LARGE-SAMPLE TESTS/CI'S FOR $p_1 - p_2$ [DEVORE 9.4]

- **LARGE-SAMPLE z-TEST FOR $p_1 - p_2$:**

Populations:	Two Populations with proportions p_1, p_2 of some "success"
Realized Samples:	$\mathbf{x} := (x_1, x_2, \ldots, x_{n_1})$ $\mathbf{y} := (y_1, y_2, \ldots, y_{n_2})$ Samples $\mathbf{x}$ & $\mathbf{y}$ are independent of one another
Test Statistic Value $W(\mathbf{x}, \mathbf{y})$	$z = \dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}\widehat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\widehat{p}_1 := X/n_1,\ \widehat{p}_2 := Y/n_2$, $\widehat{p} := (X+Y)/(n_1+n_2)$, $\widehat{q} := 1 - \widehat{p}$
	$\min\{n_1\widehat{p}_1, n_1\widehat{q}_1, n_2\widehat{p}_2, n_2\widehat{q}_2\} \geq 10$

HYPOTHESIS TEST:	**P-VALUE DETERMINATION:**
$H_0: \ p_1 - p_2 = 0$ vs. $H_A: \ p_1 - p_2 > 0$	P-value $\approx 1 - \Phi(z)$
$H_0: \ p_1 - p_2 = 0$ vs. $H_A: \ p_1 - p_2 < 0$	P-value $\approx \Phi(z)$
$H_0: \ p_1 - p_2 = 0$ vs. $H_A: \ p_1 - p_2 \neq 0$	P-value $\approx 2 \cdot [1 - \Phi(\lvert z \rvert)]$

Decision Rule:
If P-value $\leq \alpha$ then reject H_0 in favor of H_A
If P-value $> \alpha$ then accept H_0 (i.e. fail to reject H_0)

- **LARGE-SAMPLE z-CI FOR $p_1 - p_2$:**

Given any two populations with proportions p_1 and p_2 of some "success".

Let $x_1, x_2, \ldots, x_{n_1}$ be a sample taken from the 1^{st} population.

Let $y_1, y_2, \ldots, y_{n_2}$ be a sample taken from the 2^{nd} population.

Then the $100(1-\alpha)\%$ **large-sample CI for** $p_1 - p_2$ is approximately

$$\left((\widehat{p}_1 - \widehat{p}_2) - z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}}, \quad (\widehat{p}_1 - \widehat{p}_2) + z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}} \right)$$

— OR WRITTEN MORE COMPACTLY —

$$(\widehat{p}_1 - \widehat{p}_2) \pm z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}}$$

where $\widehat{p}_1 := X/n_1$, $\widehat{p}_2 := Y/n_2$, $\widehat{q}_1 := 1 - \widehat{p}_1$, $\widehat{q}_2 := 1 - \widehat{p}_2$, $\min\{n_1\widehat{p}_1, n_1\widehat{q}_1, n_2\widehat{p}_2, n_2\widehat{q}_2\} \geq 10$

EX 9.4.1: The manufacturer of a sore throat medication has developed a new formulation that it claims to be more effective.

To evaluate the new medication, a **test group** of 120 current users try the new formulation.

After a four-week trial, 108 users from the test group indicated the new formulation was more effective.

Simultaneously, a **control group** of 150 users were given the current formulation but told it was the new one.

After a four-week trial, 114 users from the control group indicated it was an improvement.

Assume this is a **randomized clinical trial**, which means there was no bias/dishonesty/cheating involved.

Does the data suggest that the new formulation is more effective at treating a sore throat??

(Use significance level $\alpha = 0.05$)

(a) State the appropriate null hypothesis H_0 & alternative hypothesis H_A.

(b) Compute the appropriate test statistic value for this hypothesis test.

(c) Compute the resulting P-value.

(d) Make the appropriate decision.

(e) Construct the approximate two-sided 90% large-sample CI for $p_1 - p_2$.