Math 122
Introduction to Statistics
TWO PROPORTION Z TEST

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION

Recall that we have two versions of the Central Limit Theorem which relate to proportions.

Proportions 1: Suppose that x is the number of successes in simple random samples of size n selected from a population in which the proportion of success is p. If $np \geq 5$ and $n(1-p) \geq 5$, then the sample successes x are approximately normally distributed with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$.

Proportions 2: Suppose that simple random samples of size n are selected from a population in which the proportion of success is p. If $np \geq 5$ and $n(1-p) \geq 5$, then the sample proportion $\hat{p}$ is approximately normally distributed with mean $\mu = p$ and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{n}}$.

We used the first version to perform hypothesis tests about one proportion by focusing on the number x out of n independent trials. We could have used the second version with the same results (though computations by hand would have been slightly more complex). We will use the second variant to test claims about two proportions p_1 and p_2. Let $\hat{p}_1$ and $\hat{p}_2$ represent sample proportions approximating p_1 and p_2. If $\hat{p}_1$ and $\hat{p}_2$ are (approximately) normal, then so is $\hat{p}_1 - \hat{p}_2$. Moreover, we can express the mean and standard deviation of $\hat{p}_1 - \hat{p}_2$ in terms of the means and standard deviations of $\hat{p}_1$ and $\hat{p}_2$. Now, any claim about the relationship between p_1 and p_2 can be expressed as a claim about the difference $p_1 - p_2$. For example, the claim that $p_1 = p_2$ is equivalent to $p_1 - p_2 = 0$. Our approach, then, is to convert any claim about the relationship between p_1 and p_2 into a claim about $p_1 - p_2$ and then to perform a z test for this difference. Of course, the mechanics will be completely performed by technology.

CLAIMS ABOUT TWO PROPORTIONS

We will consider claims about two proportions of these forms:

$$p_1 = p_2 \text{ or } p_1 \leq p_2 \text{ or } p_1 \geq p_2 \text{ or } p_1 \neq p_2 \text{ or } p_1 < p_2 \text{ or } p_1 > p_2.$$

The Null Hypothesis H_0 and he Alternative Hypothesis H_1 will always be one of these

H_0	H_1
$p_1 \leq p_2$	$p_1 > p_2$
$p_1 = p_2$	$p_1 \neq p_2$
$p_1 \geq p_2$	$p_1 < p_2$

Example: A study addressed whether men and women prefer the color blue or the color red. One hundred men and one hundred women were asked the question, "Do you prefer blue over red?" Of the men questioned, 73 answered yes. Of the women, 61 answered yes. Treat these samples as simple random samples and use this data to test the claim that the proportion of men who prefer blue to red is greater than the proportion of women who prefer blue to red.

1. **Parameters:** Define these symbols:
   - p_1 is the proportion of men who prefer blue to red.
   - $n_1 = 100$
   - x_1 is the number of men in a simple random sample of 100 who prefer blue to red.
   - $\hat{p}_1 = x_1/n_1$
   - p_2 is the proportion of women who prefer blue to red.
   - $n_2 = 100$
   - x_2 is the number of women in a simple random sample of 100 who prefer blue to red.
   - $\hat{p}_2 = x_2/n_2$

   Our observed values of x_1 and x_2 are 73 and 61 respectively.
2. **Symbolic Claim:** The claim is $p_1 > p_2$.
3. **Opposite of the Claim:** The opposite of the claim is $p_1 \leq p_2$.
4. H_0 **and** H_1**:** H_0 is $p_1 \leq p_2$ and H_1 is $p_1 > p_2$.
5. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.
6. **Distribution and Test Statistic:** Our test statistic is a z-score associated to the difference between our sample proportions. Technology gives us a z-score of 1.8046.
7. P**-value:** Since H_1 involves $>$, this is a right-tailed test. Then

   $$P = P(z > 1.8046) = \mathit{normalcdf}(1.8046, 9) = 0.0356.$$
8. **Formal Conclusion:** Since $P < \alpha$, we reject H_0.
9. **Final Conclusion:** Since our claim is the same as H_1, we first rephrase our formal conclusion to refer to H_1. Since we are rejecting H_0, we are supporting H_1. This means that we are supporting our claim. Our conclusion is:

   > There is enough sample evidence to support the claim that the proportion of men who prefer blue to red is greater than the proportion of women who prefer blue to red.

Interpretation: A larger proportion of men than women seem to prefer blue to red.