EXAM 1  Test Form Code B

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.

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

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

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

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

Signature: ______________________

categorical data

categorical data

categorical data

categorical data

1. Suppose that you were interested in studying the relationship between someone’s political party and if they would vote for a women president. What type of graphical summary would we use?
   a.) Scatterplot
   b.) Boxplots
   c.) Contingency table
   d.) Histograms

   If you have 2 variables that are categorical, you use a contingency table to explore the relationship.
2. Suppose that someone decided to stand on a street corner and determine if drivers are texting while driving. Assume that the drivers are independent. She is going to stand on the street corner until she has counted ten drivers that were texting. Let $X$ equal the number cars that went by until she counted ten texting drivers. Is this a Binomial distribution?

a.) Yes, because the trials are independent.

b.) Yes, because there are two possible responses.

c.) No, because there is not a constant sample size.

d.) No, because the trials are supposed to be dependent.

3. The hitting probabilities for 2013 regular season for a local baseball team are below.

<table>
<thead>
<tr>
<th>Number of Bases</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.65</td>
<td>0.2</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Find the mean number of this distribution.

a.) 2

b.) 1

c.) 0.66

d.) 0.132

e.) 0.2

$$\text{Mean} = 0(0.65) + 1(0.2) + 2(0.05) + 3(0.04) + 4(0.06)$$

$$= 0.60$$

Question 4 - 5 Suppose that you have a balanced twelve sided dice. Each side was given a number 1 through 12. Suppose that you rolled 100 times. Let $X$ equal the number of times that the dice landed with the "7" on top.

4. Find the standard deviation of $X$.

a.) 8.33

b.) 7.62

c.) 5

d.) 3.73

e.) 2.76

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{100 \cdot \left(\frac{1}{12}\right) \cdot \left(\frac{11}{12}\right)} = 2.76$$

5. Find the probability that $X = 1$.

a.) 0.083

b.) 0.0015

c.) 0.0000015

d.) 3.20

$$P(X=1) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{100}{1} \left(\frac{1}{12}\right)^1 \left(1 - \frac{1}{12}\right)^{100-1}$$

$$= 100 \cdot \left(\frac{1}{12}\right)^1 \cdot \left(\frac{11}{12}\right)^{99}$$

$$= 0.0015$$
6. Which of the following is NOT a characteristic of the standard deviation?
   a.) The standard deviation can be positive or negative. [True]
   b.) The standard deviation is affected by outliers. [True]
   c.) The standard deviation works well with symmetric distributions. [True]
   d.) The standard deviation is larger for data sets that are more spread out. [True]

   See the top of page 15 in the lab workbook.

7. Suppose that X~ Normal (23, 2.4). What is the P(x < 23)?
   a.) 0%
   b.) 25%
   c.) 50%
   d.) 75%
   e.) Unknown
   Because the Normal distribution is symmetric, the mean is equal to the median.
   So, 50% of the data will be below 23 (the mean).

8. Below are the calories for 8 randomly selected appetizers at Applebees.
   1280 1330 540 1290 930 1420 940 900

   Find the standard deviation.
   a.) 299.59
   b.) 280.24
   c.) 1078.75
   d.) 123.33
   Use your calculator to find the sample standard deviation.

9. In 2013, a random sample of 164 students was selected. The least squares regression line was then calculated to predict the average math score based on the average verbal score. The least squares regression line was \( y = 209.55417 + 0.675x \). For one student, the verbal score was 590 and the math score was 570. Find the residual.
   a.) 607.8
   b.) -37.8
   c.) 37.8
   d.) -4.3
   e.) 4.3

   \( \hat{y} = 209.55417 + 0.675(590) \)
   \( \hat{y} = 607.8 \)
   \( \text{res} = \text{obs}\ y - \text{pred}\ y = 570 - 607.8 = -37.8 \)
Questions 10 - 11 Suppose that 55% of Caucasian women between the ages of 30 and 35 that take a pregnancy test are actually pregnant. Suppose that the pregnancy test RapidResults accurately identifies pregnant women 95% of the time and accurately identifies women who are not pregnant 97.5% of the time. Let $P\text{REG}$ stand for pregnant and $P\text{OS}$ stand for the test gives a positive result.

10. How would you represent the “97.5%” above using symbols?
   a.) $P(\text{PREG} \mid \text{POS})$
   b.) $P(\text{PREG}^c \mid \text{POS})$
   c.) $P(\text{POS} \mid \text{PREG})$
   d.) $P(\text{POS}^c \mid \text{PREG})$
   e.) $P(\text{POS} \mid \text{PREG}^c)$

11. What is the probability of a woman in this demographic “not being pregnant and having a negative result”? 
   a.) 0.975
   b.) 0.43875
   c.) 0.438
   d.) 0.4835
   e.) 0.01125

12. Suppose that we were interested in looking at the relationship between gender and choice of transportation to campus. Assume the following notation.

   $M =$ Male, $F =$ Female, $B =$ Bus, $W =$ Walk, $K =$ Bike and $C =$ Car.

   How would we check to see if $M$ and $B$ were independent?
   a.) $P(\text{M} \mid \text{B}) = P(\text{M})$
   b.) $P(\text{M and B}) = P(\text{M})$
   c.) $P(\text{M or B}) = P(\text{B})$
   d.) $P(\text{M and B}) = P(\text{M}) + P(\text{B})$

13. In the Census for Schools video about the Cholera epidemic, Dr. John Snow was able to stop the spread of cholera after his investigation. Why was this observational study rather than an experiment?
   a.) He wasn’t able to use randomization.
   b.) He wasn’t completely able to eliminate all lurking variables.
   c.) Simpson’s paradox was in effect.
   d.) He wasn’t trying to predict into the future.
14. Which of the following graphs has the smallest standard deviation?

a.) Graph a
b.) Graph b

c.) Graph c  

Standard deviation is the typical distance of the points to the mean. In graph c, more of the points are closer to the mean, so it has a smaller stdv.

d.) A and B have the smallest variation.

e.) They are all the same.

15. Below are the responses from the Beginning of the Semester survey.

What is the approximate range of text messages for males?

a.) 500  b.) 800  c.) 200  d.) 100  e.) 75

Range = 300 - 0 = 300
16. A blogger wants to estimate the proportion of Americans that plan on voting in the next election. He decides to post a survey on his blog. Amazingly, three hundred thousand people respond. Ninety-nine percent of the sample claims that they plan on voting in the next election. The blogger claims that since he had such a large sample he can conclude that 99% of Americans plan on voting in the next election. Is this true?

a.) Yes, he has such a large sample that he can make a conclusion about all Americans.
b.) Yes, he has such a large sample and the poll was posted on the internet so anyone could have responded.
c.) No, this is an example of a convenience study.
d.) No, this is an example of extrapolation.
e.) No, this is an example of Simpson’s Paradox.

Questions 17 – 18 Below are the results of the General Social Survey from 2012. Respondents were asked if they strongly agreed, agreed, disagreed or strongly disagreed with the following statement.

“The job of a scientist would be boring.”

The responses have been broken down by gender.

<table>
<thead>
<tr>
<th></th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5</td>
<td>35</td>
<td>156</td>
<td>26</td>
<td>222</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>44</td>
<td>170</td>
<td>32</td>
<td>251</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>79</td>
<td>326</td>
<td>58</td>
<td>473</td>
</tr>
</tbody>
</table>

17. What proportion of males disagreed with the statement?
   a.) $156/473$  
   b.) $156/326$  
   c.) $156/222$  
   d.) $222/473$  
   e.) $326/473$

18. Given that someone strongly disagreed, what is the conditional proportion that they are female?
   a.) $32/58$  
   b.) $32/251$  
   c.) $32/473$  
   d.) $251/473$  
   e.) $58/473$

19. A farmer was trying to determine the best type of corn to grow with the most effective fertilizer. He selected three varieties of corn (Silver Queen, Jubilee, and Golden Cross Bantram). There are about 100 plants of each variety. He then selected two types of fertilizer one with 16 parts potassium and another with 12 parts potassium. He divides each varietal into two groups, each group gets a different type of fertilizer. What is the number of treatments?
   a.) 2
   b.) 3
   c.) 6
   d.) 50
   e.) 300

   Variety of corn $\Rightarrow$ 3 levels
   Type of fertilizer $\Rightarrow$ 2 levels
   Treatments $\Rightarrow 3 \times 2 = 6$
20. For the "Beginning of the Semester Survey" the correlation between number of friends in Facebook and the number of text messages received was 0.199. The least squared regression equation was 

\[ y = 565.10 + 0.6799x. \]

What would be the value of \( R^2 \)?

a.) 0.0396  
b.) 0.4461  
c.) 0.4623  
d.) 0.6799

\[ (0.199)^2 = .0396 \]

Question 21 - 22. The Bayley Scales Test for Early Development has a score that has a normal distribution. The mean is equal to 100 and the standard deviation is equal to 15.

21. What proportion of scores are lower than 81?

a.) -1.27  
b.) 0.1020  
c.) 0.0668  
d.) 0.1210  
e.) 0.90

\[ z = \frac{81 - 100}{15} = -1.27 \]

22. What is the score at the 95\(^{th}\) percentile?

a.) 135  
b.) 125  
c.) 115  
d.) 105  
e.) 95

\[ z = 1.645 \]

\[ x = z \sigma + \mu = 1.645(15) + 100 = 124.67 \]

Questions 23- 24. An ABC News Report from Sept. 2, 2014, by Gina Jabbour, MD, summarized the results from an NIH-funded study by Bazzano et al. They studied the difference in weight loss for 73 low fat dieters and 75 low carb dieters. After one year, the low carb dieters in this study had lost 8 more lbs. than the low fat dieters.

23. What is the response variable?

a.) Weight loss  
b.) Type of diet  
c.) Amount of Exercise  
d.) 148 participants

24. Is the "8lbs" a statistics or a parameter?

a.) Statistic  
b.) Parameter
Questions 25-27 A baseball enthusiast wanted to predict the number of games a team won based on the number of strikeouts from the pitcher. Below is the regression output.

![Fitted Line Plot](image)

25. Interpret the slope (if applicable).
   a.) The average increase in the number of strike outs is 2.813 for every additional game won.
   b.) The predicted number of games won when the strikeouts equal 0 is 3.984.
   c.) For each additional strike out, the number of wins tends to increases by 0.05656 on average.
   d.) The predicted number of strikeouts is 0.05656 when the number of games won is 0.
   e.) Should not be interpreted.

26. Predicting the number of wins when a team had 330 strikeouts would be an example of?
   a.) An appropriate use of the least squares regression equation
   b.) Extrapolation ➔ predicting too far from the observed data
   c.) Simpson’s Paradox
   d.) Misuse of Cause and Effect
   e.) Volunteer Sampling

27. Find the value of $r$, the correlation coefficient.
   a.) 0.621
   b.) -0.621
   c.) 0.562
   d.) 0.386
   e.) -0.386

\[
\sqrt{0.386} = 0.621
\]
28. A survey of homes for sale in Albuquerque in 2014 was taken. Jorge was interested in predicting the price of the house based on the square footage. He found that the average price of the homes was $259,120 and the standard deviation was $76,460.255. The average square footage was 1,943.7 and the standard deviation was 255.69. The correlation coefficient was 0.559. Find the least squares regression equation.

\[ \hat{y} = -65789.97 + 167.16x \]

b.) \[ \hat{y} = 259147.62 + 0.014210x \]

c.) \[ \hat{y} = 259123.63 + 0.00187x \]

d.) \[ \hat{y} = 43316442.9 + 167.16x \]

e.) \[ \hat{y} = -4331255.5 + 0.00187x \]

\[ b = r \frac{s_y}{s_x} = 0.559 \left( \frac{76460.255}{255.69} \right) = 167.16 \]

\[ a = \bar{y} - b \bar{x} = 259,120 - (167.16)(1943.7) = -65789.97 \]

29. Below is a stem and leaf plot for the number of strikeouts for Major League Baseball for 40 pitchers.

**Stem-and-Leaf Display: Strike Outs**

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>22223</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4555</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>66677777</td>
</tr>
<tr>
<td>(10)</td>
<td>1</td>
<td>889999999</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>004</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>222233</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>45</td>
</tr>
</tbody>
</table>

Find the median.

a.) 18
b.) 19
\[ \text{Median} = \frac{180 + 180}{2} = 180 \]
d.) 190
\[ \text{Median} = \frac{180 + 180}{2} = 180 \]
e.) 185
30. Below is a histogram of the results of the “Beginning of the Semester Survey”. This graph represents the amount of time spent studying.

Describe the shape.
- Bell shaped
- Bimodal
- Right Skewed
- Left Skewed
- Uniform

31. What measure of spread is considered resistant to outliers?
- Mean
- Median
- Mode
- Range
- IQR

32. Exam scores for a Biology class range from 58 to 96. The average on the exam is usually about 78. What would be a realistic value of the standard deviation?
- -10
- 0
- 6
- 15
- 50

33. Which of the following variables would be considered a categorical variable?
- Weight of fish
- Length of fish
- Species of fish
- Speed of fish through water