Evanston Township High School

Data Analysis and Statistics

Small Student Project

Instructor: Jiangtao Gou

February 10th, 2013

Confucius (551–479 BCE) taught,

I hear, I forget.
I read, I remember.
I do, I understand.

0.1 Organization

A Four-Step Process of Organizing a Statistical Problem by David S. Moore

- **STATE**: What is the practical question, in the context of the real-work setting?
- **PLAN**: What specific statistical operations does this problem call for?
- **SOLVE**: Make the graphs and carry out the calculations needed for this problem.
- **CONCLUDE**: Give your practical conclusion in the setting of the real-world problem.

In this course project, you may need to describe how you collect your data in the second step of Moore's four-step process.

0.2 My Expectation

The goal of my course is

> To enable students to discover that statistics can be an important tool in daily life.

You may work individually or work as a group. Please note that each group should contain at most three students. Please form your teams no later than late February.

Basically, first you need to collect your data, either from any public database or an experiment/survey designed by yourself. Second, you write a report to clearly present your results. You may follow the four-step process as a reference. Each group will hand in one report. Third, you will be invited to show your work in class.

0.3 Experience Sharing

Each group will be invited to present his/her/their result in 3 to 5 minutes on March 15th. I will hold the final class as a standard statistical conference or colloquium. Say, I will be the session leader, and I will briefly introduce each group's work and welcome the presenters to share their results with all of us. Let us call the final class "2013 Evanston Township High School Statistics Research Colloquium", 2013 ETHS-SRC for short.

Sample Course Project

Blood Types

by Jiangtao Gou

1 Question Statement

The ABO system is the most important blood-group system in human-blood transfusion.

Recently I read an article from `http://bloodcenter.stanford.edu/about_blood/blood_types.html`, where I learned that in the United States, blood type O and blood type A are much more popular than blood type B and blood type AB.

I am very curious about this result. I want to use my own data to persuade myself.

2 Statistical Plan

I designed a questionnaire to ask his/her blood type, and randomly gave to 15 classmates. No one refused to answer my question. In this report, I list my data anonymously in Table 1.

I summarized my data in Table 2 and Table 3.

Table 1: Students' Blood Types

Student	Gender	Blood Type
A	M	A
B	M	A
C	M	O
D	F	AB
E	F	A
F	F	O
G	F	B
H	M	A
I	M	A
J	M	B
K	M	O
L	M	A
M	F	B
N	F	O
O	M	O

Table 2: Blood Types

Blood Type	Student Count	Percentage
O	5	33.3
A	6	40.0
B	3	20.0
AB	1	6.7
Total	15	100

Table 3: Blood Type and Gender

Blood Type	M	F
O	3	2
A	5	1
B	1	2
AB	0	1

3 Problem Solution

We would like to know that whether blood type O and A are more popular than blood type B and AB in the United States.

Null Hypothesis
- H_0: Blood type O/A and blood type B/AB are equally popular in the United States. $p = 0.5$.

Alternative Hypothesis
- H_a: Blood type O/A are more popular than blood type B/AB in the United States. $p \neq 0.5$.

The estimated proportion of blood type O/A is

$$\hat{p} = \frac{5+6}{15} = 0.733. \tag{1}$$

The estimated standard deviation is

$$\hat{s} = \sqrt{\frac{0.733 \times 0.267}{15}} = 0.114. \tag{2}$$

The Z-statistic is
$$Z = \frac{0.733 - 0.5}{0.114} = 2.05. \tag{3}$$

Because $2.05 > 1.96$, we reject H_0 under 95% significant level.

4 Conclusion

We conclude that O/A are more popular than blood type B/AB in the United States by using a significant level 95% hypothesis testing procedure.

We are also curious about the relation between blood type and gender. However, we have not learned enough statistical knowledge to analyse this two-variable question. We will solve it when we get into college and take a complete introductory level statistics course.