PROJECT 1 SOLUTIONS

Statistics 401: Fall 2016

This quiz is worth 29 points.

1. From the article *Monitoring the Health of Canine Heroes*:

   (a) The scientific method is:

      i. Observe some phenomenon
      ii. State a hypothesis explaining the phenomenon
      iii. Collect data
      iv. Test: Does the data support the hypothesis?
      v. Conclusion. If the test fails, go back to step 2.

   (b) The scientists in the article followed the scientific method by:

      i. Observe that "Many human responders who assisted at the World Trade Center have suffered from respiratory (pulmonary) problems."
      ii. Professor Otto "was expecting all sorts of pulmonary problems" in the search and rescue dogs who worked at Ground Zero compared to other search and rescue dogs.
      iii. Data was collected from 95 search-and-rescue dogs who worked at the World Trade Center and the Pentagon in 2001 and also from search-and-rescue 55 dogs who had not worked at Ground Zero: "Each year, veterinarians took blood samples and chest X-rays and collected information about the dogs health and behavior. Whenever dogs enrolled in the study pass away, researchers performed a full autopsy."
      iv. The data from the 9/11 dogs compared to other dogs did not support Professor Otto's hypothesis.
      v. The data failed to suggest that dogs that responded to the 9/11 disaster were more likely to be diagnosed with cancer or respiratory problems than search-and-rescue dogs that werent involved.

   (c) Search-and-rescue dogs at Ground Zero and the Pentagon after 9/11 are the individuals of interest.

   (d) The explanatory variable is whether, in 2001, the dog worked at the World Trade Center (Ground Zero), the Pentagon, or neither. The response variable is whether or not a dog "suffered from respiratory (pulmonary) problems."

   (e) Table 1 shows that both of the variables from 1d are categorical.

Table 1: Variables in the Search and Rescue Dog Study

Variable	Type	Properties
Location in 2001	Categorical	3 levels: Ground Zero, Pentagon, neither
Pulmonary problems	Categorical	2 levels: Yes, No

   (f) The "300 search-and-recue dogs (that) were deployed to the World Trade Center and the Pentagon" in 2001 is the finite population of interest.

(g) Although the article does not explicitly say, the following statement in the article suggests that the researchers employed a **voluntary response sample**: "For search-and-rescue dog handlers, though, helping the dogs was reason enough to join the 9/11 canine health study." That is, Professor Otto's team may have contacted ALL of the search-and-rescue dogs in the government's database and asked for volunteers to participte in the study. In this scenario, **selection bias** will be present if those who volunteered are systematically different from those who did not (e.g., dogs in certain geographic locations or or dogs with extroverted handlers). On the other hand, Professor Otto's team may have tried to implement a **stratified random sample** (with srata equal to 9/11 work or not) by randomly selecting animals from the **sampling frame** of search-and-rescue dogs in the government's database. Unfortunately, some handlers likely refused to participate in the study. In this case, the sample could suffer from **non-response bias** if those who participated in the study are systematically different from those dogs who were selected to be in the study but failed to respond. Either of the previous sampling plans could also suffer from **measurement bias** if different veterinarians examined the X-rays, performed physicals and did the behavioral assessments.

(h) The article reports on an observational study of search-and-rescue dogs because no treatment was applied by the researchers.

2. (a) (2 pts) Table 2 displays a list of the variables and variable types from the Jellyfish study.

Table 2: Variables in Jellyfish Study

Variable	Type	Properties
Case	Categorical	46 levels: $1, 2, 3, \ldots, 46$
Location	Categorical	2 levels: Dangar I., S. Bay
Breadth	Numerical	continuous
Length	Numerical	continuous

(b) (2 pts) The investigator wants to know whether or not the location of the jellyfish can be determined given its dimensions. Thus, the response variable is Location. The explanatory variables are Breadth and Length.

(c) (2 pts) The R commands to read in the data and then display the data are:

```
D=read.table("jellyfish.txt",header=TRUE)
attach(D)
D
```

3. (6 pts) The mean length of a jellyfish from Dangar Island is 12.34mm and the mean length of a jellyfish from Salamander Bay is 18.98mm. The difference between these is "large with respect to the standard deviation" of each (3mm and 1.94mm respectively), so we may conclude that, on the average, longer jellyfish are from Salamander Bay. We'll learn later in this course how "t-procedures" can be used to quantify "large with respect to the standard deviation". See the Appendix for the code and output from R.

Appendix A

```
# Problem 2c
> D=read.table("jellyfish.txt",header=TRUE)
> attach(D)
> D

# Problem 3
> tapply(Length,Location,mean)
      DI       SB
12.34091 18.97917
> tapply(Length,Location,sd)
      DI       SB
3.005496 1.936375
```