**A New Kind of Statistical Problem**

**Example** SupposeFred claims that he can toss a fair coin in such a way that it is more likely to produce a head than produce a tail. My response: prove it. How? I have Fred toss the coin 10 times and count the number of times it comes up heads. Suppose it comes up heads 10 times. Would that convince me of his power to influence the outcome? Suppose he produces 8 heads. How convincing is that?

**Issue:** Even if Fred has no ability to influence the outcome, it is possible that he produce 10 (or 8) heads by chance with a fair coin. Can we quantify the strength of the evidence for his claim?

Two conflicting hypotheses:

1. Fred cannot influence the outcome of a toss.
2. Fred can influence the outcome of a toss.

The default hypothesis (**Null Hypothesis = H0**)

Fred cannot influence the toss

What requires strong evidence ”proof” (**Alternative Hypothesis = Ha**)

Fred can influence the toss

Let p = probability that a toss produces a head if Fred tosses it.

H0 : p = .5

Ha : p > .5

If Fred produces 8 heads out of 10 tosses, how strong is the evidence for Ha : p > .5? What is the probability that, if the null hypothesis is true, he would produce at least 8 heads? Let X = number of heads produced in 10 tosses. If p = .5, what is P(X ≥ 8)?

**Using simulation**

> sim<-do(1000)\*rflip(10,prob=.5)

> tally(~heads, data=sim)

1 2 3 4 5 6 7 8 9 10

12 40 131 191 244 210 123 43 5 1

So, 49/1000 sets of 10 tosses produce at least 8 heads. So, P(X≥8) is approximately 5%.

**p-value**

The **p-value of data** is the probability of getting data as strongly supporting the alternative hypothesis by chance if the null hypothesis is true. In our example, the p-value is 5%. The smaller the p-value, the stronger the data supports the alternative hypothesis. Is 5% convincingly strong? It’s a bit of a judgment call. A common rule is that if the p-value is ≤ 5%, reject the null hypothesis in favor of the alternative hypothesis. With that rule, Fred’s performance is convincing.

**Types of Errors**

**Type I Error**: Reject the null hypothesis when the null hypothesis is true

**Type II Error:** Fail to reject the null hypothesis when the alternative hypothesis is true.

You can think of a hypothesis test as a “criminal” trial. The null hypothesis (which requires strong evidence to reject) is that the defendant is not guilty; the alternative hypothesis is that the defendant is guilty.

H0: Defendant is innocent

Ha: Defendant is guilty

Type I Error:

Type II Error:

It takes evidence beyond a reasonable doubt to “prove” guilt. Convicting an innocent person is a greater error than letting a guilty person go free in the US criminal court system. Equivalently, a Type I error is a worse error than a Type II error. The testing (trial) procedure is designed to control the rate of Type I errors.

The procedure illustrated above is called **Hypothesis Testing**.

**General format for Population Propotions**

The value of a population proportion p is at issue.

H0 is an equality claim about the parameter

Ha is a strict inequality claim about the parameter

H0 : p = p0

Ha :

The null hypothesis is rejected in favor of the alternative hypothesis only if the data provides strong evidence against the null hypothesis in favor of the alternative hypothesis. The **p-value** of the data is the probability, assuming that H0 is true, of getting data that supports Ha as strongly as our data does by chance. The smaller the p-value, the stronger the evidence against H0 and in favor of Ha. What p-values should lead to rejection of H0 is a judgment call. In the past, one standard was: If the p-value ≤ 5%, reject H0. Now, studies often simply report the p-value. How small the p-value must be often depends on the cost of making a Type I and Type II error.

**Example**

H0 : p = p0

Ha : p > p0

**Data**: A sample of size n

**Sample statistic**:  = sample proportion

Values of  larger than p0 count as evidence for the alternative hypothesis. The p-value of the data is the probability of getting a sample proportion at least as large as the proportion of our data by chance if the null hypothesis is true. To compute this probability, we need the sampling distribution for the sample proportions for all random samples of size n from a population with proportion p0.

**An approximate procedure using the standard normal distribution (gives good results if sample size is ≥ 30.)**

Test Statistic z =

If n ≥ 30, the test statistic above has (approximately) a standard normal distribution.

**Example**: We test Fred’s ability to toss a head more than 50% of the time by having him toss the coin 50 times. Suppose he produces 30 heads on the 50 tosses. What is the p-value?

**Step 1**: Write down the hypotheses

**Step 2**: Calculate the value of the test statistic

**Step 3:** Find the p-value.

**Step 4:** Make the decision

Suppose he tosses 40 heads. What is the p-value?

**The Binomial Distribution (Revisited)**

Consider a population of individuals and a random variable X defined on that population that only produces the values 0 and 1. A random variable of this type is called a **Bernoulli random variable**.The population proportion p associated with X is the proportion of individuals in the population for which X = 1. Individuals with value 1 are called **successes** and those with value 0 are called **failures**. For a random sample of size n, let the random variable Y be the number of individuals in that sample for which X = 1.

The random variable Y has a **binomial distribution** with parameters p (success probability) and n (sample size):

Y is binomial(n,p).

The possible values for Y are 0,1,2,…,n. The probability function for Y is

p(y) = 

**Example:** A coin has a 20% chance of producing a head. If it is tossed 6 times, what is the probability of getting 4 heads? At least 4 heads?

**Example:** We test Fred’s ability to toss a head more than 50% of the time by having him toss the coin 6 times. Suppose he produces 5 heads. What is the p-value?

**Example**: Company A provides parts for a machine produced by Company B. Company A claims that, on average, less than 5% of these parts are defective. Company B decides to test Company A’s claim by taking a sample of 50 parts and determining the percent of these parts that are defective. Unless the sample provides strong evidence that the defect rate is less than 5%, Company B will reject the shipment.

1. What is the parameter? What are the null and alternative hypotheses?
2. If 2 of the parts are defective, should Company B accept Company A’s claim? Find the p-value and decide to accept or reject.

**R has the binomial distribution built in.**

dbinom(x,n,p) = P(Y = x), where x is an appropriate integer

pbinom(x,n,p) = cumulative distribution function. **P(Y <= x)**

qbinom(x,n,p) = quantile function

rbinom(k,n,p) produces a sample of size k

**Return to the example** at the top of the previous page and let R do the work.

**Prob of exactly 4 heads**

> dbinom(4,6,.2)

[1] 0.01536

**Prob of at least 4 heads**

> dbinom(4,6,.2)+dbinom(5,6,.2)+dbinom(6,6,.2)

[1] 0.01696

> 1-pbinom(3,6,.2)

[1] 0.01696

**R also has the hypothesis test for population proportions built in.**

binom.test(x , n , p , alternative = “two.sided” or “less” or “greater”),

where x is the number of successes in the sample, n is the sample size, p is the H0 value of the population proportion. **Note that R does the calculation of p-value exactly, not using the normal approximation.**

**Example:** a. A random sample of 40 Calvin students contained 23 females and 17 males. How strongly does this evidence support the hypothesis that more than half of Calvin students are female?

Parameter: p = proportion of all Calvin students that are female

H0 : p = .5

Ha : p > .5

* binom.test(23,40,.5,alternative="greater")

data: 23 out of 40

number of successes = 23, number of trials = 40, **p-value = 0.2148**

alternative hypothesis: true probability of success is greater than 0.5

95 percent confidence interval:

0.4331391 1.0000000

sample estimates:

probability of success

0.575

**The p-value of 21% is too large to conclude the more than half of the entire student body are females.**

b. Suppose the sample size was 400, of which 230 were female

* binom.test(230,400,.5,alternative="greater")

data: 230 out of 400

number of successes = 230, number of trials = 400, **p-value = 0.001565**

alternative hypothesis: true probability of success is greater than 0.5

95 percent confidence interval:

0.5327784 1.0000000

sample estimates:

probability of success

* 0.575

**The p-value of 0.15% is really small, so the evidence is convincing more than half of the entire student body are females. This illustrates that the value of by itself doesn’t determine the p-value (strenght of the evidence). The sample size plays an important roll.**

**Exercises 12**

1. A poll of 600 adults in Michigan found that 318 of the 600 would approve a ballot measure to legalize and tax marijuana.

1. **Without using R**: Compute the 95% confidence interval. What is the margin of error?
2. **Using R**: Is this data strong evidence that at least 50% of Michigan adults would approve such a ballot measure. State the hypotheses, compute the p-value, and state your conclusion using a complete sentence.

2. Without using R (calculator is allowed):

1. A fair coin is tossed 12 times. What is the probability of getting 6 heads?
2. A jar contains 5 red and 10 blue chips. Six chips are drawn **with replacement**. What is the probability of getting 4 red chips?

3. Using R: A biased coin has a probability = .4 of producing a head. It is tossed 20 times. Let Y be the number of heads that result.

1. What is P(Y = 7)?
2. What is P(Y ≥ 7)?
3. What is P(5 ≤ Y ≤ 15)?
4. Company B receives a large shipment of parts from Company A. Company B accepts the fact that any large shipment of parts will contain some defective parts. It is willing to accept a shipment that contains at most 10% defective parts. It is costly for Company B to mistakenly reject a shipment that actually meets its acceptable defective rate. So, it requires strong evidence that the defect rate exceeds 10% before it decides to reject a shipment. Suppose Company B examines 20 randomly selected parts from the shipment and finds that 3 of them are defective. How strongly does this data support the claim that the defect rate for the entire shipment is greater than 10%?
5. Set up the null and alternative hypotheses in terms of p, the defect rate for the complete shipment.
6. Find the p-value.
7. Decide whether the company should reject the shipment based on the data.
8. Repeat Exercise (4) assuming Company B samples 200 parts and finds that 30 are defective.
9. The p-value depends not only on the sample proportion, but also on the sample size. Suppose the hypotheses are

H0 : p = .3

Ha : p < .3

and  = .25. For which of the sample sizes n = 20, 100, 200, 500 would this value of  provide sufficiently strong evidence for p < .3 to lead to reject of the null hypothesis? Give the p-values in each case along with you answer to the question.