Section 7.1 The Difference Between Two Means
Large Sample Hypothesis Tests

If $X \sim (\mu_x, \sigma_x^2)$ and $Y \sim (\mu_y, \sigma_y^2)$

If n is large (n_x, n_y)

CLT kicks in!

$\bar{X} \sim N(\mu_x, \frac{\sigma_x^2}{n_x})$ and $\bar{Y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$

How is X-Y distributed?

$$X - Y \sim (\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$$

Using CLT and the work above...

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

If I want to build a hypothesis test about how μ_x compares to μ_y, we can use the following test statistic:

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Example: **A test is conducted to compare the effectiveness of two cholesterol-lowering drugs.**

75 patients were given drug X and reduced cholesterol by an average of 40 with a standard deviation of 12.

100 patients were given drug Y and reduced cholesterol by an average of 48 with a standard deviation of 15.

Can we conclude the mean reduction using drug Y is greater than that of drug X.

$H_0: \mu_y \le \mu_x \rightarrow \mu_y - \mu_x \le 0$

$H_1: \mu_y > \mu_x \rightarrow \mu_y - \mu_x > 0$

$$z = \frac{(48-40) - 0}{\sqrt{\frac{15^2}{100} + \frac{12^2}{75}}} = 3.92$$

p-value: normcdf(3.92, L.N., 0, 1)

$\approx$ hella small

$\boxed{\text{Conclusion.}}$

Confidence Interval for the Difference Between Two Means

$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}$$

Section 7.2 The Difference Between Two Proportions

Confidence Interval for a Difference Between Proportions:

If $X \sim Bin(n_X, p_X)$ and $Y \sim Bin(n_Y, p_Y)$ and n in each case is large (both samples contain at least 10 successes and 10 failures) then the CI for $p_X - p_Y$ would be:

$$(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_x}(1-\widehat{p_x})}{n_x}+\frac{\widehat{p_y}(1-\widehat{p_y})}{n_y}}$$

Hypothesis Test for a Difference Between Proportions:

Note that $\hat{p}_X - \hat{p}_Y$ are distributed as follows (if n is large in each case):

$$\widehat{p_X} - \widehat{p_Y} \sim N\left(p_X - p_Y, \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right)$$

So in general we would think that for a hypothesis test for a difference in proportions, we would have the following test statistic:

$$z = \frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}}}$$

This test statistic will change however if H0: $p_X = p_Y$

$$Z = \frac{\hat{P}_X - \hat{P}_Y}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$$

But what is p? We are only assuming $p_X = p_Y$, not that they are equal to a particular value. We need $\hat{p}$.

$$\hat{p} = \frac{X+Y}{n_X + n_Y}$$

Example: Is cocaine deadlier than heroin? A study reported that rats with unlimited access to cocaine had poorer health, had more behavior disturbances, and died at a higher rate than did a corresponding group of rats given unlimited access to heroin. The death rates after 30 days on the study were as follows:

	% Dead at 30 days
Cocaine Group	90
Heroin Group	36

Suppose that 100 rats were in each group. Is there evidence to suggest that cocaine is deadlier than heroin?

$H_0: P_C \le P_H$

$H_1: P_C > P_H$

$\hat{p} = \frac{126}{200}$

$$Z = \frac{\hat{P}_C - \hat{P}_H}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{50}\right)}} = 7.909$$

p-value: HELLA small!

There is overwhelming evidence to suggest cocaine is deadlier than heroin.

Section 7.1 and 7.2 due Monday.