Applied Statistics  
Final Exam

Name:________________________  
ID:________________________

Carefully Read The Instructions!

Instructions: This exam will last 120 minutes and consists of 9 problems and two extra credit problems, each one being worth 10 points. Provide solutions to the problems in the space provided. All solutions must be sufficiently justified to receive credit. You may use a calculator on this exam. You may **not** use any outside notes or texts. Good Luck!

**Advice**: If you get stuck on a problem don’t panic! Move on and come back to it later.

<table>
<thead>
<tr>
<th>Page Number</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Extra Credit1</td>
<td></td>
</tr>
<tr>
<td>Extra Credit2</td>
<td></td>
</tr>
</tbody>
</table>
1. Suppose that a discrete random variable $X$ has pmf

$$p(x) = \begin{cases} 
.25 & \text{if } x = 1, \\
.13 & \text{if } x = 2, \\
.14 & \text{if } x = 3, \\
.20 & \text{if } x = 4, \\
.28 & \text{if } x = 5.
\end{cases}$$

(a) Compute the CDF of $X$

(b) Compute $Var(X)$

(c) Compute $E[(2X - 2)^2]$
2. A rare disease afflicts only .1% of the population. A test has been developed to test for the presence of the disease in individual patients. If the test is performed on an individual who has the disease, the probability of a positive test result (i.e. the test says that they have the disease) is .88. If the test is performed on an individual who does not have the disease, the probability of a positive test result is .08. Suppose the test is performed on a randomly selected individual.

(a) What is the probability that the test result will be positive?

(b) Suppose the test comes back positive. What is the probability that the individual actually has the disease?
3. A bag contains 8 red marbles, 5 green marbles, and 7 blue marbles.

(a) Suppose that two marbles are removed from the bag at random. What is the probability that at least one of them is red?

(b) Suppose that 6 marbles are removed from the bag at random, what is the probability that all of them are blue?

(c) Suppose that one marble is removed from the bag at random, the color of the marble is recorded, and then the marble is put back into the bag. If this process is repeated independently 10 times, what is the probability that the color green is recorded more than once?
4. Suppose that a population has a continuous distribution whose probability density function is
\[ f(x) = \begin{cases} 
    x^2 & : 3 \leq x < c \\
    0 & : otherwise 
\end{cases} \]

(a) Compute \( c \) so that the function above is a valid probability density function.

(b) Suppose \( X_1, \ldots, X_{40} \) is a sample of size 40 from this population. What is the approximate distribution of the sample mean \( \bar{X} \)? Be sure to reference any theorems/results which are used to compute this distribution.

(c) What is the approximate probability that \( \bar{X} \) is less than 4?
5. Suppose $X$ and $Y$ are independent, identically distributed exponential(2) random variables, i.e. they both have density function

$$f(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} &: 0 \leq x < \infty \\ 0 &: otherwise \end{cases}$$

(a) Compute the joint density of $X$ and $Y$.

(b) Compute $E(XY)$.

(c) Compute $P(X < Y + 3)$ (Hint: draw the region of integration)
6. Suppose that 7 adult male racoons were removed from the population at random and weighed in an attempt to estimate \( \mu \), the intrinsic mean weight of adult male racoons. The weights of the 7 racoons are displayed in the table below. It is assumed that the weight of adult racoons is normally distributed.

<table>
<thead>
<tr>
<th>Racoon Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (lbs)</td>
<td>12.1</td>
<td>4.3</td>
<td>6.4</td>
<td>12.7</td>
<td>15.3</td>
<td>5.6</td>
<td>16.2</td>
</tr>
</tbody>
</table>

(a) Compute a 95% confidence interval for \( \mu \) if it is assumed that the standard deviation of the population is \( \sigma = 3.5 \).

(b) Compute a 95% confidence interval for \( \mu \) if the standard deviation \( \sigma \) is assumed to be unknown.

(c) Compare the confidence intervals computed in part a) and b). Account for their differences by briefly explaining the difference between a \( t \) distribution and the standard normal distribution.
7. Nabisco claims that the average number of chocolate chips in a bag Chips Ahoy! chocolate chip cookies is 23. The Better Business Bureau believes that Nabisco is actually ripping us off, and that the true average number of chocolate chips in a bag is less than 23. To prove their claim, the number of chocolate chips in 50 bags of Chips Ahoy! were counted and the sample mean $\bar{X}$ was 21.3 chips. Assume that the standard deviation of chocolate chips in a bag of Chips Ahoy! is known to be 3.2.

(a) Construct appropriate null and alternative hypotheses for this situation. Make sure the null hypothesis is simple.

(b) Perform the test that you outlined above at level $\alpha = .05$ based on the sample data. Make note of any approximations used and validate them by referencing the appropriate theorems.
8. A coin is flipped 400 times, and 217 of those times the outcome is heads. Suppose $p =$ probability that the coin when flipped will come up heads.

(a) Using this data test at level $\alpha = .01$

- $H_0 : p = .5$
- versus
- $H_A : p > .5$

(b) Derive a 90% confidence interval for $p$ based on this sample.

(c) Briefly describe any approximations used above.
9. Suppose $X_1, \ldots, X_{25}$ is a sample from a $N(\mu_1, 6)$ distribution and $Y_1, \ldots, Y_{36}$ is a sample from a $N(\mu_2, 8)$ distribution. Assume the two sample are taken independently from each other. Computed from these samples, $\bar{X} = 13.3$ and $\bar{Y} = 12.4$.

(a) Test at level .1 using this data

$H_0 : \mu_1 = \mu_2$
versus
$H_A : \mu_1 \neq \mu_2$

(b) Compute the P-value of this test based on the sample data.
10. (Extra Credit 1) How many ways are there to arrange 7 books on 3 shelves if each shelf can support all 7 books. Suppose the order in which the books are placed on the shelves matters. (i.e. book 1 to the left of book 2 is on shelf 1 is different from book 2 to the left of book 1 on shelf 1).
11. (Extra Credit 2) Suppose $X_1, \ldots, X_n$ is a sample of size $n$ from a $N(\mu, \sigma^2)$ distribution where $\sigma$ is known. Suppose we test at level $\alpha$

$H_0 : \mu = \mu_0$

versus

$H_A : \mu \neq \mu_0$

by rejecting $H_0$ if

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$

What is the probability of a Type I Error? Also, derive an expression, showing all steps of your derivation, for $\beta(\mu')$ where

$$\beta(\mu') = P(\text{Type II Error when } \mu = \mu' \neq \mu_0)$$