Mathematics 231

Lecture 26

Liam O'Brien

Announcements

- Reading
  - Today M&M 7.2 447-467
  - Next class M&M 7.1 428-435

    M&M 7.3 474-477

Topics

- Hypothesis testing for comparing two means

Comparison of Two Population Means

- So far, we have considered the comparison of the mean of a single population to some null value, μ_0.
- However, many times we want to compare the means from two different populations, μ_1 and μ_2.
- Example: Comparison of means for a treatment and control group; developed vs. developing countries, etc.
- Ordinarily, we want to know if μ_1 and μ_2 are equal.

Comparison of Two Population Means

- Given SRSs from the two populations, μ_1 and μ_2 can be estimated by their respective sample means.
- **Question:** Is the difference in sample means so large that it is unlikely to have occurred by chance alone?
- To answer this, the form of the test statistic depends on how the data were collected:
  1. Independent samples
  2. Paired samples

Independent Samples

- The two underlying populations of interest are independent.
- The population distributions are assumed to be normal.
- Given SRSs of size n_1 from population 1, and n_2 from population 2, we want to test:

  $H_0: \mu_1 = \mu_2$ against $H_A: \mu_1 \neq \mu_2$
- If the two population means are identical, we would expect the sample means to be relatively close to each other.

Independent Samples

We would want to reject $H_0 : \mu_1 = \mu_2$ if $\bar{x}_1$ and $\bar{x}_2$ are too far apart, or eqivalently, if $\bar{x}_1 - \bar{x}_2$ is far from 0.

Note: The standard deviations of the two populations σ_1 and σ_2 may or may not be equal.

Need to consider two cases:

(1) Equal standard deviations: $\sigma_1 = \sigma_2 = \sigma$

(2) Unequal standard deviations: $\sigma_1 \neq \sigma_2$

Independent Samples: Equal Sds

To evaluate $H_0 : \mu_1 = \mu_2$, use test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

is a pooled, or combined, estimate of σ^2.

Independent Samples: Equal Sds

The pooled estimate $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

combines information from both samples to produce a better estimate of σ^2.

This is sensible since s_1^2 and s_2^2 estimate the same thing.

Independent Samples: Equal Sds

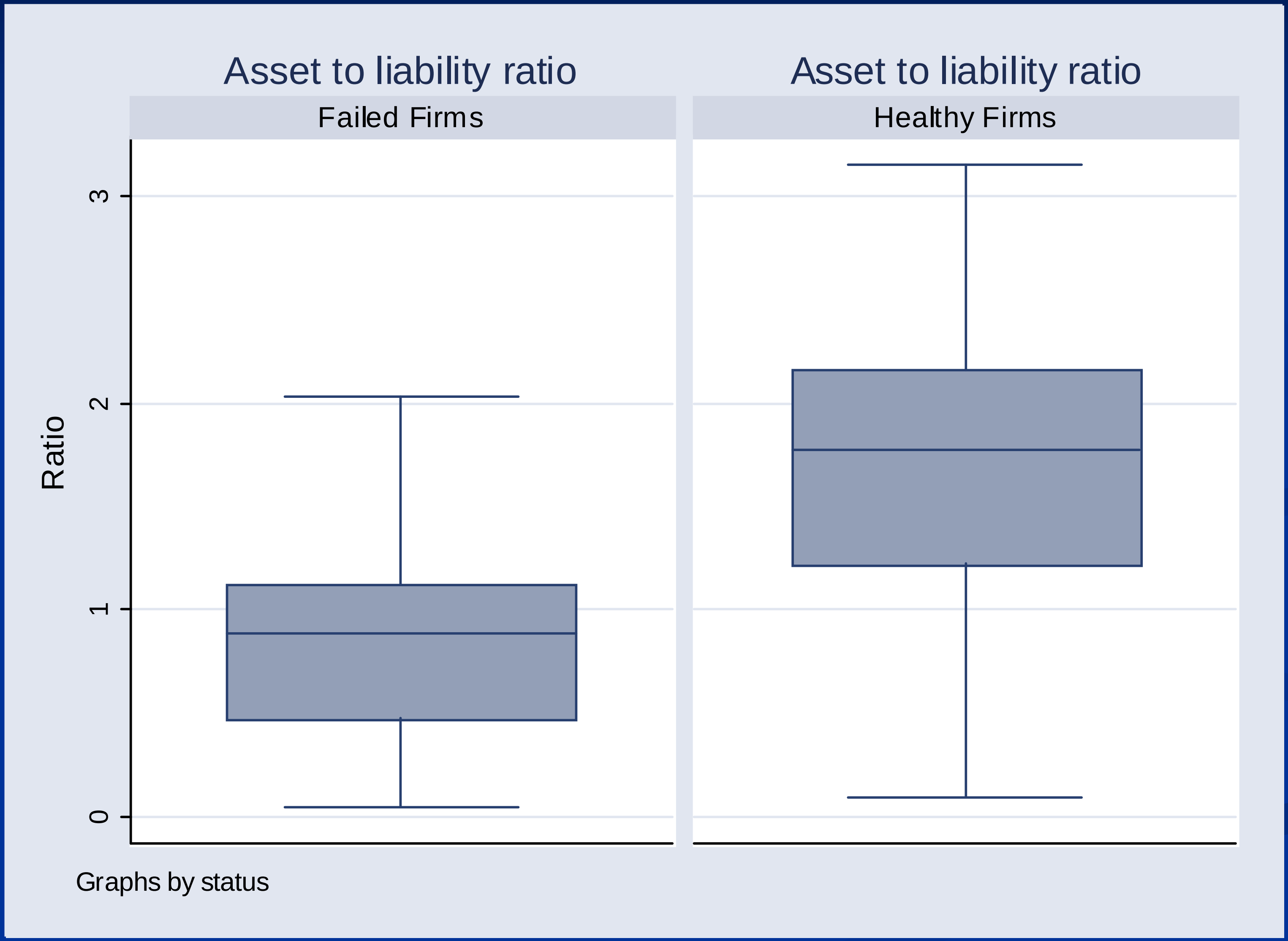
Under $\mathrm{H}_0 : \mu_1 = \mu_2$, the test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

has a t-distribution with $n_1 + n_2 - 2$ df.

This is called a two-sample t-test.

Example: Assets and Liabilities



Example: Assets and Liabilities

Among the healty firms:

$$n_1 = 68$$

$$\bar{x}_1 = 1.73$$

$$s_1 = 0.639$$

Among the failed firms:

$$n_2 = 33$$

$$\bar{x}_2 = 0.824$$

$$s_2 = 0.481$$

Two-Sample T-Test in Stata

- If you don't have a dataset click on **Statistics > Summaries, Tables & Tests > Classical Tests of Hypotheses > Two-sample mean comparison calculator**
- Enter the sample sizes, sample means, sample sd's, and select whether the variances are assumed to be equal or not.

Two-Sample T-Test in Stata

```
. ttesti 68 1.72 0.639 33 0.824 0.481

Two-sample t test with equal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      68        1.72    .0774901        .639    1.565329    1.874671
       y |      33        .824    .0837314        .481    .6534448    .9945552
---------+--------------------------------------------------------------------
combined |     101    1.427248    .0721645     .725244    1.284075     1.57042
---------+--------------------------------------------------------------------
    diff |                .896    .1257124                .6465593    1.145441
------------------------------------------------------------------------------
Degrees of freedom: 99

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
       t =   7.1274                t =   7.1274                t =   7.1274
   P < t =   1.0000          P > |t| =   0.0000            P > t =   0.0000
```

Example: Assets and Liabilities

- Since $p < 0.001$, we can reject H_0 in favor of H_A.
- Note: We can construct a 95% CI for $\mu_1 - \mu_2$.

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = (0.657, 1.15)$$

where t^* is a t-multipler with $n_1 + n_2 - 2$ df.

Example: Assets and Liabilities

- The test statistic is,

$$s_P^2 = \frac{(68-1)(0.639)^2 + (33-1)(0.481)^2}{68+33-2} = 0.352$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{1.72 - 0.824}{\sqrt{0.352\left(\frac{1}{68} + \frac{1}{33}\right)}} = 7.12$$

Independent Samples: Unequal SDs

To evaluate $H_0: \mu_1 = \mu_2$ use the test statistic,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

There is no common estimate for the standard deviation.

Independent Samples: Unequal SDs

Under $H_0: \mu_1 = \mu_2$ the test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$ has a t-distribution with

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$ degrees of freedom.

Independent Samples: Unequal SDs

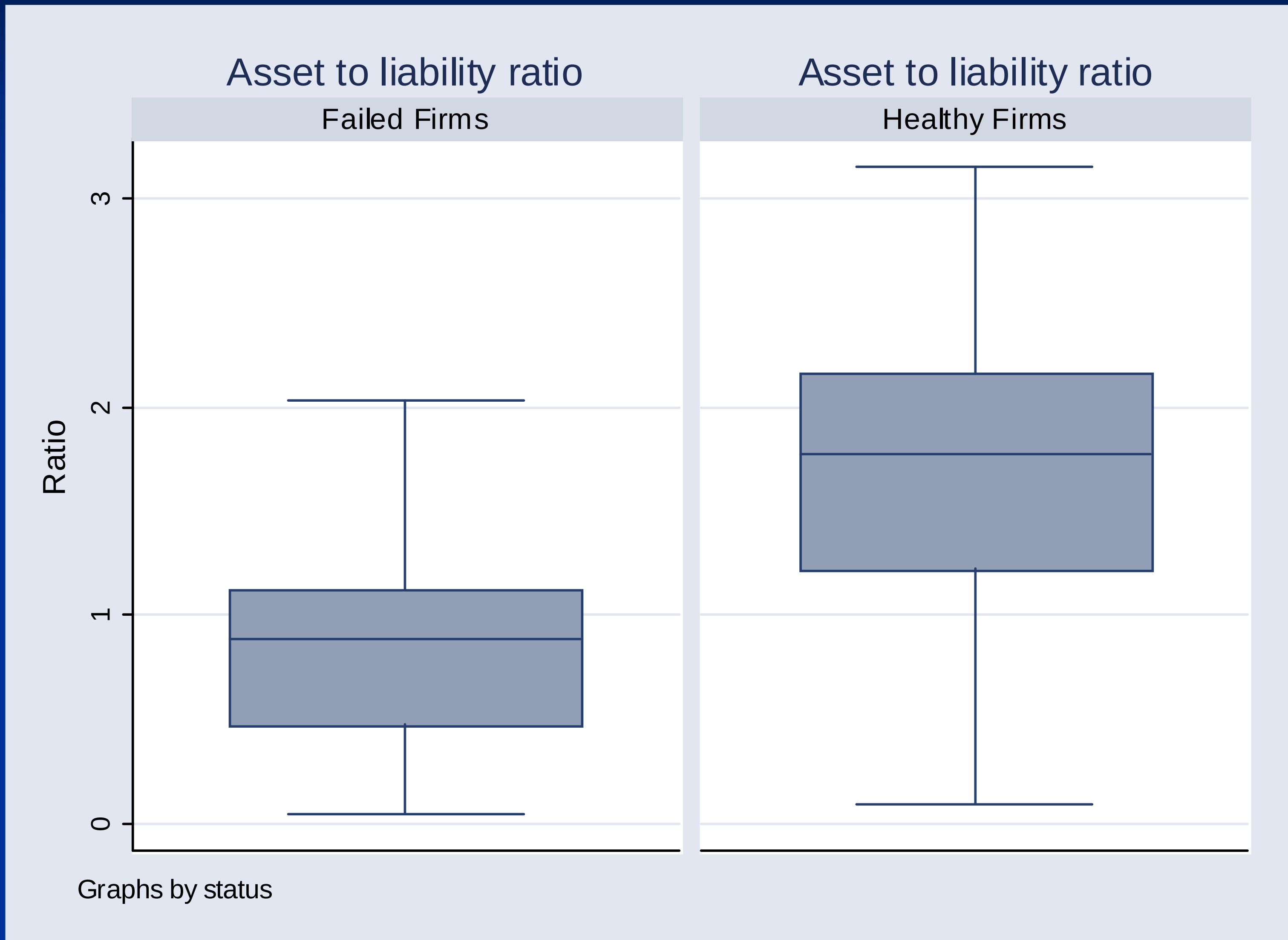
A (1-α)% CI is given by,

$$(\bar{x}_1 - \bar{x}_2) \pm t_k^* \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

with

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$ degrees of freedom.

Example: Assets and Liabilities



Example: Assets and Liabilities

- Since $p < 0.001$, we can reject H_0 in favor of H_A.
- Note: We can construct a 95% CI for $\mu_1 - \mu_2$.

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)} = (0.670, 1.12)$$

where t^* is a t-multiplier with k df.

Example: Assets and Liabilities

- The test statistic is,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1.72 - 0.824}{\sqrt{\frac{0.639^2}{68} + \frac{0.481^2}{33}}} = 7.86$$

Two-Sample T-Test in Stata

```
. ttesti 68 1.72 0.639 33 0.824 0.481, unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      68        1.72    .0774901        .639    1.565329    1.874671
       y |      33        .824    .0837314        .481    .6534448    .9945552
---------+--------------------------------------------------------------------
combined |     101    1.427248    .0721645     .725244    1.284075     1.57042
---------+--------------------------------------------------------------------
    diff |                .896    .1140862                .6690326    1.122967
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  81.6737

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
       t =   7.8537                 t =   7.8537                  t =   7.8537
   P < t =   1.0000          P > |t| =   0.0000            P > t =   0.0000
```