Math/Stat 3850 – Final Exam Practice Questions

There are 15 questions, worth a total of 150 points.
You may use R, your calculator, and any written or internet resources on this test, although you are not allowed to ask someone else for help.

(10) 1. The Intel i5 and i7 computer processors are actually the same chip. Intel makes a wafer of 190 chips, then tests them for speed. About 60% of these chips test well and are sold as i7 chips, the rest are either discarded or sold as i5 chips.
   (a) What is the expected number of i7 chips on a wafer?
   (b) What is the probability that more than 120 chips on a wafer can be sold as i7’s?

(10) 2. In some role playing games, character abilities are determined by rolling four six-sided dice, removing the smallest value, and adding the other three. For example, a roll of 2, 3, 5, 5 gives an ability 13.
   (a) What is the probability of rolling an ability of 13 or less?
       Write your code here, and give an approximate answer to within 0.01
   (b) A character has six different abilities, and if none of them are above 13 the character is discarded. What is the probability the character will be discarded?

(10) 3. Let $X$ be a random variable with the PDF $f(x) = \begin{cases} c(1-x^2) & \text{for } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$
   (a) Find $c$
   (b) Sketch the cumulative distribution function (CDF) for $X$.

(10) 4. With $X$ as in question 3:
   (a) What is $P(X > 0)$?
   (b) What is $P(X \leq \frac{1}{2})$?
   (c) What is $E(X)$?

(10) 5. In the game of Scrabble, players draw tiles randomly from a bag. Tiles have different values, given by the probability distribution:

<table>
<thead>
<tr>
<th>Tile value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.02</td>
<td>0.68</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Compute the expected value of a Scrabble tile.

(10) 6. Exact or approximate answers are acceptable.
   (a) Suppose $X$ has an exponential distribution with $\lambda = 0.1$. What is the expected value of $X$? What is the variance of $X$?
   (b) If $X_1, \ldots, X_{25}$ are i.i.d with $X_i \sim \text{Exp}(0.1)$, what is the expected value of $\bar{X} = \frac{1}{25} \sum_i X_i$? What is the variance of $\bar{X}$?
7. Suppose your population has the exponential distribution with \( \lambda = 0.1 \), and you take a sample of size 25. You wish to test \( H_0 : \mu = 10 \) versus \( H_0 : \mu \neq 10 \) at the \( \alpha = 0.05 \) level.

(a) Approximate the type I error if you use a \( t \)-test.
(b) Approximate the type I error if you use a Wilcoxon test.
(c) What should the type I error be for these tests? Which test is performing closer to its designed behavior?

8. The dataset \textbf{GAGurine} from \texttt{library(MASS)} gives the concentration of the chemical GAG in the urine of children at different ages. The aim of the study was to produce a chart to help a pediatrician to assess if a child’s GAG concentration is typical.

(a) Is GAG concentration positively or negatively correlated with age?
(b) Would it be appropriate to describe the relation between age and GAG with a linear model? Why or why not?
(c) Fit a linear model to explain the logarithm of GAG from age. What is the equation of the line?
(d) Does the linear model accurately describe the relation between age and log(GAG)?

9. The \textbf{genotype} data set from \texttt{library(MASS)} gives the weight gain of baby mice raised by foster mothers. The foster mother genotype is given by the \texttt{Litter} variable.

Test the hypothesis that all litters have the same mean weight gain. Report the \( p \)-value with your conclusion.

10. The \textbf{immer} data set from \texttt{library(MASS)} gives barley yield from various fields in 1931 (\( Y_1 \)) and 1932 (\( Y_2 \)). You want to compare the yield between the two years.

(a) What are the mean yields for each year?
(b) State \( H_0 \) and define any parameters you need to use.
(c) Carry out an appropriate hypothesis test, and summarize the results.

11. Suppose a population is normally distributed with mean 105 and standard deviation 15. You sample 40 data points and wish to test \( H_0 : \mu = 100 \) versus \( H_0 : \mu \neq 100 \) at the \( \alpha = 0.05 \) level.

(a) Approximate the type II error if you use a \( t \)-test.
(b) Approximate the type II error if you use a Wilcoxon test.
(c) Which test is more powerful in this situation?

12. Suppose \( X_1, \ldots, X_{10} \) are ten independent random variables, each with the \( t \) distribution with 5 df. Let \( X \) be the maximum of these:

\[
X = \max(X_1, X_2, \ldots, X_9, X_{10})
\]

Estimate the mean and standard deviation of \( X \) to within 0.1. Discuss the shape of the distribution of \( X \).
The “Bechdel Test” asks whether a movie features at least two women or girls who talk to each other about something other than a man or boy. The data set bechdel.csv on our course website has data on 1776 films from 1970-2013.

For questions 13-15, the relevant variables are:

- **test**: either “PASS” or “FAIL” depending on if the film passed the Bechdel test
- **reason**: the reason why the film failed the test, or “ok” if it passed.
- **budget**: the budget (in US$2013) for the film

(10) 13. State and carry out a $t$-test that the mean budget is the same for movies that pass and movies that fail the Bechdel test. Report the $p$-value in your conclusions.

(10) 14. Use ANOVA to check for a relationship between budget and the reason a film failed the test. Make sure you omit the films with reason ‘ok’ from your analysis – they passed.

(10) 15. Describe the shape of the distribution of the budget variable. This may cast doubt on the conclusions of problems 13 and 14. What could you do about it?