




Estimation and Testing for a Binomial Proportion From a Single Sample

Introduction	2
Validity Conditions for Using the Normal Distribution	3
Hypothesis Testing	4
Confidence Intervals	3
Appendix: Explanation of the Binomial Mass Function	10

Special sections:

	Core Concepts: In this chapter there may be material that we consider core material for <i>any</i> statistical analysis. So if you are reading for a refresher in general, don't skip these section.
	Some passages herein are for the statistical aficionado , and can be skipped by others.
	Nerd alert... There are some passage herein that are particularly nerdy, and can usually be skipped. The photo to the left will be your warning.

Introduction.

Is the mortality rate among Bonneville Cutthroat Trout (BCT) in the Thomas Forks irrigation canals (in SW Wyoming) less than 30%?¹ In the summer of 2002, Amy Schrank² tracked 40 radio-tagged BCT. Over the course of the summer, 9 (22.5%) died. We will use these data to illustrate confidence interval construction and hypothesis testing for Binomial proportions.

Binomial proportions data arise when each datum is recorded as belonging in one of two categories. For Amy's data, D (for dead) and S (for survived) might serve as labels. Other common pairs are Present/Absent, Male/Female, Success/Failure. Assuming each datum is independent of the others (i.e. that the survival or death of one fish does not depend on the survival or death of any other fish), this data is readily modeled by the so-called Binomial distribution. For more details about that distribution, see the Appendix.

A natural estimator for the true population proportion from the data is the sample proportion $\hat{p} = \frac{Y}{n}$, where Y is the observed number of events³. One feature of Binomial data is that the standard error of the sample proportion is determined solely by p and n : we estimate it with $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.



Nerd alert... The bit about the connection between the SE of a proportion and the SE of a mean can be skipped without it holding your life back in any particular way.

You might recall that the SE of a mean from a single sample is estimated by $SE(\bar{Y}) = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$, where s is the sample standard deviation. As it happens, the formula for the variance of a sample of 0 and 1 values is $\hat{p}(1-\hat{p})$. So the formula here is not as arcane as it might first appear. Further, \hat{p} is the sample mean of the data, if the data are viewed as 0s and 1s. Try calculating $\hat{p}(1-\hat{p})$ for a variety of values of \hat{p} , and you will see that it is at its largest when $\hat{p} = 0.50$ and diminishes as it gets smaller or larger⁴.

¹ I chose 30% (and a one-tailed test) for no particular biological reason... (so let's pretend that the historical value has been 30% and we are interested in seeing whether a management intervention has reduced the mortality rate.

² Amy was then a Ph.D. student at the University of Wyoming.

³ When I'm being generic, I'll use the term "event" to correspond to "the event of interest" (which in Amy's case is death of a fish, since she is measuring the mortality rate.

⁴ Keep in mind that it is trapped between 0 and 1...

Back to our regularly scheduled programming...

This direct connection between the estimator and its SE has ramifications when testing hypotheses, more on which later.

Example: the estimated mortality rate for the Bonneville cutthroat trout is $9/40 = 0.225$ (or 22.5%). The SE of this rate is estimated to be $SE(\hat{p}) = \sqrt{\frac{.225(1-.225)}{40}} = 0.066$ (Or 6.6%).

Validity Conditions for Using the Normal Distribution.

The usual tool for testing hypotheses or making confidence intervals for Binomial proportions is the Normal distribution. Its use as a model for the distribution of a proportion is justified if the following conditions are met.

1. The sample should be no larger than 10% of the population (if the sample *is* larger, the sampling distribution of the proportion becomes non-Normal)⁵.
2. On the other hand, the sample needs to be large enough that the expected number⁶ of 1's and 0's are both larger than 10. For sake of a label, we'll call this the "10/10" condition⁷.
3. The data come from a suitable random sample of independent observations.

Some statisticians recommend a "5/5" rule instead of 10/10; my experience suggests that 5 is too low. However, it is true that the cut-off points used here do not represent absolutes. It is simply that as the numbers of both successes and failures get larger, the Normal approximation gets better. Judge for yourself.⁸

⁵ This condition is only relevant if you are sampling from a finite population. An example might be taking a sample from the population composed of all the students in an introductory statistics class. The population size might be (say) $N = 150$. In that case a sample of size $n = 75$ would constitute $\frac{1}{2}$ of the population.

⁶ The "expected number" is determined by the underlying probability of a "1" and sample size. Since we don't know that probability (else we would not be sampling to estimate it), we have to use our observed tallies.

⁷ Scott, who has higher standards than does Ken, prefers a "15/15" condition.

⁸ We have an interactive Excel tool (**Distribution of p**) that will give you the chance to develop that judgment.

We will note here that we are using the standard Normal distribution (so-called z distribution) rather than the t . The reason⁹ lies behind the fact that the estimate of the sample SD is driven by knowing the sample proportion; i.e. the SD estimate is not a separate, independent calculation.

Hypothesis testing for a single Binomial Proportion.

If you are rusty on the core concepts and methods behind hypothesis tests, we recommend you read **Hypothesis Testing** (in the **Big Ideas** folder).

The conceptual starting point for a hypothesis test is to assume the null is true and ask, in a particular way (measured by the p -value¹⁰), whether or not your data is consonant with that hypothesis. That conceptual starting point and the fact that the SD of a proportion is determined by the proportion itself conspire to create the SE formula in step 3.

The method whose details are shown in these steps is the standard method which assumes Normality for the sampling distribution of the sample proportion. These days, statistics packages usually offer the alternative of doing the calculations using the Binomial distribution directly. We have no strong preference between the two methods; we note, however, that should the p -values differ markedly, it is likely that the Normal approximation version is not valid, likely due to a small sample size. In that case the p -value arising from using the Binomial distribution is likely more valid.

⁹ We omit the technical details in order to avoid injurious eyeball spinning...

¹⁰ Confusing notation that will require you to be alert... The statistic arising from a hypothesis test is conventionally labelled "the p -value". Also convention is to label the sample proportion of events of interest with the letter p (usually shown as \hat{p} which might help a little). Still, for Binomial proportion data, you'll have to pay attention to be sure which p means what... good luck. There is over 100 years of tradition behind this notation. Our pledge: we will fight to change it when the U.S.A. switches to the metric system ☺.

Steps:

1. State the appropriate alternate and null hypotheses (take care in deciding whether the alternate should be two-tailed or one (and if one, which direction). Choose an alpha-level.

The cutthroat trout research question asked whether the mortality rate is now less than 30%. That informs the alternate hypothesis¹¹: $H_a : p < 0.30$. Given that, we get

$H_o : p \geq 0.30$, the nullity of the alternate. We will use 0.3 as the test value. For sake of avoiding flak from our colleagues¹², let's use alpha = 0.05.

2. Check the validity conditions.

Our example: if the null hypothesis is correct (using 0.3 as the representative value), then we might, on average, expect $0.3 \times 40 = 15$ mortalities and 25 survivors. This meets the 10/10 rule and so using the Normal distribution ought to be valid.

3. Compute the p-value of the test. This can be done for you easily using a statistics package. That being said, here are the elements, using our trout data setting as an example.

a. Compute the SE: $SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.30 \times 0.70}{40}} = 0.072$. We subscripted

SE and p to remind us that we are using the values prescribed by the null hypothesis.¹³

¹¹ Note here that we wrote p not \hat{p} . We are making a test about the true proportion in the population; we will of course use the sample estimate in evaluating that test. Note also that we wrote the alternate hypothesis first. That is indeed the logical order, since the null hypothesis gets formed from it.

¹² There is an implied jab in how we worded this. Alpha = 0.05 is indeed a standard choice, but it *is* a choice, and should be done thoughtfully. Many times, it is not: folks use 0.05 because everyone around them uses it (some might even believe that one *must* use it). Truly it is not an unreasonable choice for many cases...

If you use 0.05, nobody will blink. If you use some other value, you are certain to be questioned, "Why did you choose that alpha level?" We believe that question should *always* be present, no matter the choice.

¹³ Remember that the test proceeds by acting as though we believe the null is true, and then assessing the strength of evidence against it. In this way, a statistical hypothesis test is like a criminal court case. The case moves forward because law enforcement personnel have reason to believe the dirty deed was done. Still, "innocent until proven guilty, right?" If the defendant in question was your statistics professor, this proposition might be quite shaky. Still...

- b. Determine how far your observed proportion is (in SE-units) from the hypothesized proportion (the so-called z -value):

$$z = \frac{\hat{p} - p_0}{SE_0(\hat{p})} = \frac{0.225 - 0.30}{0.072} = -1.04..$$

- c. Determine the p -value.
- i. If your test is two tailed, remember to do the calculation to determine the probability of a bigger absolute difference (positive and negative).
 - ii. If one-tailed, do the calculation ONLY if the observed proportion is going in the same direction as suggested by the alternate hypothesis. If it is not, you clearly have no evidence against the null in favor of your alternate. In our example, the observed proportion is indeed less than 0.3, so we need to assess the level of evidence. The p -value using the Binomial distribution itself is 0.196; that using the z distribution is 0.150.

4. Make appropriate conclusions.

Here, the two p -values differ, but not importantly. Using either one, we do not have enough evidence (against $\alpha = 0.05$) with which to reject the null hypothesis. The fact of our observed value being lower than 0.30 appears to have just been random chance.

Confidence Intervals for a single Binomial Proportion

Recall that, with an ideal confidence interval procedure, the percent of intervals containing the parameter being estimated would be precisely your chosen confidence level. For example, if you are making a 95% C.I. for a proportion, it ought to be the case that, using that procedure, 95% of intervals so created would actually contain the population proportion. We refer here to the nominal confidence level (your choice) and the attained confidence level. Ideally, they are the same.

There are many methods proposed for computing C.I.s for Binomial proportion. I will consider three here. First, the standard method, which, ironically, is losing its status as the standard. Second, there is a method that uses the Binomial distribution directly, the so-called

“exact” method¹⁴. This is currently the default in many statistics packages, including Minitab. The third method, which I believe will become the standard¹⁵ is called the Agresti-Coull interval (after its creators) or the “+4” method (after an easy-to-implement approximation), valid for 95% intervals. For our purposes here, we will use 95% as our choice, in part because it is complementary to the alpha (0.05) we used in the hypothesis test.

The standard method uses the formula $\hat{p} \pm z_{CL} \times SE(\hat{p})$, where

- $\hat{p} := \frac{Y}{n}$ is the observed proportion of events of interest (there being Y of them in the sample of size n). For our example, $\hat{p} = 0.225$.
- z_{CL} is a value from the standard normal distribution that determines how many SEs wide to make the interval¹⁶. For our example, $z_{.95} = 1.96$.
- $SE(\hat{p}) = \sqrt{\frac{.225 \times .775}{40}} = 0.066$.

Thus for our example, the classically constructed 95% CI is

$\hat{p} \pm z_{CL} \times SE(\hat{p}) = 0.225 \pm 1.96 \times 0.066 = 0.225 \pm 0.129 = (0.096, 0.35)$. For reporting purposes, we might round this to (0.10, 0.36). Note, though that the validity conditions for using the Normal distribution are shaky, since there were only 9 mortalities. Not fatally flawed, but also not ideal.

One alternative is to use the Binomial distribution directly (the so-called “exact” or Clopper-Pearson method^{17, 18}). It produces conservative intervals (meaning that the attained confidence level is always at least as high as the nominal one). For small sample sizes, it is quite

¹⁴ Cautionary note: This is called the “exact” method because it uses the Binomial distribution formulae itself, not some approximation. Ironically, it does not produce confidence intervals that attain *exactly* (say) the 95% confidence level. The method tends to produce intervals with a higher attained confidence level than the one you choose, implying that you are getting intervals that might be a tad wider than you need.

¹⁵ At least, it will be so for those who choose 95% as their confidence level, which is what happens approximately 99.999% of the time...

¹⁶ For instance, if your choice is a 90% interval, $z = 1.645$. That is because mean plus and minus 1.645 SEs captures the middle 90% of the distribution.

¹⁷ Charles Clopper and Egon Pearson are the two statistician who worked out the details for that method; the “exact” in the name does not imply an exact connection between nominal and attained confidence levels; rather, the computation is “exact” in that it doesn’t make any assumptions about the sampling distribution for the sample proportion (beyond that the Binomial distribution is an appropriate model for the data).

¹⁸ Egon Pearson, a famous statistician in his own right, was the son of the illustrious Karl Pearson. Karl and Ronald A. Fisher are responsible for much of the foundational work in statistical science. They also famously butted heads frequently (big egos, anyone?).

conservative. Some statisticians prefer it since the attained level is never smaller than the nominal; others, noting that this implies wider intervals than are desired, don't prefer it. We tend to fall in the latter camp. For our example, the exact confidence interval is (0.11, 0.38). This is shifted slightly, and slightly wider than the classical interval.

There is a third alternative that uses slightly different arithmetic for forming the interval (but otherwise follows the standard steps that use the Normal distribution). Alan Agresti and Brent Coull (1998) showed that superior confidence interval behavior (i.e. the actual confidence level is closer to the stipulated level) can be had by doing the usual confidence interval calculation on fudged data: for a 95% C.I., add 2 successes and 2 failures (apparently increasing the sample size by 4)¹⁹. Symbolically, that amounts to

- Define $p_{AC} := \frac{Y+2}{n+4}$.
- Use as the SE, $SE(p_{AC}) := \sqrt{\frac{p_{AC}(1-p_{AC})}{n+4}}$.

Then apply the classical formula using these values. We emphasize that you are not changing your estimate ($\hat{p} = 0.225$ is still it), it is simply a different algorithm for getting a confidence interval, an algorithm with apparently better properties.

For the Bonneville cutthroat trout data, that would be 11 1s and 33 0s. The resulting 95% CI is (0.122, 0.378). The Agresti-Coull method (also known in intro stat textbooks by the name “+4 method”, applicable for 95% confidence intervals) is becoming widely recommended, as it balances ease of implementation and technical properties. We repeat: **this ONLY works if you choose 95% as your confidence level.**

¹⁹ We note that this is not adding fake data. The data are the data are the data. This is simply a different way to get from data to confidence interval.

References:

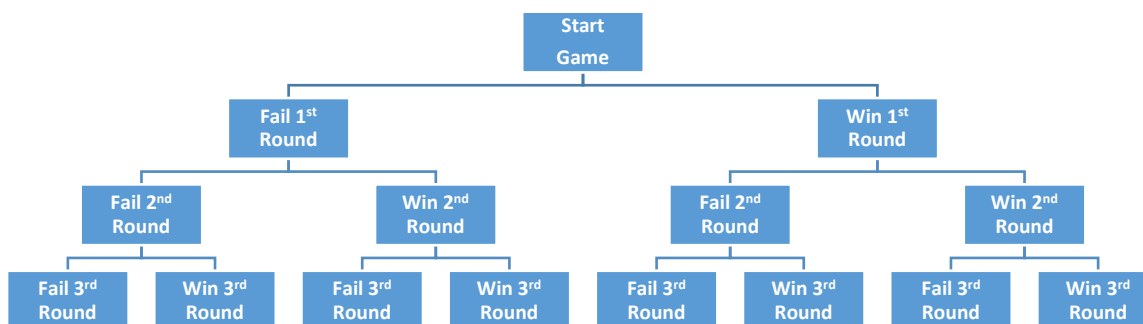
Agresti, A., and B. Coull, 1998. "Approximate is better than 'Exact' for Interval

Estimation of Binomial Proportions." *The American Statistician*, 52: 119-126.

Brown, L. D., Cai, T., and DasGupta, A. (2001), "Interval Estimation for a Binomial Proportion (with discussion)," *Statistical Science*, 16, 101-133.

Appendix: Explanation of the Binomial Mass Function²⁰

For sake of ease of illustration, suppose you have three chances to win a prize (you may win 0, 1, 2, or 3 times) in some game, with probability 0.7 of winning each time. Each trial is independent of the previous trials.



Result	FFF	FFW	FWF	FWW	WFF	WFW	WWF	WWW
Probability	.3×.3×.3	.3×.3×.7	.3×.7×.3	.3×.7×.7	.7×.3×.3	.7×.3×.7	.7×.7×.3	.7×.7×.7

So, for instance, the chance of winning twice can be reckoned by noticing that there are three ways it can occur (shown in green in the last row), and that each of them has probability $.3 \times .7 \times .7 = .147$ (the order of the two wins and one loss doesn't matter).

Imagine trying this so-called “branching tree” method to figure out the probability of (say) 3 successes in ten tries, with some specified probability. If that prospect leaves you gasping for air, it should! Without deriving it, let us show you a formula that works. And let's make it generic: sample size is n , number of successes Y (Y being one of 0, 1, 2, ..., n), and the probability of success is labeled p .

The probability of any one sequence containing Y successes is $p^Y (1-p)^{n-Y}$ (compare that, say, to our observed $.3 \times .7 \times .7$).

²⁰ As the Binomial Distribution is more formally called...

The number of ways that Y successes can appear in n trials is

$$\frac{n \times (n-1) \times (n-2) \dots \times 1}{[Y \times (Y-1) \times (Y-2) \dots \times 1] \times [(n-Y) \times (n-Y-1) \times (n-Y-2) \dots \times 1]}$$

This generic formula falls apart (apparently) if n and/or Y are small; I wrote it the way I did just for illustration. For instance to calculate the number of ways of getting 3 successes in 10 tries, it

is $\frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times (7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} = \frac{10 \times 3 \times 4}{1} = 120$ (with some obvious cancellations). For our

small ($n = 3, Y = 2$) example, it yields $\frac{3 \times 2 \times 1}{(2 \times 1) \times (1)} = 3$ (check that against the tabled example).

The mechanics here are clear, but the notation is still a little unwieldy. Long, long ago, someone came up with notation called *factorial* (which uses the exclamation mark “!”), defined as follows: For any integer²¹ n , $n! := n \times (n-1) \times (n-2) \dots \times 1$, with “1!” and “0!” both defined to just be 1. Then the notation for counting the number of ways that Y successes can appear in n trials is simply presented as $\frac{n!}{Y!(n-Y)!}$. Much tidier, but apparently not tidy enough.

Mathematicians and statisticians sometimes write this in an even more abstract form: $\binom{n}{Y}$ (read as “ n choose Y ”).

Hence the formula for calculating the probability of Y successes in n trials (with chance of success p) is usually given as

$$f(Y|n, p) = \binom{n}{Y} p^Y (1-p)^{(n-Y)} = \frac{n!}{Y!(n-Y)!} p^Y (1-p)^{(n-Y)}$$

The notation on the left side of the equals sign reads as “probability of Y , given²² n and p ”.

²¹ This factorial business is defined for non-integers also, but we don’t need to deal with that here.

²² That is the spoken meaning of the vertical bar “|”.