Hypothesis Test - Difference in Independent Population Means $\mu_x - \mu_y$

I) Small Sample Case ($n \leq 20$ and $m \leq 20$)

EX: A researcher wishes to assess a "new" teaching method for "slow learners." A random sample of 8 students use the new method, and a random sample of 12 students use the "standard" teaching method. After 6 months, an exam is administered to each student. Does the data indicate that the new teaching method is
preferable?  ($\alpha = .05$)

Exam Scores
New Method (GROUP 1)
80  76  80  66  79  79  81  76
$\bar{X} = 77.125$
$S_x = 4.853$  \hspace{1cm} n = 8

Standard Method (GROUP 2)
79  73  72  62  76  68
70  86  75  68  73  66
$\bar{Y} = 72.333$
$S_y = 6.344$  \hspace{1cm} M = 12
1) \( H_0: \mu_x - \mu_y = 0 \)

2) \( H_1: \mu_x - \mu_y > 0 \) \((\mu_x > \mu_y)\)

3) \( \alpha = .05 \)

4) Test Statistic

\[
T = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}
\]

where

\[
s_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2}}
\]
Here

\[ s_p = \sqrt{\frac{(8-1)4.853^2 + (12-1)6.344^2}{8 + 12 - 2}} \]

= \frac{\sqrt{33.754}}{}

= 5.81

So

\[ T = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \]

= \frac{77.125 - 72.333}{5.81 \sqrt{\frac{1}{8} + \frac{1}{12}}}
5) p-value

Find prob. of getting 1.825 or larger for value of test stat., assuming H0 true. We will examine values from t-dist with df = n+m-2 = 18 d.f.
Since Ha: \( \mu_x - \mu_y > 0 \) we will try to find \( P(T \geq 1.825) \).

Looking for 1.825 along the row for 18 d.f. we see \( P(T \geq 1.73) = 0.05 \)

Since 1.825 > 1.73

Then \( P(T \geq 1.825) < 0.05 \)
6) Decision
   Reject $H_0$ if $p$-value $\leq \alpha$
   Since we found $p$-value $< 0.05$
in step 5, we reject $H_0$

7) Conclude:
   At the 5% significance level, the evidence indicates that the new teaching method produces higher mean exam score than the standard teaching method.

Note: 1) The data from each pop. must be normally dist.

2) The pop. std. dev. must be equal, $\sigma_1 = \sigma_2$
Hypothesis Test - Difference in Independent Pop. Means $\mu_x - \mu_y$

II) Large Sample Case ($n \geq 20$ and $m \geq 20$)

Ex: A college statistics professor conjectures that students with good high school math background (2 or more math courses) perform better in college statistics course than students with a poor high school math background (1 or fewer math courses). He randomly selects 35 students with good math background, and 45 students with poor math background, and records Exam 1 scores from a college statistics course. Test the hypothesis that the mean score of the "good background" students
will be higher than the mean score of "poor math background" students (use $\alpha = 0.10$). Summary data follows:

Two or more math courses:
$n = 35 \quad \bar{X} = 84.2 \quad S_x = 10.2$

One or fewer math courses:
$m = 45 \quad \bar{Y} = 73.1 \quad S_y = 14.3$

1) $H_0: \mu_x - \mu_y = 0$

2) $H_a: \mu_x - \mu_y > 0$

3) $\alpha = 0.10$

4) Test Statistic

\[ Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \]
where \( \mu_x - \mu_y \) is the hypothesized difference in Ho.

\[
T = \frac{(84.2 - 73.1) - 0}{\sqrt{\frac{10.2^2}{35} + \frac{14.3^2}{45}}}
\]

\[
= \frac{11.1}{\sqrt{7.51679}}
\]

\[
= \frac{11.1}{2.74168}
\]

\[
= 4.0486
\]

5) *p*-value

We must find probability of getting 4.05 or larger value for the test.
statistic (since \( H_A: \mu_X - \mu_Y > 0 \)), under the assumption that \( H_0 \) is true:

\[
P(T \geq 4.05) \approx
\]

\[
P(Z \geq 4.05) = 1 - P(Z < 4.05)
\]

\[
\approx 1 - 1
\]

\[
\approx 0
\]

So it is highly unlikely we would get a value 4.05 or larger for test statistic if \( H_0 \) true.

This provides evidence that \( H_0 \) is likely false.

6) Decision

Reject \( H_0 \) if \( p\)-value \( \leq \alpha \)

Since \( \alpha < 0.10 \)

we reject \( H_0 \)
7) Conclusion
At the 10% significance level there is evidence that students with good high school math background have a higher mean score on Exam 1 than students with poor high school math background.

Assumption:
Independent random samples from the populations