Math Stat 465

Notes on the sample size for confidence intervals for p

In formula (8.12), p.389 (the sample size needed for a confidence interval for p), just use the "+" sign. The $\pm$ is an error.

By factoring out z^2, this formula may be simplified to

$$n = \left(\frac{z}{w}\right)^2 \left[2\hat{p}\hat{q} - w^2 + \sqrt{4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^2}\right]. \tag{1}$$

This may be a more useful form for computation. (It depends on the quantities z, w, and $\hat{p}\hat{q}$.) **Caution**: w is the desired **width**, not the radius, of the interval.

As the text notes, formula (1) involves us in a "Catch-22": In order to calculate $\hat{p}$ we have to know n, but **the formula for n requires us to know the estimate $\hat{p}$ already**.

Here are some possible ways to determine the desired n from the formula:

- We may use a value of $\hat{p}$ from a previous sample.
- We may use a value of p that we believe to be true for theoretical reasons.
- When we have no prior information about p, we usually use the value $\hat{p} = 1/2$. In most cases this produces the largest value of n, and so is a conservative choice. When $\hat{p} = 1/2$, formula (1) further simplifies dramatically to

  $$n = \frac{z^2(1 - w^2)}{w^2}, \tag{2}$$

  since the entire radical $= 1/2$. This is definitely the formula of choice when there are no assumptions about p. Formula (2) is not given in the text.

The text (page 389) gets an approximation from (1) by neglecting certain terms:

$$n = \frac{4z^2\hat{p}\hat{q}}{w^2}.$$

If we use $\hat{p} = \hat{q} = 1/2$, this becomes $n = z^2/w^2$, which is slightly bigger than (2). Thus formula (2) is preferable.