Problem Set 2

Due Friday August 9 at the beginning of class

1. Uniform Colors. A study conducted by Hill and Barton (Nature, 2005) investigated whether Olympic athletes in certain uniform colors have an advantage over their competitors. They noticed that competitors in the combat sports of boxing, tae kwon do, Greco-Roman wrestling, and freestyle wrestling are randomly assigned red or blue uniforms. For each match in the 2004 Olympics, they recorded the uniform color of the winner.

The observational units in this study are the matches, and the variable is whether the match was won by someone wearing red or someone wearing blue—a categorical variable. Let's suppose that going into this study, the researchers wanted to see whether one color had an advantage over the other. In other words, competitors that wear one of the uniform colors will win a majority of the time.

Researchers Hill and Barton used data collected on the results of 457 matches and found that the competitor wearing red won 248 times, whereas the competitor wearing blue won 209 times.

We will carry out a calculation to assess whether or not the observed data provide evidence in support of the research conjecture, using the usual strategy: Form specific hypotheses (null and alternative), determine the statistic, calculate the could-have-been outcomes of the statistic under the null model, and assess the strength of evidence against the null model by estimating the p-value or the standardized statistic.

One sided vs. two-sided tests

One factor that influences strength of evidence is whether we conduct a two-sided ($\neq$) or a one- sided test ($<$ or $>$). Why would the researchers do a two-sided test and what are the implications?

(a) State the null and a one-sided alternative hypotheses *in words*.

(b) We will let p represent the probability that a competitor wearing a red uniform wins. Using this, restate the hypotheses using symbols.

(c) What is the statistic we will use? Calculate the observed value of the statistic in this study.

(d) Describe how you could use a coin to develop a null distribution to test our hypothesis.

(e) Using Matlab, calculate the one-sided p-value for the observed data and write a conclusion.

Also write down the mean and standard deviation from your null distribution when the *proportion of successes* (i.e. percent red wins) is used for the variable on the horizontal axis. You will need this later.

(f) Sketch the null distribution you calculate (by hand or with Matlab), with proportion of successes on the horizontal axis. Indicate the value on the horizontal axis that is used to calculate the p-value.

Now perform the same analysis for a two-sided alternative hypothesis.

(g) State the null and the two-sided alternative hypotheses *in words* and *in symbols.*

(h) Calculate the two-sided p-value for the observed data and write a conclusion. Describe how this differs from the one-sided p-value in (e).

(i) Sketch the null distribution you calculate (by hand or with Matlab), with proportion of successes on the horizontal axis. Indicate the values on the horizontal axis that are used to calculate the p-value.

(j) Explain how switching from a one-sided to a two-sided test influences the strength of evidence against the null.

Distance between statistic and null hypothesis parameter value (effect size)

A second factor that influences the strength of evidence against the null is the distance between the observed sample statistic and the value of the parameter specified under the null hypothesis. For this study the null value was 0.50 and the observed sample statistic was about 0.543 (or 54.3% of the competitors wearing red won their matches). Suppose a larger proportion of competitors wearing red won their matches. If fact, suppose 57% of the 457 matches were won by a competitor wearing red.

(k) Calculate the p-value for this new data, for both one-sided and two-sided alternative hypotheses.

(l) Are your p-values larger or smaller than your original one? Explain why this makes sense.

(m) Write a sentence explaining the relationship between the distance between the observed sample statistic and the value of the parameter specified under the null hypothesis to the strength of evidence against the null hypothesis.

Sample size

The third factor we will look at that influences strength of evidence against the null hypothesis is the sample size. As we said earlier, the data for this study came from four combat sports in the 2004 Olympics. One of those sports was boxing. The researchers found that out of the 272 boxing matches, 150 of them were won by competitors wearing red. This proportion of $150/272 \approx 0.551$ is similar to the overall proportion of times the competitor wearing red won. Let's see what the smaller sample size does to the strength of evidence. We will test the same hypotheses as above, but with just the boxing matches as our sample.

(n) Compare the null distribution you generate in this case to that generated in (e). In particular, how do the center (mean) and variability (standard deviation) compare?

(o) What are your new p-values, for one- and two-sided tests? Are they larger or smaller than your original p-value from (e) and (h)? Explain why this makes sense.

(p) Write a sentence explaining the relationship between sample size and the strength of evidence against the null hypothesis.

2. Sampling Distribution The mean and standard deviation of the lifetime of a battery in a portable computer are known to be 3.5 and 1.0 hours, respectively.

(a) Approximate the probability that the mean lifetime of 25 batteries exceeds 3.25 hours

(b) Approximate the probability that the mean lifetime of 100 batteries exceeds 3.25 hours

(c) Comment briefly on why the answers to parts (a) and (b) differ.

3. An executive is worried that she may be laid off by her employer before her stock options become vested and she will lose a significant portion of her compensation. Stock options are vested after 6 years, and she has been an employee for 4 years, so she has 2 years until vesting. She used LinkedIn to locate 6 former executives of the company and finds that the length of their employment was approximately normally distributed, with a mean of 5.25 years and sample standard deviation of 1.25 years.

(a) Test the hypothesis that the mean time for executive employment is actually 6 years or more versus less than 6 years. Use $\alpha = 0.05$. Start by stating null and alternate hypotheses.

(b) What is the p-value for the test in part (a)?

4.Lottery Some states that operate a lottery believe that restricting the use of lottery profits to supporting education makes the lottery more profitable. Other states permit general use of the lottery income. The profitability of the lottery for a group of states in each category is given below.

State Lottery Profits

For Education		For General Use	
State	% Profit	State	% Profit
New Mexico	24	Massachusetts	21
Idaho	25	Maine	22
Kentucky	28	Iowa	24
South Carolina	28	Colorado	27
Georgia	28	Indiana	27
Missouri	29	Dist. Columbia	28
Ohio	29	Connecticut	29
Tennessee	31	Pennsylvania	32
Florida	31	Maryland	32
California	35		
North Carolina	35		
New Jersey	35		

Source: *New York Times*, National Section, October 7, 2007, p. 14.

Test at the α=0.01 level whether the mean profit of states using the lottery for education is higher than that of states permitting general use. Assume that the variances of the two random variables are equal.

5. A stomach flu outbreak appears among Harvard students. Public health officials would like to determine if the outbreak was spread through one particular house's dining hall, or by casual contact among students throughout the College. To address this question, they studied samples of students: 55 were students who ate in the dining hall; among these, 24 got sick. Of 149 sampled students who did not eat in the dining hall, 54 got sick.

(a) Can we conclude with 95% confidence that the fraction (or percentage) of students who got sick is higher among students eating in the dining hall than among students eating elsewhere?

(b) What is the p-value for this test?

6. MATLAB problem.

The population proportion hypothesis tests we covered in class uses the normal approximation to the binomial distribution, bin(n,p) ~ norm($\mu=np$,$\sigma=\text{sqrt}(np(1-p))$), which is taken to be reasonably accurate when both $np>5$ and $n(1-p)>5$.

How large is the error when using this approximation for various values of n and p?

Produce a plot showing the largest difference between the binomial distribution bin(n,p) and the normal distribution norm(np, sqrt($np(1-p)$)) for each value of n from 1 to 50. Produce plots for three different values of p = 0.1, 0.2, and 0.5.

Useful code snippets:

```
x=0:n;
binos=binopdf(x,n,p);
norms=normpdf(x,n*p,sqrt(n*p*(1-p)));
```

Hand in:

- 3 plots: max. error vs. n (2 to 50) for p=0.1, 0.2, and 0.5; <u>indicate on each plot</u> (by hand is OK) the range of values of n where the condition {both $np>5$ and $n(1-p)>5$} applies
- Your MATLAB script
- A brief comment on the size of the largest error you can expect when the condition {both $np>5$ and $n(1-p)>5$} is met, and the size of the error for all values of n and p.