Calculus 111, section 08.y Proportions and Sample Size

notes by Tim Pilachowski

If you haven't done it yet, go to the Math 111 page and download the handout: Central Limit Theorem supplement. Today's lecture will use the material on pages 6 of this supplement and pages 15-19 of Appendix E of your text. Tonight's homework assignment is #4 and #5 from the supplement, along with #1, #3 and #5 from the text.

Before this, when we have looked at a binomial experiment, like tossing a coin, we defined our random variable X in terms of number of successes, for example X = number of heads.

We're going to change that a little, and define a new random variable,

$$Y = \text{proportion of heads} = \frac{\text{number of heads}}{\text{number of tosses}} = \frac{X}{n}.$$

By repeating the experiment (tossing a coin) for n trials, we are creating a sampling distribution.

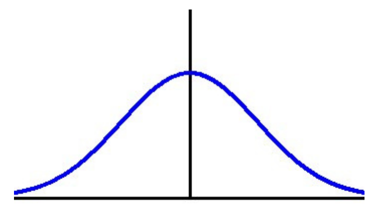
At the center of this sampling distribution we'll have $\frac{\mu}{n} = \frac{np}{n} = p$, i.e. a proportion of successes.

The standard deviation (standard error) for this sampling distribution will be $\frac{\sigma}{n} = \frac{\sqrt{npq}}{n} = \frac{\sqrt{npq}}{\sqrt{n^2}} = \sqrt{\frac{pq}{n}}$.

As the number of trials n increases, the shape of the distribution becomes increasingly closer to a normal distribution.

(See page 6 of the supplement for a demonstration and interpretation of this idea of proportion applied to tossing a coin 40 times.)

Example A. One hundred students answer a True-False question by guessing. What is the probability that more than 60% guess correctly?   *answer*: 0.0228



We're going to rearrange the z-score formula a little.

Appendix E, pages 15-19, of your text deals largely with this concept of margin of error, which it designates as $d = z * \sqrt{\frac{pq}{n}}$. (When news reports provide information about a survey or poll, they often say something like, "with a margin of error of 3%.)

Specifically, if we want to estimate a population proportion by taking a sample, and want to have our sample proportion end up within a specified margin of error of the actual (but unknown) population proportion, how large must our sample be?

We'll algebraically solve for n.

We can refine this to make it easier to remember and apply.

Since we don't know the population parameter p, what values should we use for p and q in the formula?

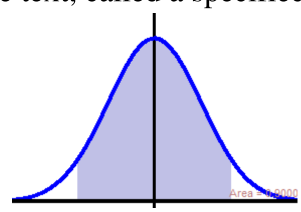
p	0	0.2	0.4	0.5	0.6	0.8	1
q	1	0.8	0.6	0.5	0.4	0.2	0
pq	0	0.16	0.24	0.25	0.24	0.16	0

The most conservative values are $p = q = 0.5$, $pq = 0.25$.

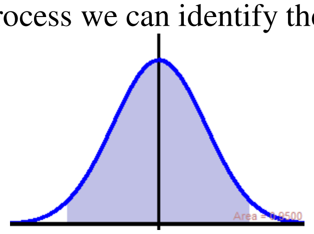
$$\text{sample size } = n = \frac{z^2 pq}{d^2} = \frac{z^2}{d^2} * \frac{1}{2} * \frac{1}{2} = \frac{z^2}{4d^2}$$

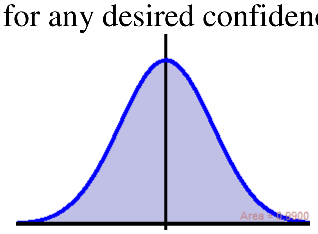
The value we use for z will be determined by the specified level of confidence (in the text, called a specified " % sure" or " % certain").

For a 90% confidence level (i.e. we want to be 90% sure of our result), 90% in the middle of the probability distribution will put the remaining 10% split between the two tails. Our question becomes P ($Z <$ what value) = 0.0500 ?



Using a similar process we can identify the critical value for z for any desired confidence level.





A more important conclusion, e.g. a life-and-death situation, would call for a smaller margin of error and/or a higher confidence level.

Example B-1. The re-election campaign of Senator Phil E. Buster will conduct a survey to find out how he is faring among the voters. If campaign staff wants the results to be 95% certain, with a margin of error of 2%, how large should the sample size be? *answer*: 2401

Example B-2. The re-election campaign of Senator Phil E. Buster will conduct a survey to find out how he is faring among the voters. The campaign staff wants the results to be 95% certain. How would the sample size calculation change if the desired margin of error, instead of being 2%, were 1%? 5%? *answers*: 9604, 385

Example B-3. The re-election campaign of Senator Phil E. Buster will conduct a survey to find out how he is faring among the voters. The campaign staff wants a margin of error of 2%. How would the sample size calculation change if the confidence level, instead of being 95%, were 90%? 99%? *answers*: 1692, 4161