Math 316

Section 9.3 Other than 95% Confidence Levels

I think it makes a little more sense to discuss the Section 9.3 ideas before talking about the Section 9.2 ideas, which are just a special case (95% confidence) of the 9.3 ideas.

Where "proportion" means "the proportion of the population that is whatever it is that we are interested in" (for example, in **HW 9.3.1**, "proportion" is "what proportion of the children are iron deficient"), then we have the following notation:

p is the true proportion

$\hat{p}$ is the estimated proportion from the sample

So if n is the sample size and y is the number in the sample that is "whatever we are interested in," then $\hat{p} = \frac{y}{n}$.

A usual, we can use information from a sample to make an estimate about the population. For example, in the past, given sample mean $\bar{y}$ and standard error $SE_{\bar{Y}}$ (where $SE_{\bar{Y}} = SD/\sqrt{n}$) we have $\mu = \bar{y} \pm t_{\alpha/2} \cdot SE_{\bar{Y}}$.

A dichotomous variable is a categorical variable which has only two possible values. Example: a flower is red or it is not. We can normally approximate a binomial (a.k.a. dichotomous) distribution. This is covered in Section 5.4, which we skipped. This means we get to use a z value (rather than a t value) in finding the confidence interval.

Using the above ideas and notation, we find confidence interval $p = \hat{p} \pm z_{\alpha/2} SE_{\hat{P}}$ where $SE_{\hat{P}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. (Notice that as usual SE is smaller for larger n.) Let's work **HW 9.3.1**.

Actually, it turns out that a slightly modified version of $\hat{p}$ may be a better estimate of p. This Wilson-Adjusted Sample Proportion is

$$\tilde{p} = \frac{y + 0.5\left(z_{\alpha/2}^2\right)}{n + z_{\alpha/2}^2} \text{ and } SE_{\tilde{P}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z_{\alpha/2}^2}}$$

which leads to

$$p = \tilde{p} \pm z_{\alpha/2} SE_{\tilde{P}}$$

where $z_{\alpha/2}$ in all three formulas is found in Table 3 by the desired level of confidence. Let's rework **HW 9.3.1** using these modified values. This is actually how HW 9.3.1 should have been worked, using the modified values. We can also find **one sided confidence intervals** which we'll do in class.

We can estimate the sample size needed to get a certain interval width. Let's do so in the previous example, **HW 9.3.1**.

Section 9.2 95% Confidence Interval for a Population Proportion

One common confidence level is 95% (so a significance level of $\alpha = .05$), in which case $z_{\alpha/2} = 1.96$, in which case we have the modified sample proportion is

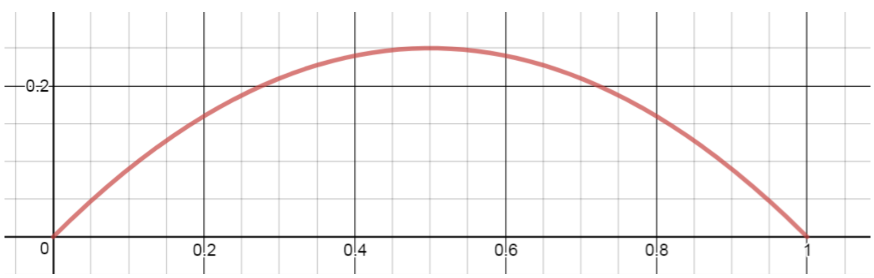
$$\tilde{p} = \frac{y + 0.5(1.96^2)}{n + 1.96^2} \approx \frac{y+2}{n+4}$$

as given on page 355, and

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+1.96^2}} \approx \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

as given on page 361. Section 9.2 deals with this particular case of 95% confidence. Section 9.3 (that we discussed above) deals with the more general case: confidence of any (not just 95%) level.

By the way, for a given value of $n + z_{\alpha/2}^2$, $SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+z_{\alpha/2}^2}}$ is largest when $\tilde{p}$ is ≈ 0.5 and smallest when $\tilde{p}$ is around 0 or 1, as seen in the plot of $\tilde{p}(1-\tilde{p})$ below.



One consequence of this is that if you are unsure what $\hat{p}$ (i.e. $\tilde{p}$) is, to be conservative (to make sure that sample size n is sufficiently large), use $\tilde{p} = 0.5$. See **HW 9.2.12**.

I've decided we'll not cover Section 9.1. To me it's more confusing than helpful.