

Why $np < 10$?

Standard inference procedures are based on certain theoretical conditions. If the conditions hold for a particular situation, then using the corresponding procedure is reasonable. That is, significance levels and confidence levels provide accurate and meaningful information about the conclusions of the procedures. However, as the necessary conditions are relaxed, what happens to these values? In hypothesis tests, α is supposed to measure the rate at which the test falsely rejects a true null hypothesis. Likewise, the confidence level is supposed to measure the rate at which confidence intervals contain the true population parameter in repeated sampling. When conditions are not met, what happens to error rates, and why?

This article focuses on hypothesis tests. Results for confidence intervals would be similar. See also John Lieb's article, "Violating Conditions in Significance Testing," at this site for discussions of similar questions related to t-tests. All figures in this article were generated using TI-Interactive! software.

For large-sample inference on one proportion, one common set of conditions, each apparently related to sample size, is:

- (1) $np \geq 10$ and $n(1 - p) \geq 10$, and
- (2) $N \geq 10n$ (if sampling without replacement).

For such tests n represents sample size, N is population size, and p is the proportion of successes. Some texts use 5 instead of 10 in condition (1).

Why are these conditions necessary? Why are there two conditions instead of just one? Why don't books agree on the "magic" number? What happens if the conditions are not met?

This article discusses violations of condition (1) from several points of view. It is suggested that this material be used when the z-test for proportions is first introduced. A theoretical treatment of condition (2) appears in lecture notes from John Cryer and Jeff Witmer on the topic of the hypergeometric distribution in the Introduction to Theory of Inference at http://phywww1.ncssm.edu/green/Math/Stat_Inst/Notes.htm

Background:

The z-test for proportions uses the normal approximation to the binomial and the facts that the mean of a binomially distributed random variable is np and its standard deviation is $\sqrt{np(1 - p)}$ to describe the test statistic. The effectiveness of the approximation can be judged in two ways:

Qualitatively: examine overlaid graphs of the two distributions for shape and domain
Quantitatively: compare α to the actual rate of rejecting a true null hypothesis about p

In fact, the quantitative evaluation can actually be accomplished in two ways, by direct calculation from the theoretical distributions and by simulation. This activity addresses the theoretical approach.

In order to assure that condition (2) plays no role, we assume that sampling is from an infinite population (i.e., with replacement). With that restriction, samples of size n drawn independently from a population having proportion p of “successes” produce a sampling distribution of “number of successes” x having a binomial distribution with mean np and variance $np(1-p)$.

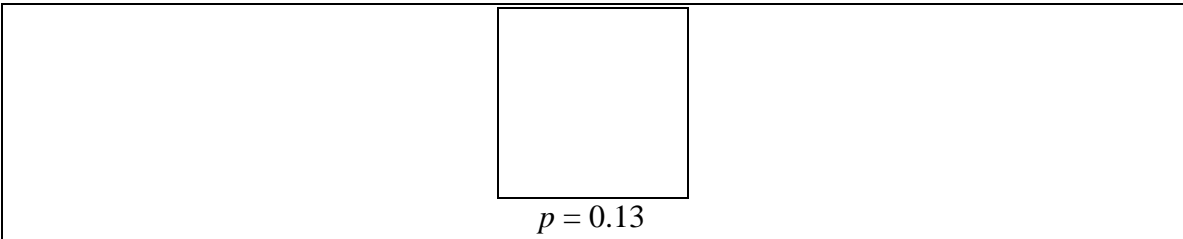
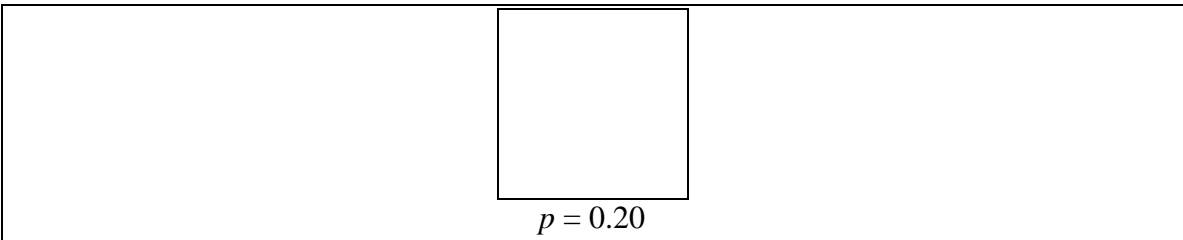
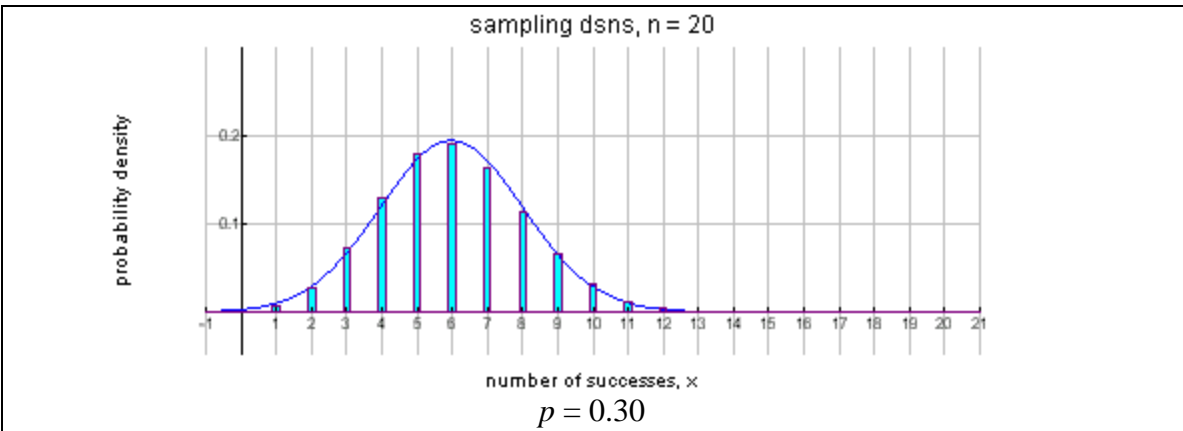
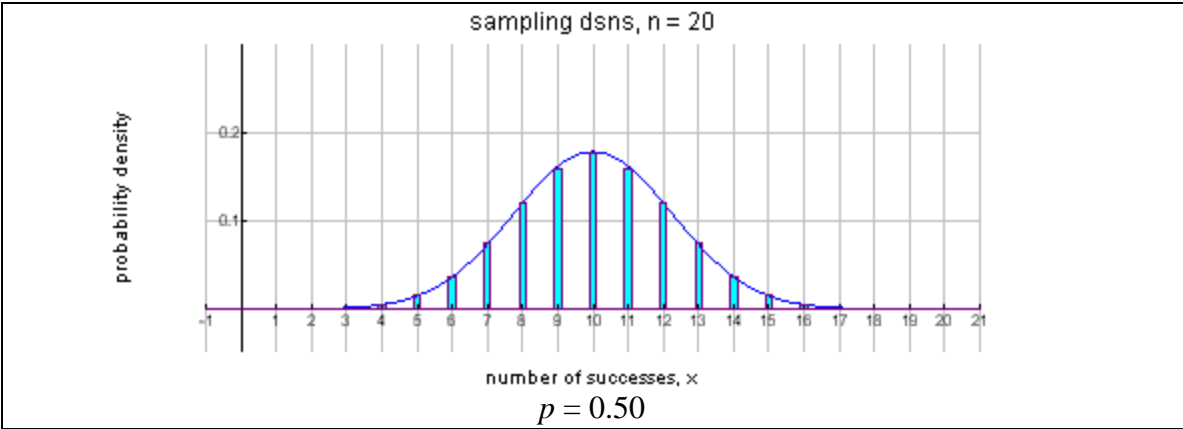
Graphs:

The graph of the distribution of x and the graph of a $N(np, \sqrt{np(1-p)})$ random variable should be comparable if the normal approximation is good. To evaluate the quality of the approximation graphically, compare overlaid plots for a variety of values of n and p . In particular, since the condition we are examining sets a lower bound on np , comparing graphs for low values of np is most instructive.

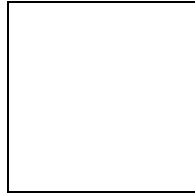
Two TI-Interactive! documents, “np demo 20” and “np demo 50” are available. The documents display overlaid graphs of the corresponding binomial and normal distributions for $n = 20$ and $n = 50$, respectively. Each document permits the user to change the value of p by way of a “slider” and observe the corresponding changes in the two distributions’ graphs. For classes in which this software is not available, similar explorations may be done on calculators or using other computer software. With calculators, of course, the dynamic aspects of the slider are lost.


Note that the graphs of the binomial distributions are constructed using “narrow” histogram bars. This is to emphasize the discrete “integer values only” aspect of such distributions. We will say more about that later.

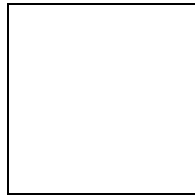
Since $(20)(0.5) = 10$, we might expect that only the $p = 0.5$ graphs are reasonably similar for $n = 20$. A little experimentation shows that such is not the case! In fact, for p as low as 0.3 (i.e., $np = 6$) the two graphs appear essentially identical. With p around 0.2, we see that the left tail of the normal distribution noticeably enters the $x < 0$ region, something the binomial never does. As p decreases further, its histogram bars begin to “poke through the top” of the normal approximation, and the asymmetry of the binomial becomes more apparent. The figures below are typical.

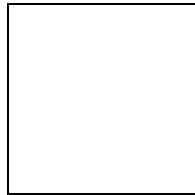


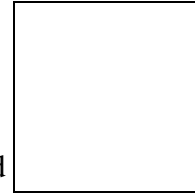
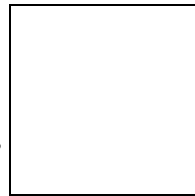
Experimentation with graphs for $n = 50$ show similar behavior. For $np = 10$, the normal and binomial graphs are remarkably alike, and the left tail of the normal goes negative around $np = 5$. Analogous behavior is observed as $n(1-p)$ decreases (that is, when p is near 1).



Hypothesis Tests and  **:**

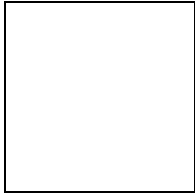


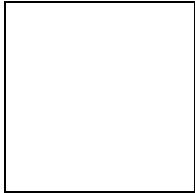
In hypothesis tests,  is supposed to measure the rate at which the test falsely rejects a true null hypothesis. For a one-sided ($<$) z-test of proportion we reject

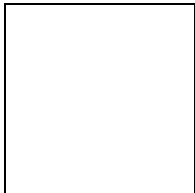


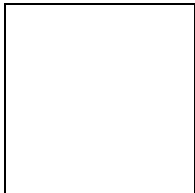
the null hypotheses for all $x < L$, where L satisfies  and . A

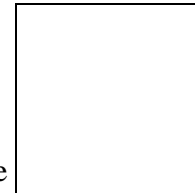
one-tailed ($>$) z-test would reject for $x > R$, defined analogously. In short, then, based on



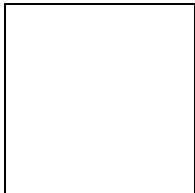
 we determine a “cut point” beyond which an observed sample statistic x leads to our rejecting the null hypothesis. (Note: This process uses the number of successes, x , as the test statistic rather than the more familiar proportion of successes,



. For a fixed n the two procedures are equivalent since x/n .)



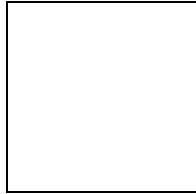
Any practical hypothesis test will sometimes incorrectly reject its null hypothesis due to random variation among samples. The probability of that occurring is supposed to be



. However, since x is actually binomially distributed, the exact probability

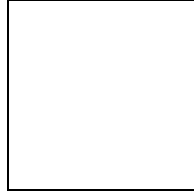
of incorrectly rejecting the null hypothesis for a given p is obtained from the binomial distribution function. That is, the *test* is designed and carried out using the normal approximation, but the *actual* “error rate” comes from a binomial distribution.

The TI-Interactive! documents “np demo 20” and “np demo 50” also permit the user to



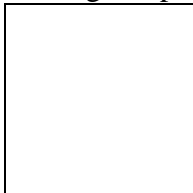
specify a desired value. The program displays the corresponding cut points for both left- (l) and right-tailed (r) tests, together with the actual error rates obtained from the binomial distribution. Again, similar results may be obtained using calculators or other computer software.

One observation is immediately obvious and is easily predicted from the earlier graphical explorations. That is, as the cut point for the rejection region moves below 0 or above 1,

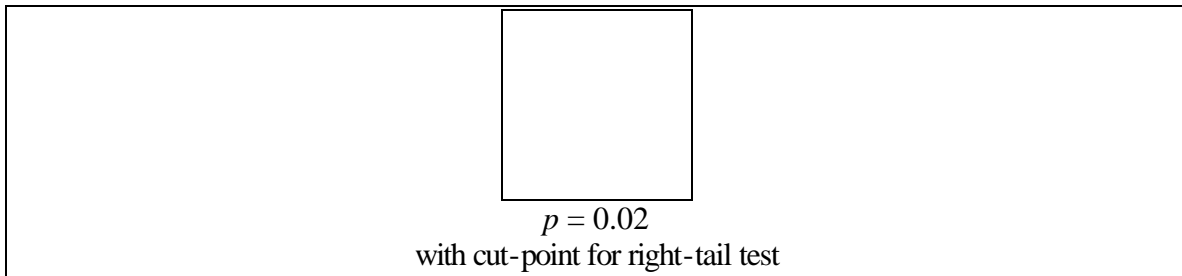


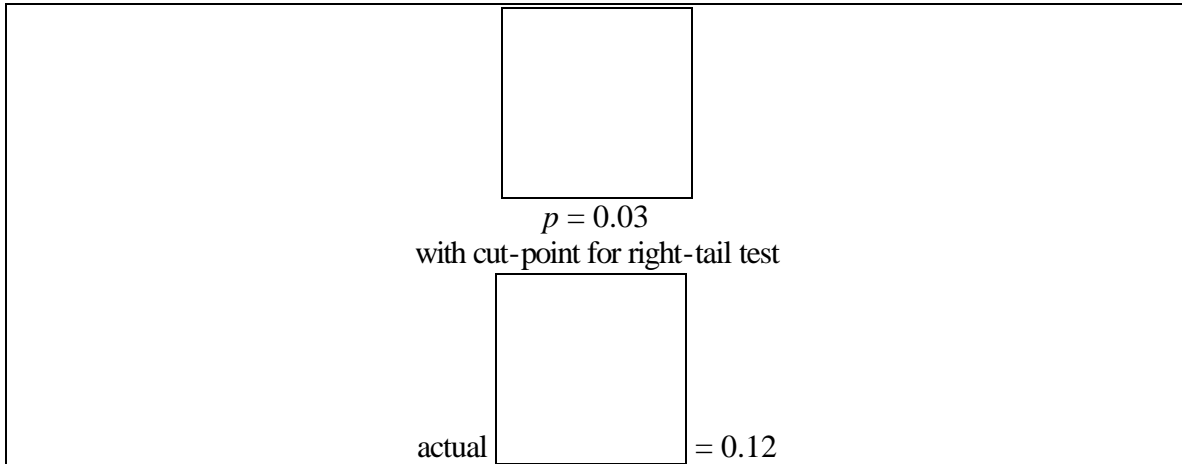
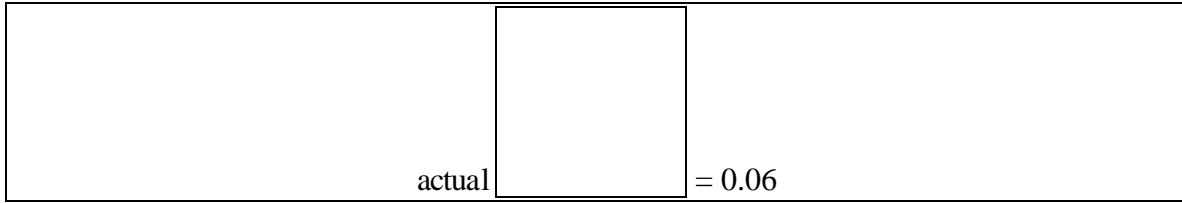
the actual probability of rejection is 0, not . However, it is instructive to carry out this comparison of significance levels slowly, varying p by 0.01 at a time, if possible.

We might expect that as np monotonically decreases the disparity between the value of

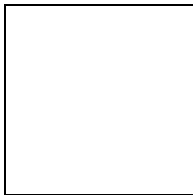


used to set the cut point and the actual probability of rejecting the null hypothesis would steadily increase. Such is not the case! In fact, with $n = 20$ and $p = 0.02$ (so $np = 0.4$!) the actual significance level for a right-tailed test is about 0.06. By comparison, the corresponding error rate for $p = 0.03$ is about 0.12, twice as high! What’s going on?



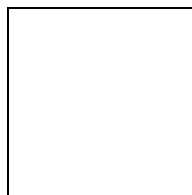


Remember the comment above about the “integers only” nature of binomial random variables. As p varies, probabilities for each x value change (look at $x = 2$ above). However, so do the mean and standard deviation of the underlying distribution. That changes the corresponding left and right cut points. As the cut points “move past” integer values, discrete “chunks” of probability are gained or lost. Thus not only do the probabilities for individual x values change, the x -values included in the rejection region change. So the behavior of the actual probability of rejection is highly unpredictable, with some values of np producing actual error rates near or below the specified



while nearby values have much larger error rates.

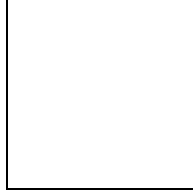
Other observations from this exploration include the following:

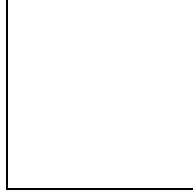


- z-tests may have error rates lower than ; not all “bad” np values increase errors.
- Larger n means more values of x are possible, so the binomial probabilities are “spread across more x values, making each histogram bar relatively smaller. This

decreases the variability of error rates as np changes. In fact, the reason for requiring np to be above some minimum value is more to eliminate this unpredictability than that small np is always bad.

- Except in the most extreme cases (gross violations having $np \ll 5$), differences



between  and the actual error rate are generally less than 0.05. This makes it very difficult to see these discrepancies using simulation without very large numbers of replications.