Exam 3 Practice Questions

1. (9 pts, 3 pts each) Some of the statistical inference techniques that we have studied so far include:
   
   A. One-sample \( z \)-procedures for a proportion
   B. One-sample \( t \)-procedures for a mean
   C. Two-sample \( z \)-procedures for comparing proportions
   D. Two-sample \( t \)-procedures for comparing means
   E. Paired-sample \( t \)-procedures

   For each of the following questions, identify (by letter) the procedure that you would use to investigate that question. Also indicate (either in symbols or in words) the null and alternative hypotheses to be tested in each case.

   a) Some critics of the military have claimed that members of the U.S. Armed Forces have an average IQ less than 100. Suppose that you take a random sample of members of the U.S. Armed Forces and measure their IQs, in order to test this claim.

   b) Researchers investigated whether drivers tend to take longer to react to a stimulus if they are talking on a cell phone than if they are listening to an audiobook. A sample of 32 drivers participated in a simulated driving device under both conditions (talking on phone, listening to book), and researchers recorded how long it took to react to a stimulus under each condition.

   c) Do the Brooklyn Nets allow significantly fewer points on average during games in which Kevin Garnett plays as opposed to games in which Kevin Garnett does not play? To investigate this question, you record the number of points allowed by the Nets in every game played this year, along with whether or not Garnett played in the game.

2. I recently read an article in Science magazine that described an experiment concerning people’s willingness to be organ donors. All of the subjects in the experiment were told that to imagine that they have moved to a new state and have applied for a driver’s license, and they must make a decision about whether to become an organ donor. Some of the subjects were randomly assigned to be told that the default option is not to be a donor, and the rest of the subjects were told that the default option is to be a donor. All subjects were then given the choice of whether or not to become an organ donor. The researchers suspected that a higher proportion of people are willing to be donors when the default option is to be a donor.

   a) Identify the explanatory and response variables in this study.

   b) State (in symbols) the appropriate null and alternative hypotheses (in symbols) for testing the researchers’ suspicion.

   c) The article reported that 42\% of the subjects in the “default is not to be a donor” group decided to become a donor, compared to 82\% in the “default is to be a donor” group. What
further information would you need in order to conduct the test (i.e., to calculate a test statistic and p-value)?

3. I once collected data in class on how long (in seconds) it took for a chocolate chip to melt in your mouth and for a peanut butter chip to melt in your mouth. I took the differences in these times (chocolate minus peanut butter) for each person. The sorted data, and a dotplot, for the 31 differences appear below:

![Dotplot of chocolate minus peanut butter melting times](image)

```
-41 -36 -35 -33 -31 -28 -25 -25 -20 -20
-17 -17 -16 -14 -11 -7 -6 -5 -5 -4
-4 -2 1 3 6 15 17 21 30 36
67
```

a) Explain what the value -41 means in terms of the student who produced that value and his/her melting times.

b) Determine the five-number summary of these differences. (Show how you do these calculations.)

The mean of these 31 differences is -6.65 seconds, and the standard deviation is 23.61 seconds.

c) Conduct a test of whether the sample data provide strong evidence of a difference in melting times of chocolate and peanut butter chips on average. Report the hypotheses, test statistic, and p-value as accurately as you can.

d) Determine and interpret a 95% confidence interval based on the 31 differences.

e) Summarize your conclusion from this analysis.

f) Now suppose that you were to re-do this analysis after removing the outlier value of 67. Indicate how each of the following would change. Circle your answers. Do not bother to explain or perform any calculations.

<table>
<thead>
<tr>
<th>Calculation</th>
<th>Decrease</th>
<th>Increase</th>
<th>Remain the same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>Decrease (more negative)</td>
<td>Increase (less negative)</td>
<td>Remain the same</td>
</tr>
</tbody>
</table>
4. The following Minitab output comes from analyzing data collected by a naturalist named Bumpus on the lengths (in millimeters) of sparrows, some of which had survived a severe winter storm and some of which had perished. 

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>lengths (died)</td>
<td>24</td>
<td>162.00</td>
<td>2.41</td>
<td>0.49</td>
</tr>
<tr>
<td>lengths (survived)</td>
<td>35</td>
<td>159.06</td>
<td>2.81</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Difference = \( \mu \) (lengths (died)) - \( \mu \) (lengths (survived))

95% CI for difference: (1.57150, 4.31421)

a) Write out the formula (with numbers plugged in) that would have produced the confidence interval reported here. (Do not bother to work out the calculations, though; just show the formula with the numbers plugged in.)

b) Interpret the confidence interval given here. Be sure to address whether the sample data provide evidence that the population mean length of sparrows who survived is different from that of sparrows who perished. If there is evidence of a difference, also address which type (those that survived or those that perished) tends to be longer.

c) Is it reasonable to conclude from this study that the sparrow’s length caused it to die or survive? Explain briefly.

5. In April of 2006, the National Cancer Institute released a report about a study that compared two drugs (raloxifene and tamoxifen) intended to reduce a woman’s risk of developing breast cancer. Women who participated in the study were randomly assigned to take one of these drugs on a daily basis. It turned out that 167 of the 9745 women who took raloxifene developed breast cancer, compared to 163 of the 9726 women who took tamoxifen. (For simplicity, refer to the drugs as R and T.) Conduct a hypothesis test of whether these proportions differ significantly. Include the following components:

a) null and alternative hypotheses
b) test statistic
c) P-value
d) test decision at \( \alpha = .05 \) level
e) check of technical conditions
f) summary of conclusion in context

6. (10 pts) Students in an introductory statistics class were asked how many states they have visited. The following output pertains to the sample results:
| states | 50  | 13.36 | 1.03  | 7.27  | 2.00  | 7.00  | 12.00 | 20.00 | 31.00 |

a) (3 pts) Determine a 90% confidence interval for the population mean number of states visited among all students at this university.

b) (2 pts) Check and comment on whether the technical conditions of this confidence interval are satisfied.

c) (1 pts) For what proportion of students in the sample is the number of states visited within the interval from a)?

d) (2 pts) Should you expect your answer to c) to be close to 90%? Explain why or why not.

e) (2 pts) Based on your interval, what can you say about the $p$-value if you were to conduct a two-sided significance test of whether the population mean differs from 10? Explain briefly, without conducting a test or doing new calculations.