Hypothesis Testing with Two Samples

MAT 152 – STATISTICAL METHODS I

LECTURE 2

INSTRUCTOR: DUSTIN ROTEN

FALL 2020

Two-Sample Hypothesis Testing

If two groups are independent simple random samples from two distinct populations:

1. If the sample sizes are small, the distributions should be normal
2. If the sample sizes are large, the distributions are not important

(The test comparison of two independent population means with unknown and **possibly unequal population standard deviations** is called the Aspen-Welch t-test.)

Differences depend on sample means and sample standard deviations. Very different means can occur if the variations are large.

Null and alternative hypotheses are still constructed.

Example

A study is done by a community group in two neighboring colleges to determine which one graduates students with more math classes. **College A** samples 11 graduates. Their average is 4 math classes with a standard deviation of 1.5 math classes. **College B** samples 9 graduates. Their average is 3.5 math classes with a standard deviation of 1 math class. The community group believes that a student who graduates from **College A** has taken more math classes, on average. Both populations have a normal distribution. Test at the 1% significance level.

Q1: Is this a test of two means or two proportions?

Two means

Q2: Are the population standard deviations known or unknown?

Unknown

Q3: Which distribution should be used?

Student's *t*

Statistic	College A	College B
n	11	9
$\bar{x}$	4	3.5
s	1.5	1
Std. Err.	$\frac{s_A}{\sqrt{n}} = \frac{1.5}{\sqrt{11}} \approx 0.452$	$\frac{s_B}{\sqrt{n}} = \frac{1}{\sqrt{9}} \approx 0.\overline{3}$
ν	10	8

Example (Cont.)

A study is done by a community group in two neighboring colleges to determine which one graduates students with more math classes. **College A** samples 11 graduates. Their average is 4 math classes with a standard deviation of 1.5 math classes. **College B** samples 9 graduates. Their average is 3.5 math classes with a standard deviation of 1 math class. The community group believes that a student who graduates from College A has taken more math classes, on average. Both populations have a normal distribution. Test at the 1% significance level.

How do we construct H_0 and H_a for two means?

Q4: What is the random variable?

Look at the differences in the two means: $\bar{X}_A - \bar{X}_B$

Q5: What are the null and alternative hypotheses?

$H_0: \mu_A \leq \mu_B$

$H_a: \mu_A > \mu_B$

Q6: Will this be a right-, left-, or two-tailed test?

Right-tailed test

Let's rewrite our hypotheses:

$H_0: \mu_A \leq \mu_B$	$H_0: \mu_A - \mu_B \leq 0$
$H_a: \mu_A > \mu_B$	$H_a: \mu_A - \mu_B > 0$

We are interested in the differences:

$H_0: \Delta\mu \leq 0$

(The difference of the means is less than zero; μ_A is smaller)

$H_a: \Delta\mu > 0$

(The difference of the means is greater than zero; μ_A is larger)

Example (Cont.)

We are now ready to test! We assume that H_0 is true. (No difference in the mean $\Delta\mu = 0$).

Assuming the null hypothesis to be true, we need to determine the t-score of our sample mean. In our case: $\bar{x}_A - \bar{x}_B = \Delta\bar{x} = 4 - 3.5 = 0.5$.

- Normally, $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ where $\frac{s}{\sqrt{n}}$ is the standard error (SE). The variance would be $SE^2 = \left(\frac{s}{\sqrt{n}}\right)^2 = \frac{s^2}{n}$.
- We now have two sample variances that we must combine!

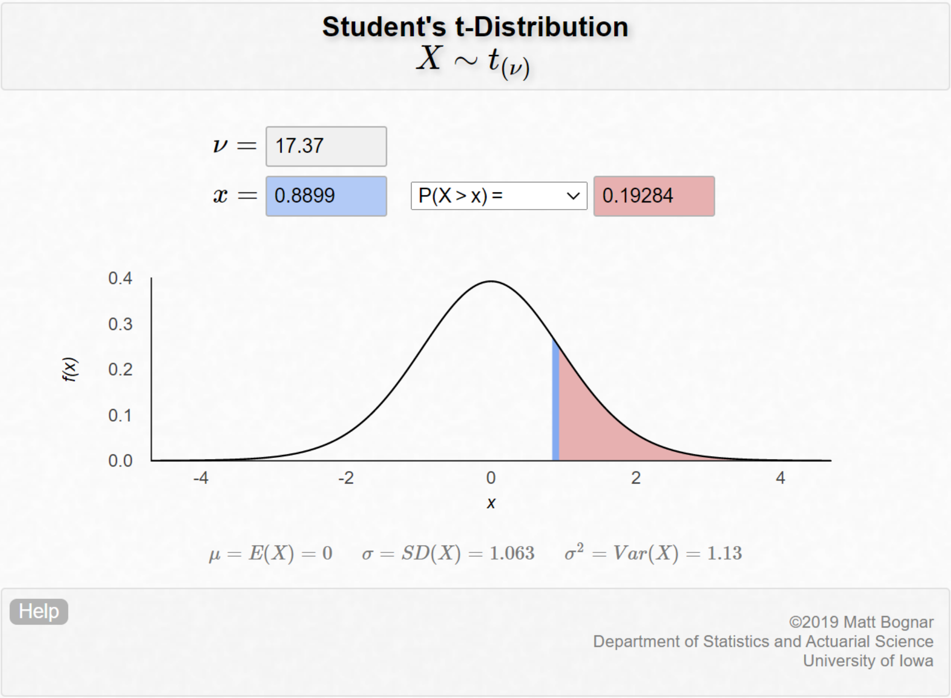
$V_A + V_B = \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$. We can take the square root of this combined variance to get our combined standard error:

$$SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

Example (Cont.)

Q7: What is the t-score of our value? (Assuming $\Delta\mu = 0$)

$$t = \frac{\Delta\bar{x}-\Delta\mu}{SE} = \frac{\Delta\bar{x}-\Delta\mu}{\sqrt{SE_A^2+SE_B^2}} = \frac{0.5-0}{\sqrt{0.452267^2+0.333333^2}} \approx 0.8899$$



Q8: How many degrees of freedom are there? Aspin-Welch approximates ν as follows:

$$\nu = df = \frac{\left(SE_A^2+SE_B^2\right)^2}{\left(\frac{SE_A^4}{\nu_A}\right)+\left(\frac{SE_B^4}{\nu_B}\right)}$$

$$\nu = df = \frac{\left(SE_A^2+SE_B^2\right)^2}{\left(\frac{SE_A^4}{\nu_A}\right)+\left(\frac{SE_B^4}{\nu_B}\right)} = \frac{\left(0.45^2+0.33^2\right)^2}{\left(\frac{0.45^4}{10}\right)+\left(\frac{0.33^4}{8}\right)} \approx 17.37$$

Example (Cont.)

A study is done by a community group in two neighboring colleges to determine which one graduates students with more math classes. **College A** samples 11 graduates. Their average is 4 math classes with a standard deviation of 1.5 math classes. **College B** samples 9 graduates. Their average is 3.5 math classes with a standard deviation of 1 math class. The community group believes that a student who graduates from **College A** has taken more math classes, on average. Both populations have a normal distribution. Test at the 1% significance level.

We now know that $P(\Delta\bar{x} > 0.5) = 0.19284$ or roughly 19.28%.

At the 1% significance level: $\alpha = 0.01$ and $\alpha < p - value$.

At the 1% significance level, there is NOT sufficient evidence provided by the samples to conclude that a student who graduates from College A has taken more math classes, on average, than a student who graduates from College B.

Known Population Standard Deviations

In the previous problem, we did not know the populations' standard deviations. In rare cases where this is known a normal distribution can be used.

$$\Delta\bar{X} \sim N\left[\Delta\mu, \sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}\right]$$

Standard Deviation: $\sigma = \sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$

Z-score: $\mathrm{z} = \frac{\Delta\bar{x} - \Delta\mu}{\sigma} = \frac{\Delta\bar{x} - \Delta\mu}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}}$

Review

Two samples from two populations can be compared. This is done by comparing the differences between their means: μ_1 and μ_2.

The same process applies:

1) Assume the null hypothesis is true.

2) Calculate the probability of the random sample occurring. (t-test or z-test)

$$t = \frac{\Delta\bar{x} - \Delta\mu}{SE} = \frac{\Delta\bar{x} - \Delta\mu}{\sqrt{SE_A^2 + SE_B^2}}$$

$$\nu = df = \frac{\left(SE_A^2 + SE_B^2\right)^2}{\left(\frac{SE_A^4}{\nu_A}\right) + \left(\frac{SE_B^4}{\nu_B}\right)}$$