Derivation of the Standard Error formula for comparing two proportions

Distribution of the Difference is Normal

If we are considering the difference of the values in the two populations...

$$H_0 : p_1 - p_2 = 0$$
$$H_A : p_1 - p_2 > 0 \ \ (or\ p_1 - p_2 < 0 \ \ or\ p_1 - p_2 \neq 0)$$

...then analysis is similar to the 'speed dating problem: what is the probability that the woman's height is larger than the man's. We are combining two different distributions.

If the original distributions are Normal, then the distribution of the difference is also Normal:

$$\mu_{diff} = \mu_1 - \mu_2$$

The variances add: (assuming distributions are independent - Pythagorean Theorem of Statistics)

$$Var_{diff} = Var_1 + Var_2$$
$$\sigma^2{}_{diff} = \sigma_1{}^2 + \sigma_2{}^2$$

Each distribution is a sampling distribution of a sample proportion (which should match the population means):

$$\mu_1 = p_1 = \frac{x_1}{n_1} \qquad \mu_2 = p_2 = \frac{x_2}{n_2}$$

If we knew the true population proportions, we could compute the standard deviations:

$$SD_1 = \sigma_1 = \sqrt{\frac{p_1 q_1}{n_1}} \qquad SD_2 = \sigma_2 = \sqrt{\frac{p_2 q_2}{n_2}}$$

But we only have estimates of the populations from samples, so we use the standard error:

$$SE_1 = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}} \qquad SE_2 = \sqrt{\frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Putting this all together:

$$H_0 : p_1 - p_2 = 0$$
$$H_A : p_1 - p_2 > 0 \ \ (or\ p_1 - p_2 < 0 \ \ or\ p_1 - p_2 \neq 0)$$

$$\mu_{diff} = \mu_1 - \mu_2 \qquad \sigma^2{}_{diff} = \sigma_1{}^2 + \sigma_2{}^2$$

$$\mu_{diff} = p_1 - p_2 \qquad SE^2{}_{diff} = SE_1{}^2 + SE_2{}^2$$

$$\mu_{diff} = \frac{x_1}{n_1} - \frac{x_2}{n_2} \qquad SE^2{}_{diff} = \left(\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}}\right)^2 + \left(\sqrt{\frac{\hat{p}_2 \hat{q}_2}{n_2}}\right)^2$$

$$SE_{diff} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$