{183}$

FYI: In this case "strong" means that the respondent strongly identified with that party.
**Question 14 - 15** In 2007, the percentage of people in Botswana expected to have the HIV virus was 23.9%. Assume that this percentage has not changed. Suppose that a researcher selects 5 people at random from the population of Botswana.

14. What is the expected number of people in the sample to have the HIV virus?
   a.) 1  
   b.) 1.195  
   c.) 2  
   d.) 2.195  
   e.) Cannot be determined

\[ \mu = np = 5 \times 0.239 = 1.195 \]

Do not round.

15. What is the probability that two of the people sampled have the HIV virus?
   a.) 0.02517  
   b.) 0.020  
   c.) 0.2517  
   d.) 0.40  
   e.) None of the above.

\[ P(X = 2) = \binom{5}{2} \times 0.239^2 \times (1 - 0.239)^{5-2} = \frac{5!}{2! \times (5-2)!} \times 0.239^2 \times 0.761^3 = 0.2517 \]

**Questions 16 – 17** In the 2012 Beginning of the Semester Survey for STA 2023 students, respondents were asked, “How many continents have you visited? (Count North America as 1 continent visited.)” Answer the following questions about the boxplots below.

16. Which data set has a larger range?
   a.) Females  
   b.) Males  
   c.) The values are the same.  
   d.) Can’t be determined.

\[ \sqrt{\text{Females}} \Rightarrow \text{Range} = \text{max} - \text{min} = 7 - 1 = 6 \]

\[ \text{Males} \Rightarrow 5 - 1 = 4 \]

17. For females, what percentages of students have visited 2 or less continents?
   a.) 0%  
   b.) 25%  
   c.) 50%  
   d.) 75%  
   e.) 100%
18. Because of increasing nuisance of spam email messages, many start-up companies have emerged to develop e-mail filters. One such filter was recently advertised as being 95% accurate. The way the advertisement is worded, 95% accurate could mean that "95% of the valid email is allowed through". Let S denote {message is spam}, and let B denote {filter blocks message}. Using these events and their complements, identify the probability stated in bold. (Homework Question 5.29)

\[ \text{valid email} = S^c \quad \text{(not span)} \]
\[ \text{allowed through} = B^c \quad \text{(not blocked)} \]
\[ P(B^c | S^c) \]

19. According to Statistics Canada, in 2004 the median household in Canada was $58,100 and the mean was $76,100. What would you predict about the shape of the distribution? (Homework problem 2.41)

a. Skewed to the left
b. Skewed to the right
c. Bell shaped
d. Uniform

If the distribution was bell-shaped, the mean would be almost equal to the median, but you can't go 3 standard deviations to the left of $58,100 w/o getting a negative income. So dist.

20. A random number generator is used to generate a real number between 0 and 2, equally likely to fall anywhere in this interval. Find the probability that a number falls between 0.5 and 1.

a. 0.25
b. 0.5
c. 0.75
d. 1.0

21. The slope of the regression equation and the correlation are similar in the sense that (Homework problem 3.107)

a. They do not depend on the units of measurement
b. They both must fall between -1 and 1.
c. They both have the same sign.
d. neither can be affected by severe regression outliers.

\[ b = r \frac{S_y}{S_x} \quad \text{since } S_y \text{ and } S_x \text{ are always positive, } b \text{ and } r \text{ must have the same sign.} \]
22. Below is a scatterplot using data from the STA 2023 Beginning of the Semester Survey 2012. The questions asked on the survey were, “How much change do you have in your pocket right now?” and “How many religious services do you attend per week?”

If the point (12, 420) is removed, it is likely that . . .
   a.) That the value of r would increase.
   b.) That the value of r would decrease.
   c.) That the value of r would remain the same.


   “Young women at high familial risk for breast cancer may see an even greater risk from diagnostic scans that expose them to chest radiation, including mammograms, researchers found. Any diagnostic use of radiation before age 30 increased breast cancer risk by 90 percent for carriers of BRCA1 or BRCA2 mutations, according to a retrospective analysis of three nationwide cohorts in Europe by Flora E. van Leeuwen of the Netherlands Cancer Institute in Amsterdam, and colleagues. “

What is the response variable?

a.) Radiation
b.) BRCA 1 and BRCA 2 gene mutations
c.) Cancer risk
d.) Family history
24. The high temperature was recorded for the first seven days of 2009 at the Physics Building at the University of Florida. The temperatures are below. Find the standard deviation of this sample.

\[ \text{\text{\text{S_x = 1.40 Use your calculator.}}} \]

\[ \begin{align*}
\text{a.)} & \quad 1.00 \\
\text{b.)} & \quad 1.10 \\
\text{c.)} & \quad 1.20 \\
\text{d.)} & \quad 1.30 \\
\text{e.)} & \quad 1.40
\end{align*} \]

25. The Clear Blue Pregnancy Test claims that it is 99% accurate. Suppose that this means that 99% of women that are pregnant are accurately identified as being pregnant and suppose that 95% of women that are not pregnant are accurately identified as not being pregnant. Suppose that 70% of women that use a pregnancy test are actually pregnant. What proportion of all women that use the test are not pregnant AND get a positive reading?

\[ \begin{align*}
\text{P} & = \text{pregnant} \\
\text{t} & = \text{test is positive} \\
P(\text{t} = 1 | \text{P}) & = .99 \\
P(\text{t} = 1 | \text{P}^c) & = .95 \\
P(\text{P} | \text{t}) & = .7
\end{align*} \]

\[ \begin{align*}
\text{test Pos } P(\text{t} = 1 | \text{P}) & = .99 \\
P(\text{t} = 1 \wedge \text{P}) & = .7(0.99) \\
& = .693
\end{align*} \]

\[ \begin{align*}
\text{test Neg } P(\text{t} = 0 | \text{P}) & = .01 \\
P(\text{t} = 0 \wedge \text{P}) & = .7(0.01) \\
& = .007
\end{align*} \]

\[ \begin{align*}
\text{test Pos } P(\text{t} = 1 | \text{P}^c) & = .05 \\
P(\text{t} = 1 \wedge \text{P}^c) & = .3(0.05) \\
& = .015
\end{align*} \]

\[ \begin{align*}
\text{test Neg } P(\text{t} = 0 | \text{P}^c) & = .95 \\
P(\text{t} = 0 \wedge \text{P}^c) & = .3(0.95) \\
& = .285
\end{align*} \]

\[ \begin{align*}
\text{P(Not P and t = 1)} & = \text{P(t = 1 \wedge \text{P}^c) + P(t = 1 \wedge \text{P})} \\
& = .015 + .693 \\
& = .708
\end{align*} \]
26. Here is a stem and leaf plot of the heights (inches) of students who completed the Beginning of the Semester Survey in the Fall 2012. Find Q₃.

**Stem-and-Leaf Display: Height**

<table>
<thead>
<tr>
<th>Stem-and-leaf of Height</th>
<th>N  = 222</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf Unit</td>
<td>1.0</td>
</tr>
<tr>
<td>N* = 3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>4</th>
<th>555</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>26</td>
<td>6</td>
<td>00000000000111111</td>
</tr>
<tr>
<td>67</td>
<td>6</td>
<td>222222222222222333333333333333333</td>
</tr>
<tr>
<td>47</td>
<td>6</td>
<td>444444444444444444444555555555555555555</td>
</tr>
<tr>
<td>108</td>
<td>6</td>
<td>666666666666666666666666666667777777777777777</td>
</tr>
<tr>
<td>67</td>
<td>6</td>
<td>88888888888888888888888888888888888888888888888888888888</td>
</tr>
<tr>
<td>45</td>
<td>7</td>
<td>00000000001111111</td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>2222222222222223333333</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>444445</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>677</td>
</tr>
</tbody>
</table>

- a.) 67 inches
- b.) 68 inches
- c.) 69 inches
- d.) 70 inches
- e.) 71 inches

**Position of Median**

\[
\frac{222 + 1}{2} = 111.5
\]

**Position of Q₃**

\[
\frac{111 + 1}{2} = 56
\]

\[6.9 = Q₃\]
Questions 27 – 29 Below is a plot of the relationship between the protein content and the cost of a Subway sandwich. The amount of protein is measured in grams and the cost in dollars.

27. Interpret the y-intercept, if applicable.
   a.) The predicted amount of protein in a sandwich which costs 0 dollars in 4.33 ounces.
   b.) The predicted amount of protein is a sandwich which costs $4.22 is 0 grams.
   c.) As the cost increases by 1 dollar, the amount of protein increases by 7.562 g on average.
   d.) As the amount of protein increases by 1 gram, the amount of protein decreased by 4.218g.
   e.) Do not interpret.

28. One sandwich costs $3.69 and had 26 grams of protein. Find the residual.
   a.) 23.69  
   b.) 2.01  
   c.) -2.01  
   d.) 2.31  
   e.) -2.31

29. Find the value of correlation.
   a.) 0.769  
   b.) -0.769  
   c.) 0.349  
   d.) -0.34  
   e.) 0.756

\[
pred_y = -4.218 + 7.562x 
= -4.218 + 7.562(3.69) 
= 23.69 
\]

\[
\text{residual} = obsy - pred_y 
= 26 - 23.69 
= 2.31 
\]
30. In 1995 in the United Kingdom, the Equality Code used by legal profession added a section to make members more aware of sexual harassment. It states “research for the Bar found that over 40 percent of female junior tenants said they had encountered sexual harassment during their time at the Bar.” This was based on a study conducted at the University of Sheffield that sent a questionnaire to 334 junior tenants at the Bar, of whom 159 responded. Of the 159, 67 were female. Of those females, 3 said they had experienced it as a slight problem. Which of the following is a correct statement about these results?

a. The quoted statement might be misleading because the nonresponse was large.
b. No one was forced to respond, so everyone had a chance to be in the sample, which implies it was a random sample.
c. This was a retrospective case-control study, with those who received sexual harassment as the cases.
d. It is impossible to learn useful results about a population based only a sample of 159 people.

31. On September 12th 2012, Gallup reported “Those with incomes of less than $24,000 per year report using their strengths an average of 5.8 hours a day compared with averages of 6.7 to 7.0 for higher-income groups.” These results were based on a “random sample of 5049 adults, aged 18 and older, living in all 50 US states and the District of Columbia, selected using random digit dial sampling”.

Are the values “5.8, 6.7 and 7.0” examples of statistics or parameters?

a.) 5.8 is an example of a statistic and 6.7 and 7.0 are examples of parameters
b.) 6.7 and 7.0 are examples of parameters and 5.8 is an example of a statistic

32. In 2008, the General Social survey asked 504 women respondents if they had health insurance from their employer. Out of 504 people, 373 said yes that they had health insurance from their employer. From this we can predict that between 70 to 78% of women in the United States has health insurance from their employer. The statement in bold is an example of what area of statistics.

a.) description
b.) design
c.) inference
d.) data file

33. Which of the following is a categorical variable?

a.) miles traveled by the Olympians to London
b.) hours trained for the Olympics
c.) sport
d.) amount of fluids consumed before competing