Math 251, Final Exam

Instructions: Do 11 of the following 12 questions. Each question is worth 10 pts. The first two questions each consist of several multiple choice, true/false or short answer questions. Please do your best by justifying all of your problem solutions with appropriate work.

1. (i) (2 pts) Given a normal probability distribution, which of the following is true?
   (a) The median is to the right of the mean.
   (b) The mean is to the right of the median.
   (c) The mean is equal to the median.
   (d) Any of the above can occur depending on whether the normal distribution is symmetric, negatively or positively skewed.
   Answer: 

(ii) (2 pts) According to Chebyshev’s theorem, at least what proportion of data in any distribution lies within 5 standard deviations of the mean?
   Answer: 

(iii) (1 pt) (True or False) Two events in a sample space are independent if they never occur at the same time.
   Answer: 

(iv) (1 pt) (True or False) In an hypothesis test on the mean, a Type I error occurs if the null hypothesis is rejected when it is true.
   Answer: 

(v) (2 pts) For events A and B in a sample space S, we are told $P(A) = .6$ and $P(B) = .3$ and $P(A \text{ and } B) = .2$. Which of the following is true?
   (a) $P(A|B) = 2/3$.
   (b) $P(A \text{ or } B) = .9$
   (c) A and B are mutually exclusive.
   (d) None of the above are true.
   Answer: 

(vi) (2 pts) Which of the following is true about a binomial random variable for 10 trials with probability of success on each trial given as .1?
   (a) The probability of no successes is $(.1)^{10}$
   (b) The 10 trials must be identical and independent from one another.
   (c) The probability of 10 successes is $(.9)^{10}$
   (d) All of the above.
   Answer: 

2. (i) (2 pts) Suppose that the maximum error $E = 4$ for a 95% confidence interval for the mean using a sample of size $n = 100$. What sample size would you expect to use from the same population for a 95% confidence interval of the mean to have a maximum error $E = 1$?

**Answer:**

(ii) (2 pts) According to the empirical rule for bell shaped distributions, approximately what percentage of data lies within one standard deviation of the mean?

**Answer:**

(iii) (2 pts) Suppose a weight of 50 pounds is at the 67th percentile for Springer Spaniel dogs. In a collection of 500 randomly selected Springer Spaniel dogs, how many would you expect to weigh more than 50 pounds?

**Answer:**

(iv) (1 pt) What z-score would you assign to a number that lies 3.5 standard deviations below the mean?

**Answer:**

(v) (1 pt) (True or False) In hypothesis testing, the null hypothesis should be rejected if the level of significance is smaller than the $P$-value.

**Answer:**

(vi) (2 pts) Unaware that 33% of the 8,000 voters in his district still support him, a politician decides to estimate his political strength. A sample of 900 voters shows that 36% support him. Which of the following is true?

(a) The parameter of interest is 36%.

(b) The statistic of interest is 33%.

(c) The sample size is 8000.

(d) None of the above.

**Answer:**
3. (a) (2 pts) In how many ways can the state of Nevada make license plates of the from yxx-aaa where the first character is a number 1 - 9, the second and third characters are numbers 0 - 9, and the last three characters are letters A - Z (all uppercase)?

(b) (2 pts) Given that the probability that a newborn child is female is 48% and 52% that the child is male. What is the probability that exactly 11 of the first 20 babies born in 2003 were female?

(c) (2 pts) In how many ways can three medals (gold, silver and bronze) be awarded in a 100m race that contains 9 runners?

(d) Consider the random variable $x$ whose probability distribution is given by the following table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x)$</td>
<td>.3</td>
<td>?</td>
<td>.3</td>
<td>.1</td>
</tr>
</tbody>
</table>

(i) (1 pt) Find $P(x = 4)$. Answer =

(ii) (3 pts) Find the expected value of $x$ and the standard deviation of $x$. 
4. Consider the sample of 30 numbers

   31 37 44 46 47 51 52 65 68 68 69 70 71 72 73
   75 75 76 76 77 78 79 80 81 82 83 83 85 88 89

(a) (3 pts) Construct a stem and leaf plot for this data with stems 3,4,5,6,7,8

(b) (7 pts) Construct a relative frequency histogram for the data where the first class has limits 30 – 39. State the class width, and list the limits, boundaries and relative frequencies for all classes.

5. (10 pts) Given the data 9,12,15,17,18,19,23,45,52,61,63,63. One has \( \sum x = 397 \) and \( \sum (x - \mu)^2 \approx 5206.917 \). Find:

(a) the mean

(b) the sample variance

(c) the population variance

(d) the sample standard deviation

(e) the median

(f) the 50th percentile of the data

(g) the mode
6. The following data is for investigating the relation between salary in thousands (x) and average number of absences per year (y).

<table>
<thead>
<tr>
<th>Salary (x)</th>
<th>20</th>
<th>23</th>
<th>28</th>
<th>30</th>
<th>33</th>
<th>35</th>
<th>37</th>
<th>40</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences (y)</td>
<td>2.4</td>
<td>2.2</td>
<td>1.9</td>
<td>2.1</td>
<td>1.5</td>
<td>1.4</td>
<td>1.3</td>
<td>0.5</td>
<td>0.4</td>
</tr>
</tbody>
</table>

For this data: \( \sum x = 288, \sum x^2 = 9660, \sum y = 13.7, \sum y^2 = 24.93, \sum xy = 398.2 \).

(a) (4 pts) Find the equation of the least squares regression line.

(b) (2 pts) Does the data appear to be positively or negatively correlated? Explain.

(c) (2 pts) How many absences per year would you expect from an employee that makes $38,000 per year?

(d) (2 pts) Given that the correlation coefficient is \(-0.945\), determine how well the data fits the line of best fit. Explain your answer.

7. A business employs 300 men and 700 women. Of these employees, 60 men and 210 women have been working there for more than 10 years. Let \( A \) be the event the employee is a woman, and \( B \) be the event the employee has been employed for more than 10 years.

(a) (3 pts) Find \( P(A) \) and \( P(B) \).

(b) (3 pts) Find \( P(A \text{ and } B) \) and \( P(A \text{ or } B) \).

(c) (2 pts) Are the events \( A \) and \( B \) independent? Explain.

(d) (2 pts) Find the probability an employee is female, given that the length of employment is more than 10 years.
8. In a recent survey of 180 randomly selected engineering graduates starting work in the United States, it was found that the average annual starting salary for the sample was $43,200 with a standard deviation of $6,000.

(a) (7 pts) Conduct an hypothesis test to determine whether the average starting salary for engineers is more than $42,500 per year using a significance level of $\alpha = .05$. Make sure to state $H_0$, $H_a$, the critical region, and your conclusion.

(b) (3 pts) If the true mean starting salary for engineers in the United States is actually $43,000 with standard deviation $6,000, what is the probability that a random sample of 180 starting engineers will have a sample mean starting salary less than $44,000?

9. A “Gallop” poll of 1100 randomly selected Americans found that 47.3 percent think the Elian Gonzales story was the top news event of 2000.

(a) (5 pts) Conduct an hypothesis test to determine if the true population proportion of Americans who think the Elian Gonzales story was the top news event of 2000 is less than .50. Report the P-value of your test and conclusion at a level of significance of $\alpha = .05$.

(b) (2 pts) Based on this information, would you be comfortable saying that less than 50 percent of Americans think the Elian Gonzalez story was the top news event in 2000. Explain.

(c) (3 pts) What sample size would be needed so that you could estimate the true proportion in a population within ±.02 with 99 percent confidence?
10. In a 1993 survey of 150 Education graduates and 200 Social Science graduates, the following data were obtained for their average starting salaries.

<table>
<thead>
<tr>
<th>Major</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>24,500</td>
<td>2600</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>21,500</td>
<td>2400</td>
</tr>
</tbody>
</table>

(a) (5 pts) Conduct an hypothesis test to determine whether $\mu_1 - \mu_2 < 3800$ using a 5% level of significance where $\mu_1$ is the population mean salary for the Education graduates and $\mu_2$ is the population mean salary for the Social Science graduates. Make sure to state $H_0$, $H_a$, the rejection region, and your conclusion.

(c) (5 pts) Find a 99 percent confidence interval for $\mu_1$, the mean salary of Education graduates. Does the confidence interval suggest that $\mu_1 > 24,000$? Explain.

11. (a) (2 pts) Find the value $z_c$ needed in the formula for constructing a 94% confidence interval from a large sample.

(b) (2 pts) What assumptions must be made on the population and/or sample when constructing a confidence interval for the mean using a large sample?

(c) (2 pts) Find the rejection region for a two-tailed hypothesis test at level of significance $\alpha = .05$ on a mean using a sample of size $n = 12$ from an approximately normal population with unknown standard deviation?

(d) (2 pts) What assumptions must be made on the population and or sample to do a hypothesis test on a proportion?

(e) (2 pts) Find the rejection region for a two-tailed test on a population proportion at a level of significance of $\alpha = .06$ assuming the appropriate conditions as requested in (d) are satisfied.
12. (a) A company wishes to check whether the mean weekly production levels at their five factories are equal using the method of analysis of variance.

(i) (2 pts) State what conditions should be satisfied by the populations and samples in order to use the method of analysis of variance.

(ii) (1 pt) State the null and alternative hypotheses.

(iii) (2 pts) Given that random samples of size 4 were obtained from each factory. What is the rejection region for the hypothesis test conducted at a level of significance of $\alpha = 0.05$.

(b) (5 pts) Suppose that in an attempt to determine whether a die is fair, you tossed it 120 times with the following results:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

Does this provide sufficient evidence at a level of significance of $\alpha = 0.05$ to conclude that the die is not fair? Be sure to find the test statistic, rejection region, and state the conclusion.