Lecture 7: Review on two sample problems

Two Large Independent Samples

Is there a difference in the average daily intakes of dairy products for men versus women? In order to answer the question, a researcher randomly selected two large samples form men and women, respectively. The results are summarized in the following

Avg Daily Intakes	Men	Women
Sample size	50	50
Sample mean	756	762
Sample Std Dev	35	30

Two Small Independent Samples

Example: Calcium and Blood Pressure

Does increasing the amount of calcium in our diet reduce blood pressure? Researchers designed a randomized comparative experiment. The subjects were 21 healthy black men who volunteered to take part in the experiment. They were randomly assigned to two groups: 10 of the men received a calcium supplement for 12 weeks, while the control group of 11 men received a placebo pill that looked identical. The experiment was double-blind. The response variable is the decrease in systolic (top number) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase appears as a negative response Here are the data:

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

Two Dependent samples (paired data)

Example: A pharmaceutical company claims that its medicine helps reduce the blood pressure. The table shows the blood pressures of 5 patients before and after taking the medicine.

Member	**1**	**2**	**3**	**4**	**5**
Before	120	150	140	135	130
After	125	130	130	140	120

At $\alpha = .05$, is there enough evidence to support the company's claim?

Large Sample Test for Difference Between Two Population Means

A random sample of size n_1 drawn from population 1 with mean μ_1 and variance σ_1^2.

A random sample of size n_2 drawn from population 2 with mean μ_2 and variance σ_2^2.

Large Sample Test for Difference Between Two Population Means

Step 1: Specify the null and alternative hypothesis

- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 \neq D_0$**
  (two-tailed test)
- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 > D_0$**
  (right-tailed test)
- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 < D_0$**
  (left-tailed test)

where D_0 is some specified difference that you wish to test. $D_0 = 0$ when testing no difference.

Large Sample Test for Difference Between Two Population Means

Step 2: Test statistic for large sample sizes when $n_1 \geq 30$ and $n_2 \geq 30$

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}}$$

Step 3: Under H_0, the sampling distribution of z is approximately standard normal

Large Sample Test for Difference Between Two Population Means

Step 3: Find p-value. Compute

$$z^* = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

– $\mathbf{H_a}$: $\mu_1 - \mu_2 \neq D_0$ (two-tailed test) $\text{p-value} = 2P(z > |z^*|)$

– $\mathbf{H_a}$: $\mu_1 - \mu_2 > D_0$ (right-tailed test) $\text{p-value} = P(z > z^*)$

– $\mathbf{H_a}$: $\mu_1 - \mu_2 < D_0$ (left-tailed test) $\text{p-value} = P(z < z^*)$

Example

Avg Daily Intakes	Men	Women
Sample size	50	50
Sample mean	756	762
Sample Std Dev	35	30

- Is there a difference in the average daily intakes of dairy products for men versus women? Use $\alpha = .05$.

Small Sample Testing the Difference between Two Population Means

A random sample of size n_1 drawn from population 1 with normal distribution with mean μ_1 and variance σ_1^2.

A random sample of size n_2 drawn from population 2 with normal distribution with mean μ_2 and variance σ_2^2.

Assumption: Note that both population are normally distributed with the same population variances.

Small Sample Testing the Difference between Two Population Means

Step 1: Specify the null and alternative hypothesis

- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ **versus** $\mathbf{H_a: \mu_1 - \mu_2 \neq D_0}$ (two-tailed test)
- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ **versus** $\mathbf{H_a: \mu_1 - \mu_2 > D_0}$ (right-tailed test)
- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ **versus** $\mathbf{H_a: \mu_1 - \mu_2 < D_0}$ (left-tailed test)

where D_0 is some specified difference that you wish to test. $D_0=0$ when testing no difference.

Small Sample Testing the Difference between Two Population Means

Step 2: Test statistic for small sample sizes

$$t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where s^2 is calculated as $s^2 = \dfrac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}$

Step 3: Under H_0, the sampling distribution of t has a Student's t distribution with n_1+n_2-2 degrees of freedom

Small Sample Testing the Difference between Two Population Means

Step 3: Find p-value. Compute

$$t^* = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- $\mathbf{H_a: \mu_1 - \mu_2 \neq D_0}$ (two-tailed test) $\boxed{\text{p-value} = 2P(t > |t^*|)}$
- $\mathbf{H_a: \mu_1 - \mu_2 > D_0}$ (right-tailed test) $\boxed{\text{p-value} = P(t > t^*)}$
- $\mathbf{H_a: \mu_1 - \mu_2 < D_0}$ (left-tailed test) $\boxed{\text{p-value} = P(t < t^*)}$

Example

Two training procedures are compared by measuring the time that it takes trainees to assemble a device. A different group of trainees are taught using each method. Is there a difference in the two methods? Use $\alpha = .01$. Assume both populations are normally distributed with equal variances.

Time to Assemble	**Method 1**	**Method 2**
Sample size	10	12
Sample mean	35	31
Sample Std Dev	4.9	4.5

Solution:

The Paired-Difference Test

- We have assumed that samples from two populations are independent. Sometimes the assumption of independent samples is intentionally violated, resulting in a **matched-pairs** or **paired-difference test**.
- By designing the experiment in this way, we can eliminate unwanted variability in the experiment
- Denote data as

Pair	1	2	…	n
Population 1	x_{11}	x_{12}	…	x_{1n}
Population 2	x_{21}	x_{22}	…	x_{2n}
Difference	$d_1 = x_{11} - x_{21}$	$d_2 = x_{12} - x_{22}$	…	$d_n = x_{1n} - x_{2n}$

The Paired-Difference Test

Step 1: Specify the null and alternative hypothesis

- H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 \neq D_0$ (two-tailed test)
- H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 > D_0$ (right-tailed test)
- H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 < D_0$ (left-tailed test)

where D_0 is some specified difference that you wish to test. $D_0=0$ when testing no difference.

The Paired-Difference Test

Step 2: Test statistic for small sample sizes

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

where $\bar{d}$ and s_d are sample mean and standard deviation of the paired differences $d_1, \cdots, d_n$

Step 3: Under H_0, the sampling distribution of t has a Student's t distribution with n-1 degrees of freedom

The Paired-Difference Test

Step 3: Find p-value. Compute

$$t^* = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

- **H_a: $\mu_1 - \mu_2 \neq D_0$** (two-tailed test) $\text{p-value} = 2P(t > |t^*|)$
- **H_a: $\mu_1 - \mu_2 > D_0$** (right-tailed test) $\text{p-value} = P(t > t^*)$
- **H_a: $\mu_1 - \mu_2 < D_0$** (left-tailed test) $\text{p-value} = P(t < t^*)$

Example: A pharmaceutical company claims that its medicine helps reduce the blood pressure. The table shows the blood pressures of 5 patients before and after taking the medicine.

Member	**1**	**2**	**3**	**4**	**5**
Before	120	150	140	135	130
After	125	130	130	140	120

At $\alpha = .05$, is there enough evidence to support the company's claim?

Solution:

F-Distribution

Let s_1^2 and s_2^2 represent the sample variances of two different populations. If both populations are normal and the population variances σ_1^2 and σ_2^2 are equal, then the sampling distribution of

$$F = \frac{s_1^2}{s_2^2}$$

is called an ***F*-distribution**.

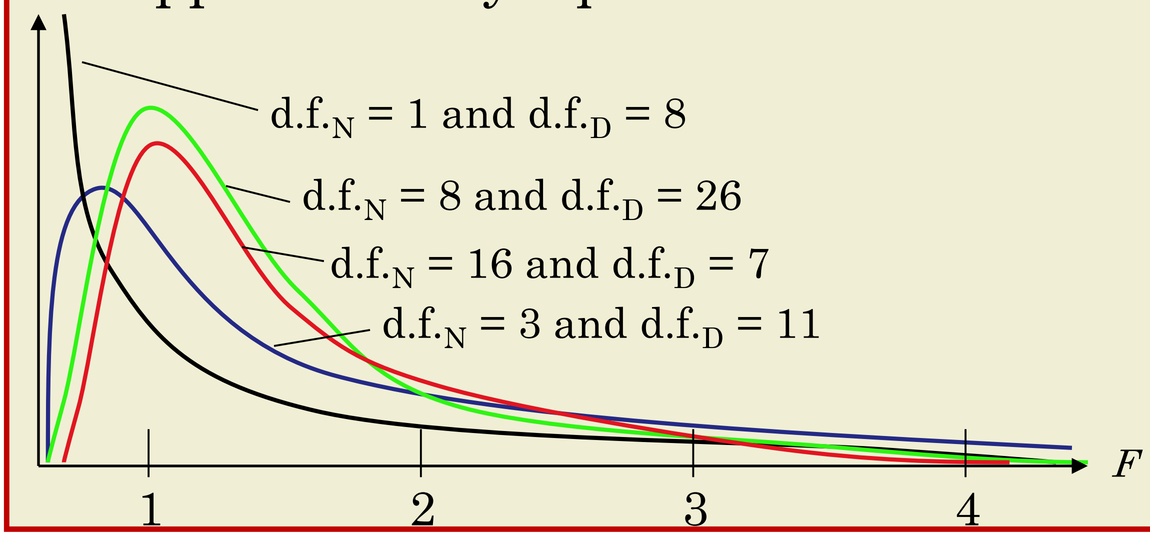
There are several properties of this distribution.

1. The F-distribution is a family of curves each of which is determined by two types of degrees of freedom: the degrees of freedom corresponding to the variance in the numerator, denoted **d.f._N**, and the degrees of freedom corresponding to the variance in the denominator, denoted **d.f._D**.

F-Distribution

Properties of the *F*-distribution continued:

2. *F*-distributions are positively skewed.
3. The total area under each curve of an *F*-distribution is equal to 1.
4. *F*-values are always greater than or equal to 0.
5. For all *F*-distributions, the mean value of *F* is approximately equal to 1.



Comparing Two Population Variances

How do we know whether $\sigma_x = \sigma_y$? Apparently, the following hypothesis needs to be tested

$$H_0 : \sigma_x^2 = \sigma_y^2$$

- Test statistic

$$F = \frac{s_x^2}{s_y^2}$$

Comparing Two Population Variances

- Finally, under the null hypothesis $\sigma_x^2 = \sigma_y^2$, one has

$$F = \frac{s_x^2}{s_y^2} \sim F_{(n-1),(m-1)}$$

- The calculation for P-value

$$\text{P-value} = \begin{cases} P(F > F_{obs}) & \text{for } H_a : \sigma_x^2 > \sigma_y^2 \\ P(F < F_{obs}) & \text{for } H_a : \sigma_x^2 < \sigma_y^2 \\ 2P\left(F > \frac{\max(s_x^2, s_y^2)}{\min(s_x^2, s_y^2)}\right) & \text{for } H_a : \sigma_x^2 \neq \sigma_y^2 \end{cases}$$

Example

Two training procedures are compared by measuring the time that it takes trainees to assemble a device. A different group of trainees are taught using each method. Is there a difference in the two methods? Use $\alpha = .01$. Assume both populations are normally distributed with **equal variances**.

Time to Assemble	Method 1	Method 2
Sample size	10	12
Sample mean	35	31
Sample Std Dev	4.9	4.5

Solution: