

**Assessing Left-Handedness with Exact and  
Approximate Confidence Intervals Appropriate  
for Samples of Different Sizes**

**An Honors Thesis By Virginia Gardner**

**Bemidji State University**

**2003**

**Honors Thesis Advisor: Dr. Derek Webb, Department of Mathematics and  
Computer Science, Bemidji State University**

## **Abstract**

The proportion of persons that are left-handed in the general population will be estimated through the use of basic sampling techniques and subsequent analysis. A brief introduction to sampling will be given and different sampling techniques will be discussed. A sampling design was created and two samples were taken: one small sample ( $n=10$ ) and one large sample ( $n=30$ ). The data was then compiled and analyzed. Minitab, a major statistical package was used. Confidence intervals for the data obtained were used to estimate the proportion of left-handed individuals in the general population. Confidence intervals were created by exact and approximate methods and the appropriateness of each method will be discussed. A comparison of weaknesses, strengths, and appropriateness of each method will follow.

## **1. Introduction**

The intention of this paper is to demonstrate the importance of appropriate samples and subsequent analysis in estimating a proportion. Surveys, questionnaires, etc. are all used in statistical research to obtain a representative sample of the population of interest without having to consider the entire population. Conclusions drawn can then be extended to represent the entire population with fairly high accuracy using confidence intervals. This will be shown by an example of the proportion of left-handed people in society. For illustration, two samples were obtained: a small and large one. Different methods of analysis were then used to estimate the proportion of left-handed people from this data. A comparison of the results will follow with an explanation of their appropriateness and accuracy.

## **2. Left-handedness in Society**

### *2.1 Historical Context*

Left-handed people are one of the few minorities that still exist today with no real sense of common identity. Discrimination is experienced differently for them than for most other minority groups. Left-handed people are able to navigate the daily routines of life as any right-handed person would. However, their discrimination lies in that they are forced to adapt to a predominately right-handed world. Historically, society has favored right-handers, through the design of simple devices like can-openers and pencil sharpeners to complex items such as computers. Even writing the English language from left to right is much more convenient for right-handers. It is important to note that there are exceptions; not all cultures use devices designed for the convenience of right-handers. For example, the Chinese writing system is vertical, making it suitable for the use of either hand. Today, society has become more aware of these inconveniences and devices have been specially designed for left-handed people.

The impact of this discrimination varies greatly from person to person. Usually, it amounts to nothing more than the awkwardness of using everyday tools. However, in some cases, severe emotional problems and serious learning difficulties may emerge. Individual responses vary from stuttering, headaches, and behavior disorders to confusing directions (Silverstein and Silverstein, 1977). Children who are forced to switch their dominant hand may also be more susceptible to dyslexia and have a greater chance of developing learning problems in school. Of course all of this depends on the child at



hand, their degree of “left-handedness,” and their reaction to this change. For most children, there are few or no significant negative responses to switching hand preference.

When dealing with any aspect of hand preference, the question of why most people are right-handed immediately arises. Even today there are no concrete answers to this question. There are, however, definite advantages to favoring one hand over the other. Everyday tasks become automatic, and a person does not have to think about which hand they are going to use for a given task. This saves a lot of time and effort, especially in life-threatening situations. The tendency to favor a certain hand probably arose to reap these benefits. Despite this, the question still remains; why are most people are right-handed?

There is still much confusion over why right-handedness is dominant and when this preference emerged. Many theories have been offered but little evidence supports most of these speculations. Early evidence from the Stone Age is incomplete and lacking. The little evidence researchers possess has led them to believe that these people lived mainly in an ambidextrous society.

Some pictorial evidence has been discovered in caves. This gives researchers a better understanding of hand preference during ancient times. Right-handed people tend to draw animal pictures facing to the left, while left-handers usually draw animals facing to the right. For example, cave drawings by the San people were found in South Africa. About 60% of these drawings are facing left. Most of these people were probably either left-handed, or the artists of the time were mostly left-handed.

Right-handed dominance emerges during the Bronze Age. Tools and weapons appear to be made for the use of the right hand. Again, there are many theories but no

definite answers to the question why. Evidently, prehistoric people sought conformity. Social pressure to fit in may have influenced handedness. People who were different than the norm were often feared and scorned. For whatever reason, the right hand seems to be favored from this period forward. Hand preference tends to be passed down from generation to generation. Children were taught to use tools geared for the convenience of their parents, who were predominantly right-handed. This begins the snowball effect of the right-hand dominance. This may be one of the many factors influencing the dominance of one hand over the other. Records show that the Egyptians, Greeks, Romans, and most Indian tribes were predominately right-handed although exceptions do exist.

## *2.2 Handedness in Today's Society*

Although the extent of the relationship between genetics and hand preference, if any, is not fully understood, it is clear that handedness is not completely fixed at birth. This has increasingly been a topic of great discussion. Children often go through phases, switching back and forth between hands or using different hands for different tasks. Parents often have mistaken this as a permanent choice of hand preference. Handedness is usually not determined until the child reaches about the age of five (Silverstein and Silverstein, 1977).

## **3. Goals of the Study**

The goal of the research conducted and the purpose of this paper is to examine different methods of analysis for samples of different sizes and the appropriateness of the

different methods. This research is not intended to produce exact figures for the actual proportion of left-handers in society, but rather focus on the methodology and mathematics behind it. Therefore, the results of the actual analysis are not intended for further use in other studies and no concrete conclusions can be made about the proportion of left-handers in society.

#### **4. Sampling Design**

##### *4.1 Background*

Two samples were taken, a large one consisting of 30 people and a small one consisting of 10 people. To avoid complications with Bemidji State University protocol, only people over the age of 18 were allowed to participate in this study. On February 16, 2003, a total of 40 people at the Paul Bunyan Mall in Bemidji, Minnesota agreed to participate in this survey. The Bemidji Mall was chosen because of its great influx of people. As people entered the mall, every tenth person was asked their hand preference. Because it has not yet been determined whether genetics influence handedness (Silverstein and Silverstein, 1977), this method was chosen in order to avoid asking members of the same family. The definition of left-handedness used in this survey was the hand used for writing. This method created a sample as random as possible given the circumstances and time restrictions. The data is presented in tables I and II in the Appendix.



#### *4.2 Potential Problems with Sampling Design*

Assessing the proportion of the adult population that is left-handed is not a straightforward problem for a variety of reasons. The first difficulty concerns the actual definition of left-handedness. What exactly determines whether or not a person is left-handed? There is no definite answer to this question. The percentage of left-handers varies from one to fifty percent, depending on the criteria used to determine handedness (Fincher, 1993). Interestingly, if one was to consider handedness to be all or nothing, i.e. all tasks are performed with only one hand; the percentage of true right-handers would drop significantly to about seven and a half percent. Handedness has been defined as anything from which hand a person uses to learn new tasks to preference for completing intricate projects. In today's society the use of both hands has become very popular and widely accepted. Most people favor a certain hand for certain tasks such as writing and eating while the other hand is specifically used for other tasks such as playing sports. Despite this vagueness, most people identify themselves as left-handed or right-handed. The hand used for writing appears to be the deciding factor (Fincher, 1993). Most researchers agree that in general about 10-11% of the population is left-handed. Only one of the forty people sampled claimed to be ambidextrous. Since middle ground for the analysis used in this paper is non-existent (i.e. a person is either right or left-handed), this person's response was dismissed and a new data point was obtained.

#### *4.3 Changes in Society's Influence*

Due to BSU protocol, I only surveyed people who were over 18. A significant number of the people surveyed appeared to be middle-aged or older. This may bias the

survey because historically people of these age groups that were left-handed were discouraged from being so. In fact, children were often punished and forced to become right-handed. One gentleman in particular stated that he was left-handed as a child but was converted to being right-handed to fit the norm of society. This brings up the interesting question of whether he was actually right or left-handed and whether preference is determined more by genetics or society. Today, left-handedness is widely accepted. Although it is not always encouraged, it is at least, for the most part, not discouraged.

#### *4.4 Other Related Factors*

Other factors that may influence the outcome of the analysis of the survey data include the area in which the survey was taken, and the time it was conducted. Since the entire sample was taken at the Bemidji Mall, it may not be a representative sample of the society as a whole. The survey was conducted from about 2:00 to 4:00 p.m. on Sunday, February 16, 2003. During this time period certain cohorts of people may frequent the mall in greater numbers than others. It has also not yet been determined how much, if any, gender, age, and cultural/genetic background, or geography influence a person's hand preference. Therefore, there may be hidden biases and related factors that are unaccounted for in the analysis. For example, in this survey, I noticed that many of those who participated were from an older generation which probably biased the data.



## 5. Estimating a Binomial Proportion

### 5.1 Confidence Intervals

Confidence intervals are used to estimate an unknown parameter in statistics. For example, when estimating the unknown proportion of left-handed people in the general population one typically starts with a point estimate. The point estimate might be 11%. This estimate is in all likelihood incorrect. Therefore, statisticians construct a confidence interval such as (9%, 13%). Confidence intervals give a range of probable values for the parameter being estimated.

Confidence intervals are always given along with a level of confidence. This is expressed as a  $100(1 - \alpha) \%$  confidence level with  $0 < \alpha < 1$ . In the previous example, (9%, 13%) might be a 95% confidence interval. This means  $\alpha=0.05$ . A 95% confidence interval can be interpreted as an interval that captures the true but unknown value of the parameter of interest 95% of the time. In other words, if 100 different samples were obtained and 100 different 95% confidence intervals were created, each estimating the same unknown parameter, 95 of the 100 would have captured the parameter within their bounds and 5 would have not captured the parameter. Therefore, 95 would be “correct” and 5 would be “incorrect.” The problem is that we never know if we have calculated a “correct” interval, we only know the level of confidence that our interval is correct.

Many methods exist for creating confidence intervals, each with their own advantages, disadvantages, and computational complexity. The survey data will be analyzed by computing two different types of confidence intervals for a binomial proportion. A comparison of these methods and the author’s recommendations will

follow.

### 5.2 Choice of Sample Size

For this paper, a sample of  $n=10$  was obtained to illustrate confidence interval methodology using a small sample size. Samples of sizes in the range of 10 are very common in many scientific fields. The Clopper Pearson method will be used to illustrate the small sample. A sample of size  $n=30$  was obtained to illustrate confidence interval methodology using a large sample size because textbooks commonly recommend a sample size of at least 30 when using the normal approximation method to construct confidence intervals. The normal approximation method will be illustrated in this paper.

### 5.3 Approximate and Exact Methods

All confidence intervals can be classified into two categories: approximate and exact confidence intervals. The accuracy of approximate confidence intervals is dependent on the true, but unknown value of  $p$ , the population proportion. For a given  $100(1 - \alpha) \%$  confidence level with  $0 < \alpha < 1$ , an approximate confidence interval method will necessarily result in probabilities below the desired level of  $\alpha$  for one or more values of the unknown parameter  $p$ . More precisely,  $\inf_{p \in \Omega} P(L_y \leq p \leq U_y) < 1 - \alpha$ , where  $L_y$  stands for the lower endpoint of the confidence interval,  $U_y$  stands for the upper endpoint of the confidence interval, and  $\Omega$  is the parameter space. Therefore, a practitioner using an approximate method cannot be certain that they have created a confidence interval with the level of confidence that they desire. This uncertainty often makes it difficult to make any concrete conclusions.

Exact methods ensure that a confidence interval is always at least at the desired level of  $\alpha$ . More precisely, these confidence intervals have the property that

$\inf_{p \in \Omega} P(L_y \leq p \leq U_y) \geq 1 - \alpha$ . Here the practitioner is assured to have at least  $100(1 - \alpha) \%$  confidence.

Even though exact methods seem to have a higher confidence, there are some disadvantages. One drawback of exact methods lies in that they may at times be too conservative. For example, if a 95% confidence interval is desired, for a given value of  $p$ , an exact method may yield a 99% confidence interval. Then, the interval will be unnecessarily conservative. As confidence levels increase, the width of the interval also increases. This implies that an exact method that produces confidence levels greater than or equal to  $100(1 - \alpha) \%$  but as close to  $100(1 - \alpha) \%$  as possible are the most desirable.

#### 5.4 Normal Approximation Method

Let  $p$  be the unknown proportion of people that are left-handed. Let  $Y$  be the random variable, which in our case, is the number of people that are left-handed in a sample of size  $n$ . Using the method of maximum likelihood estimation (Hogg and Tanis, 2001) the maximum likelihood estimator (MLE) of  $p$  is  $\hat{p} = Y / n = \text{number of successes} / \text{number of trials}$ . A binomial distribution has mean  $\mu = np$  and  $\sigma^2 = np(1-p)$ . To normalize this expression take the random variable  $Y$  minus the mean  $\mu$  divided by the standard deviation  $\sigma$  to obtain the expression  $\frac{Y - np}{\sqrt{np(1-p)}}$ . From the Central Limit Theorem, we

know that  $\frac{Y - np}{\sqrt{np(1-p)}}$  has an approximate  $N(0,1)$  distribution. Dividing the top and



bottom by  $n$ , we have  $\frac{\frac{Y}{n} - p}{\sqrt{p(1-p)/n}}$ . Note that  $\sqrt{\frac{p(1-p)}{n}}$  is the standard error of the

estimator. Because we know the approximate distribution of  $\frac{\frac{Y}{n} - p}{\sqrt{p(1-p)/n}}$ , we can

create a probability statement dependent on a given value of  $\alpha$ . Therefore, we have the

following equation:  $P(-z_{\alpha/2} \leq \frac{\frac{Y}{n} - p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}) \approx 1 - \alpha$ , where  $z_{\alpha/2}$  is the  $z$  value from

a standard normal distribution resulting in a tail probability of  $\alpha/2$ . Since we are

interested in the confidence interval for  $p$ , we will now solve this equation for  $p$  to create

an approximate  $100(1-\alpha)\%$  confidence interval. Solving for  $p$  results in the following

endpoints for our  $100(1-\alpha)\%$  confidence interval:  $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$ . Since the value

of  $p$  is unknown, in order to calculate the endpoints of the confidence interval we

substitute the MLE  $\hat{p}$  in for  $p$  to arrive at the final  $100(1-\alpha)\%$  confidence interval

formula for the normal approximation:  $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ . All confidence intervals have

the general form: (point estimator)  $\pm z_{\alpha/2} \times$  (standard error of the estimator) (Mendenhall, et al, 1999).

The MLE is a function of the unknown parameter  $p$  that locates the value of  $p$  most likely to have created the sample values. Using the MLE to estimate  $p$ , we begin with a random sample of  $n$  independent observations:  $X_1, X_2, X_3, \dots, X_n$ ;  $X_i \perp\!\!\!\perp X_j, \forall i \neq j$ .

Each  $X_i$  has an approximate Bernoulli distribution, meaning every trial has one of two

possible outcomes, success or failure. Let  $Y = \sum X_i$  with each  $X_i \sim \text{Bernoulli}(p)$ . Then

$Y_i \sim \text{Binomial}(n, p)$ . Consider the joint probability density function (pdf),

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{i=1}^n \binom{n}{y_i} p^{y_i} (1-p)^{n-y_i} = \prod_{i=1}^n \binom{n}{y_i} [p^{\sum y_i} (1-p)^{n-\sum y_i}] \quad \text{The}$$

Likelihood function created is  $L(p) = \prod_{i=1}^n \binom{n}{y_i} [p^{\sum y_i} (1-p)^{n-\sum y_i}]$ . We want to maximize

the function  $L(p)$  by taking the partial derivative with respect to  $p$ . Using calculus we will find the critical points and determine whether they are maximums or minimums.

Since it is often easier to take derivatives of log functions, consider

$$\ln L(p) = \ln \left[ \prod_{i=1}^n \binom{n}{y_i} (p^{\sum y_i} (1-p)^{n-\sum y_i}) \right] = \ln \prod_{i=1}^n \binom{n}{y_i} + \sum (y_i) \ln p + (n - \sum (y_i)) \ln(1-p).$$

Since the natural log is an increasing function, it will have the same maximum as  $L(p)$ .

Therefore, we need only to find the maximum of this function. The partial derivative of

$$\ln L(p) \text{ with respect to } p \text{ is: } \frac{d}{dp} \ln L(p) = 0 + \frac{\sum y_i}{p} + \frac{(n - \sum y_i)}{1-p}. \quad \text{To find the critical}$$

points, set the partial derivative equal to zero:

$$(1-p) \sum y_i - p(n - \sum y_i) = 0 \Rightarrow \sum y_i - p \sum y_i - pn + p \sum y_i = 0 \Rightarrow \hat{p} = \frac{\sum y_i}{n} = \bar{y}.$$

Now, we need to show that  $\hat{p}$  is a maximum. We will prove this by showing that the

second partial derivative is negative. Rewrite the first derivative as

$$\frac{d}{dp} \ln L(p) = \frac{\sum y_i}{p} + \frac{n - \sum y_i}{p-1}. \quad \text{Then } \frac{d^2}{dp^2} \ln L(p) = \frac{-\sum y_i}{p^2} - \frac{n - \sum y_i}{(p-1)^2}. \quad \text{Since we found}$$

the point estimator of  $p$  to be  $\bar{y}$ , we can rewrite this expression as

$$\frac{-\sum y_i}{\bar{y}^2} - \frac{n - \sum y_i}{(\bar{y} - 1)^2} = \frac{-(\bar{y} - 1)^2 \sum y_i - \bar{y}^2 (n - \sum y_i)}{\bar{y}^2 (\bar{y} - 1)^2} < 0. \text{ The denominator of this}$$

expression is always positive as it consists of multiplying two squared numbers. Since  $0 < \sum y_i < n$ , the numerator is negative as the first term is negative and the second term is either negative or zero. Therefore, the second partial derivative is negative and  $\hat{p}$  is a maximum.

The normal approximation method is the most commonly used technique for computing confidence intervals when estimating proportions, as it is fairly straightforward and easy to use. Approximate methods will produce decent results when  $np$  and  $n(1 - p) > 5$  (Mendenhall, et al, 1999). This means that the normal approximation method will work well when the sample size is sufficiently large and/or the probability is not too small or too large. A problem that immediately arises for my sample is that  $p$  is unknown. A search of the literature reveals that the best approximation available of the true proportion of left-handers in the general population is around 11 percent. For example, in the sample size of  $n=10$  case,  $np = (10)(0.11) = 1.1$  and  $n(1 - p) = 10(0.89) = 8.9$ . One of these is unacceptable. When  $n=30$ ,  $np = (30)(0.11) = 3.3$  and  $n(1 - p) = 30(0.89) = 26.07$ . One is still unacceptable, although greatly improved from the  $n=10$  case.

#### 5.4.1 Assumptions

Confidence intervals for the parameter  $p$ , using the two data sets obtained ( $n = 10$  and  $n = 30$ ) will be calculated for a value of  $\alpha = 0.05$ . This will result in a 95 % confidence level. The value of  $\alpha = 0.05$  was chosen because it is the most common value



used in science. The following additional assumptions will be made: the sample is from a binomial distribution and the observations are all independent of each other. Since we are working with a binomial random variable there are only two outcomes, success or failure. Since we are interested in left-handedness, we will arbitrarily define success as being left-handed and failure as being right-handed.

#### 5.4.2 Example Using Large Sample of $n=30$

Most textbooks state that the normal approximation method will yield desirable results with a sample of size 30 or greater. This method yields results closer to the desired  $100(1 - \alpha) \%$  confidence level the larger the sample size because of the central limit theorem. The central limit theorem states that as  $n$  increases the more normally distributed  $\frac{Y - np}{np(1 - p)}$  becomes and therefore, the closer the approximation method gets to the desired level of confidence.

Thirty different people were asked their hand preference. As mentioned earlier, one person claimed to be ambidextrous. That piece of data was discarded and replaced by a new one. The random variable  $Y$  is defined as the number of success, or left-handed people observed. The results of the survey are,  $n = 30$ ,  $y = 4$ , and the point estimator for  $p$  is  $\hat{p} = 4/30 \approx 0.133$ . Notice that  $n\hat{p} = (30)(4/30) = 4$  and  $n(1 - \hat{p}) = (30)(1 - 4/30) = 26$ , still unsatisfactory under the requirements that  $np$  and  $n(1-p)$  be greater than 5 as was stated earlier.

The confidence level of  $\alpha = .05$  results in  $z_{\alpha/2} = 1.96$ . The 95% confidence

interval formula is  $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ . Putting the respective values into this equation

results in  $4/30 \pm 1.96 \sqrt{\frac{(4/30)(1-4/30)}{30}}$ . Simplifying the formula yields the 95%

confidence interval [0.0117, 0.2550]. We can state that we are 95% confident that our confidence interval of [0.0117, 0.2550] captures the true but unknown proportion of left-handers. This method produced an approximate 95% confidence interval of the true but unknown parameter  $p$ .

#### 5.4.3 Example Using Small Sample of $n=10$

When working with small sample sizes, the normal approximation method oftentimes yields undesirable results. This will be demonstrated using the  $n = 10$  sample. One of the ten people who were asked their hand preference claimed to be left handed. Therefore,  $n = 10$ ,  $y = 1$ , and the point estimator for  $p$  is  $\hat{p} = 1/10 = 0.10$ . As before, for  $\alpha = 0.05$ ,  $z_{\alpha/2} = 1.96$ . The resultant approximate 95% confidence interval for  $n = 10$  is [-0.0859, 0.2859]. Notice that the lower bound for the confidence interval is less than zero which is unacceptable because a proportion is necessarily between 0 and 1. It is interesting to note that software programs, such as Minitab, will just round -0.0859 to 0 to avoid this problem. This is somewhat misleading as a practitioner may be unaware of this error resulting from the normal approximation method. The confidence level is also very questionable because the central limit theorem will not produce a good approximation for  $n < 30$  (Hogg and Tanis, 2001).

#### *5.4.4 Conclusions Concerning the Normal Approximation Method*

Confidence intervals with a confidence of 95% were created for  $n=10$  and  $n=30$ . The normal approximation method produced erroneous results for the  $n=10$  sample size. The confidence interval for the  $n=30$  sample contained endpoints between 0 and 1, but the level of confidence may not be 95%. For small samples more advanced techniques should be used and one is illustrated in the next section. For samples of 30 or more, according to the normal approximation method, should be used cautiously as the confidence level is only approximate.

#### *5.5 Clopper Pearson Method*

There are many exact and approximate alternatives to the normal approximation method that will yield better results for small sample sizes. This paper will focus on the oldest and most commonly used exact method, the Clopper Pearson Method (Clopper and Pearson, 1934).

##### *5.5.1 Example Using Large Sample of $n=30$ and Small Sample of $n=10$*

For a sample size of  $n=30$  and  $y=4$ , the Clopper Pearson method yields an exact 95% confidence interval of:  $[0.035, 0.320]$  and for a sample size of  $n=10$  and  $y=4$ , the Clopper Pearson method yields an exact 95% confidence interval of:  $[0.0025, 0.445]$ . The interval calculations were performed on Minitab. When Clopper and Pearson developed this method there were no computers or calculators. The method used by Clopper and Pearson in 1934 will be illustrated in section 5.5.3. Because this is an exact method we can state that we are 95% confident that  $[0.035, 0.320]$  captures the true but



unknown proportion of people who are left-handed given our sample of size  $n=30$  and we are 95% confident that  $[0.0025, 0.445]$  captures the true but unknown proportion of people who are left-handed given our sample of size  $n=10$ . We know that the level of confidence can be no less than 95% for each of these intervals.

### 5.5.2 A Comparison of the Normal Approximation and Clopper Pearson Methods

Below is a table compiling the information produced from each of the two methods for 95% confidence intervals for sample sizes of  $n=10$  and  $n=30$ .

Sample Size	Method	Interval	Width of Interval
$n=10$	Normal Approximation	$[-0.0859, 0.2859]$	0.3718
$n=10$	Clopper Pearson	$[0.0025, 0.445]$	0.4425
$n=30$	Normal Approximation	$[0.0117, 0.2550]$	0.2433
$n=30$	Clopper Pearson	$[0.035, 0.320]$	0.285

All methods of creating confidence intervals have the following two properties:

- 1) Confidence interval width increases as the sample size decreases.
- 2) Confidence interval width also increases as the level of confidence increases for all methods of creating confidence intervals.

Property (1) is true because the set of probable values for the unknown parameter  $p$  increases as the sample size decreases. In other words, the smaller the sample size, the less we know about the population parameter  $p$  and the more uncertain we are of its value. Property (2) is true because if we want increased confidence we must accept a larger set of probable values for the unknown parameter  $p$ . In the extreme case, a 100%

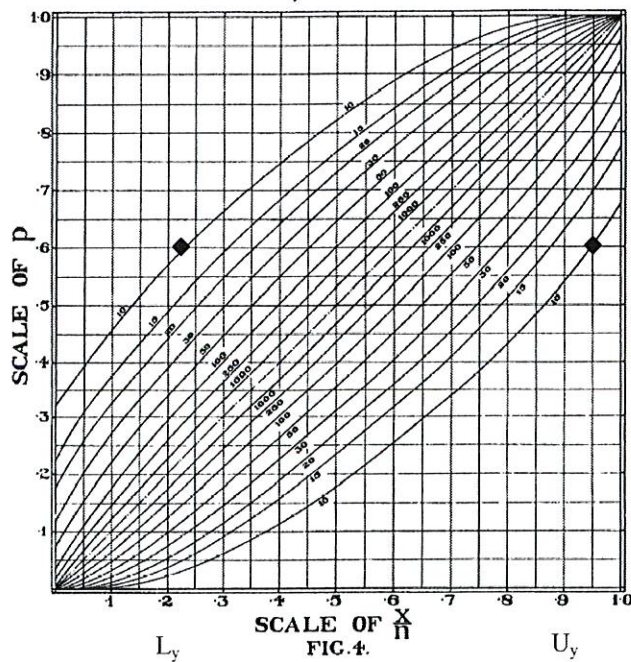
confidence interval for  $p$  is  $(-\infty, \infty)$ .

Note that for each sample size the width of the Clopper Pearson interval is greater than that of the normal approximation method. We know that the Clopper Pearson intervals are exact (i.e. the level of confidence is at least 95%) and the normal approximation intervals are not. Given (2) above, this implies that the level of confidence for the normal approximation intervals is less than 95%.

### *5.5.3 An Example of How Clopper and Pearson Computed Confidence Intervals in 1934*

The chart reproduced below (Clopper Pearson, 1934) demonstrates the method Clopper and Pearson used to present confidence intervals in 1934. Today, much of the work that they did by hand can be quickly done using statistical software such as Minitab. Notice that this particular chart is for creating confidence intervals with a confidence coefficient of 0.95 or a confidence level of 95%. An example will be given to demonstrate the technique Clopper and Pearson used to create this chart.

Suppose we let  $\alpha = 0.05$ ,  $p = 0.6$  and  $n = 10$ . Looking at the chart below, at  $p=0.6$  the values of the confidence interval endpoints of which we want to find are marked with diamonds.

CONFIDENCE BELTS FOR  $p$  (CONFIDENCE COEFFICIENT  $\cdot .95$ )

We want to find  $L_y$  and  $U_y$  of the confidence interval that would correspond to  $p = 0.6$ . Essentially, we want to find  $L_y$  and  $U_y$  so that  $P(L_y \leq p \leq U_y) = .95$ , an exact 95% confidence interval. Looking at the chart we have an idea of where  $L_y$  and  $U_y$  should lie (Clopper and Pearson did not). Since we are given that  $p = 0.6$  we have the following:

$$P(L_y \leq .6 \leq U_y) = .95. \quad (\text{Insert } p = .6)$$

$$P(10 L_y \leq 6 \leq 10 U_y) = .95 \quad (\text{Multiply by 10})$$

$$P(Y_L \leq 6 \leq Y_U) = .95 \quad (Y_L \text{ and } Y_U \text{ are values between 0 and 10})$$

$$P(y \leq Y_L) = .025, P(y \geq Y_U) = .025 \quad (\text{Assume equal tail probabilities})$$

Now that we have the equation in this form we can solve for  $Y_L$  and  $Y_U$ .

We will start by finding the value of  $Y_L$ . Looking at the chart, we have the advantage of knowing the general area of where our value should lie. We can see that  $Y_L$  lies



somewhere between the values 2 and 3. For  $Y_L=2$  we have:

$$\sum_{y=0}^2 \binom{10}{y} .6^y (1-.6)^{10-y} = 0.01229. \text{ Since } 0.01229 < .025, \text{ we know that } Y_L=2 \text{ is not large}$$

$$\text{enough. For } Y_L=3 \text{ we have: } \sum_{y=0}^3 \binom{10}{y} .6^y (1-.6)^{10-y} = 0.0516. \text{ Since } 0.0516 > .025, \text{ we}$$

know that  $Y_L=3$  is too large. Therefore, the value must lie between  $Y_L=2$  and  $Y_L=3$ .

To find the values of  $Y_L$  and  $Y_U$  we will use linear interpolation. For  $Y_L$ , we need to interpolate between the points, (2, .01229) and (3, .0516). The general equation of a line is  $y - y_I = m(x - x_I)$ , with  $m = (y_2 - y_1)/(x_2 - x_1)$ . In our example,  $m = (0.0516 - 0.01229)/(3-2) = 0.03913$ . Then, the equation of the line is  $y - 0.01229 = 0.03913(x - 2)$ . Solving for  $y$ , we have:  $y = 0.03913x - 0.06579$ . Letting  $y = 0.025$  implies  $x = 2.32$  which implies that  $Y_L = 2.32$ . Then, if  $Y_L = 2.32$ , we know that  $2.32 = 10L_Y$ , therefore  $L_Y = 0.232$  as is given on the chart by the left diamond.

To find the value of  $Y_U$  we will use the same method of interpolation. Again, from the chart we know that  $Y_U$  is between 9 and 10. For  $Y_U=9$  we have:

$$\sum_{y=9}^{10} \binom{10}{y} .6^y (1-.6)^{10-y} = 0.04636. \text{ Since } 0.04636 > 0.025, \text{ we know that } Y_U=9 \text{ is too}$$

$$\text{large. For } Y_U=10 \text{ we have: } \sum_{y=10}^{10} \binom{10}{y} .6^y (1-.6)^{10-y} = 0.00605. \text{ Since } 0.00605 < .025, \text{ we}$$

know that  $Y_U=10$  is too small. Therefore, the value must lie between  $Y_U=9$  and  $Y_U=10$ .

For  $Y_U$ , we need to interpolate between the points (9, 0.04636) and (10, 0.00605). The slope  $m = (0.00605 - 0.04636)/(10-9) = -0.04031$ . The equation of the line is  $y - 0.04636 = -0.04031(x - 9)$ , or simplified is  $y = -0.04031x + 0.40915$ . Again, let  $y = 0.025$ . Then, solving for  $x$  we have  $x = 9.5299$ , which implies  $Y_U = 9.5299$ . If  $Y_U = 9.5299$ , then

$9.5299 = 10U_Y$  implies that  $U_Y = 0.95299$  as is given in the chart by the right diamond.

Therefore, from this method we know that  $L_Y = 0.232$  and  $U_Y = 0.95299$ . The confidence interval we created is  $(0.232, 0.95299)$ .

Today, with the aid of computers and advanced calculators, the method of computing confidence intervals has been transformed. What used to involve extensive hand calculations can now be done almost instantaneously with the technology available. Charts are no longer used to display confidence intervals, rather a computer or calculator will calculate them for a given sample and level of confidence.

## 6. Conclusions

The history of left-handedness and the difficulties associated with estimating the proportion of the general public that is left-handed was explored. There is no consensus as to the true proportion of people that are left-handed. To explore statistical methods of estimating this unknown proportion  $p$  two samples were taken and analyzed using two methods of creating confidence intervals.

As demonstrated through the use of the two different methods, an approximate and an exact method, we see the importance of choosing a method based on sample size. For small sample sizes (anything below 30) the exact method as proposed by Clopper and Pearson is superior to the normal approximation method because we can be assured of at least a  $100(1 - \alpha) \%$  level of confidence. For large sample sizes the normal approximation methods yields acceptable confidence intervals with an approximate  $100(1 - \alpha) \%$  level of confidence.

Before the advent of computers the normal approximation method was an easy to

use satisfactory method. But, today, given the ease of which Clopper Pearson intervals can be calculated, there is really no reason to continue to use the normal approximation method, except as an easy way to comprehend examples in introductory statistics courses.

### References:

1. Casella, G. and Berger, R. *Statistical Inference*. 1990
2. Clopper, C.J. and Pearson, E.S. (1934). "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." *Biometrika* 26, 404-413.
3. Elias, Lorin. "Lorin Left-handedness Site." Internet.  
<http://duke.usask.ca/~elias/left/>. 3/25/03.
4. Fincher, Jack. *Lefties: The Origins & Consequences of Being Left Handed*. Barnes & Noble Books, 1993.
5. Hogg, Robert and Tanis, Elliot. *Probability and Statistical Inference*, Sixth Edition. New Jersey: Prentice Hall, 2001.
6. Mendenhall, William; Beaver, Robert J. and Beaver, Barbara M. *Introduction to Probability and Statistics*, Tenth Edition. Duxbury Press: An International Thomson Publishing Company, 1999.
7. Silverstein, Dr. Alvin, and Silverstein, Virginia B. *The Left-hander's World*. Chicago: Follett Publishing Company, 1977.



## Appendix

**Table I: Sample of size 30 taken on 16 February 2003**

1.	right	11.	right	21.	right
2.	right	12.	right	22.	right
3.	right	13.	right	23.	right
4.	right	14.	left	24.	right
5.	Left	15.	left	25.	left
6.	right	16.	right	26.	right
7.	right	17.	right	27.	right
8.	right	18.	right	28.	right
9.	right	19.	right	29.	right
10.	right	20.	Right	30.	right

Summary: 4 left-handers and 26 right-handers

**Table II: Sample of size 10  
obtained on 16 February  
2003**

1.	right
2.	right
3.	right
4.	right
5.	right
6.	right
7.	right
8.	right
9.	left
10.	right

Summary: 1 left-hander and  
9 right-handers