



Lecture 07

Statistics II

STAT-221 • UG

Dr. Rizwan Jahangir

NUST Business School

National University of Sciences & Technology

Contents

- **1. Welcome, Review & Assignment 1:**
 - ▶ Independent vs. Dependent Samples.
 - ▶ Pop-up Concept Check.
- **2. Tests of Hypothesis: Dependent Samples (Chapter 11):**
 - ▶ The "Trick" of the Difference (d).
 - ▶ Formulas for Dependent Samples.
 - ▶ Example 1: SAT Prep Course.
 - ▶ Example 2: The Schadek vs. Bowyer Revelation.
- **3. Break (15 Minutes)**
- **4. Two-Sample Tests about Proportions (Chapter 15):**
 - ▶ The Pooled Proportion (p_c).
 - ▶ The z-Statistic for Proportions.
 - ▶ Example 3: Fear of Flying (Men vs. Women).
- **5. Summary & Next Steps**



Welcome, Review

Review: Independent vs. Dependent

In Lecture 6, we learned how to compare the means of two populations.

- **Independent Samples:** The samples are drawn from two completely unrelated groups (e.g., comparing the fuel efficiency of Hondas vs. Toyotas).



Review: Independent vs. Dependent

In Lecture 6, we learned how to compare the means of two populations.

- **Independent Samples:** The samples are drawn from two completely unrelated groups (e.g., comparing the fuel efficiency of Hondas vs. Toyotas).
- We used the **Pooled Variance** (s_p^2) formula to run our t -test.



Review: Independent vs. Dependent

In Lecture 6, we learned how to compare the means of two populations.

- **Independent Samples:** The samples are drawn from two completely unrelated groups (e.g., comparing the fuel efficiency of Hondas vs. Toyotas).
- We used the **Pooled Variance** (s_p^2) formula to run our t -test.

Today's Focus: Dependent Samples

Sometimes, the two samples are explicitly linked. They are paired or matched. Most often, it is the **exact same person or object** measured in a "before and after" study.



Pop-Up Question 1: Concept Check

Discussion

Pop-Up Question!

Determine if the following business scenarios require an Independent or Dependent two-sample test:

1. You compare the average daily sales of the Islamabad branch vs. the Lahore branch.
2. You track the typing speed of 10 administrative assistants, send them to a training workshop, and track their typing speed again to see if they improved.



Pop-Up Question 1: Concept Check

Discussion

Pop-Up Question!

Determine if the following business scenarios require an Independent or Dependent two-sample test:

1. You compare the average daily sales of the Islamabad branch vs. the Lahore branch.
2. You track the typing speed of 10 administrative assistants, send them to a training workshop, and track their typing speed again to see if they improved.

Answers:

- **1: Independent.** The two branches are completely separate entities.
- **2: Dependent.** The data points are perfectly paired to specific individuals (before and after the workshop).



Tests of Hypothesis: Dependent Samples

The "Trick" of Dependent Samples

Why do we separate Dependent samples from Independent samples? Because the math is entirely different!



The "Trick" of Dependent Samples

Why do we separate Dependent samples from Independent samples? Because the math is entirely different!

- With independent samples, we had to calculate a complex pooled variance.



The "Trick" of Dependent Samples

Why do we separate Dependent samples from Independent samples? Because the math is entirely different!

- With independent samples, we had to calculate a complex pooled variance.
- With dependent samples, we use a brilliant mathematical "trick".



The "Trick" of Dependent Samples

Why do we separate Dependent samples from Independent samples? Because the math is entirely different!

- With independent samples, we had to calculate a complex pooled variance.
- With dependent samples, we use a brilliant mathematical "trick".

Instead of comparing the "Before" mean to the "After" mean, we simply subtract them to find the **Difference (d)** for each person.

This magically turns a complex two-sample problem into a simple Chapter 10 one-sample problem!



The Formulas for Dependent Samples

We only care about the column of differences (d).

1. Mean of the Differences ($\bar{d}$):

$$\bar{d} = \frac{\sum d}{n}$$



The Formulas for Dependent Samples

We only care about the column of differences (d).

1. Mean of the Differences ($\bar{d}$):

$$\bar{d} = \frac{\sum d}{n}$$

2. Standard Deviation of the Differences (s_d):

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$



The Formulas for Dependent Samples

We only care about the column of differences (d).

1. Mean of the Differences ($\bar{d}$):

$$\bar{d} = \frac{\sum d}{n}$$

2. Standard Deviation of the Differences (s_d):

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

3. The Paired t -Statistic:

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



The Formulas for Dependent Samples

We only care about the column of differences (d).

1. Mean of the Differences ($\bar{d}$):

$$\bar{d} = \frac{\sum d}{n}$$

2. Standard Deviation of the Differences (s_d):

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

3. The Paired t -Statistic:

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

Degrees of Freedom:

$$df = n - 1.$$

(Note: n is the number of **pairs**, not the total number of individual observations!)



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	100
2	1100	1150	



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	100
2	1100	1150	50
3	950	1050	



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	100
2	1100	1150	50
3	950	1050	100
4	1200	1180	



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	100
2	1100	1150	50
3	950	1050	100
4	1200	1180	-20
5	1050	1100	50



Example 1: SAT Prep Course

Conceptual Example: Summer Workshop

An educational firm offers a summer course designed to increase SAT scores. They sample 5 high school juniors, record their SAT scores, put them through the summer workshop, and record their scores as seniors.

Task: At the **0.05 significance level**, can we conclude the course **increased** scores?

Raw Data:

Student	Before (Junior)	After (Senior)	Difference (d : After - Before)
1	1000	1100	100
2	1100	1150	50
3	950	1050	100
4	1200	1180	-20
5	1050	1100	50



SAT Prep Course: Steps 1 to 4

1. Hypotheses (One-Tailed):

We want to prove scores increased (After - Before > 0).

- $H_0 : \mu_d \leq 0$
- $H_1 : \mu_d > 0$



SAT Prep Course: Steps 1 to 4

1. Hypotheses (One-Tailed):

We want to prove scores increased (After - Before > 0).

- $H_0 : \mu_d \leq 0$
- $H_1 : \mu_d > 0$

2. Significance Level: $\alpha = 0.05$.



SAT Prep Course: Steps 1 to 4

1. Hypotheses (One-Tailed):

We want to prove scores increased (After - Before > 0).

- $H_0 : \mu_d \leq 0$
- $H_1 : \mu_d > 0$

2. **Significance Level:** $\alpha = 0.05$.

3. **Test Statistic:** t -distribution.



SAT Prep Course: Steps 1 to 4

1. Hypotheses (One-Tailed):

We want to prove scores increased (After - Before > 0).

- $H_0 : \mu_d \leq 0$
- $H_1 : \mu_d > 0$

2. Significance Level: $\alpha = 0.05$.

3. Test Statistic: t -distribution.

4. Decision Rule:

- $df = n - 1 = 5 - 1 = 4$.
- One-Tailed, $\alpha = 0.05$.
- Critical t -value = **2.132**.



SAT Prep Course: Steps 1 to 4

1. Hypotheses (One-Tailed):

We want to prove scores increased (After - Before > 0).

- $H_0 : \mu_d \leq 0$
- $H_1 : \mu_d > 0$

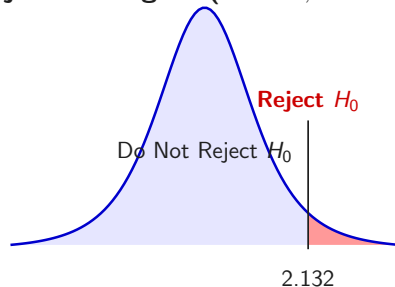
2. **Significance Level:** $\alpha = 0.05$.

3. **Test Statistic:** t -distribution.

4. **Decision Rule:**

- $df = n - 1 = 5 - 1 = 4$.
- One-Tailed, $\alpha = 0.05$.
- Critical t -value = **2.132**.

Rejection Region ($df = 4, \alpha = 0.05$)



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100		



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50		



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100		



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20		



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20	$-20 - 60 = -80$	



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20	$-20 - 60 = -80$	6400
70		



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20	$-20 - 60 = -80$	6400
70	$70 - 60 = 10$	



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20	$-20 - 60 = -80$	6400
70	$70 - 60 = 10$	100
$\Sigma d = 300$		$\Sigma(d - \bar{d})^2 = 9800$



SAT Prep Course: Calculating Statistics

Step 5: Make a Decision (Part 1 - Find $\bar{d}$ and s_d).

d	$(d - \bar{d})$	$(d - \bar{d})^2$
100	$100 - 60 = 40$	1600
50	$50 - 60 = -10$	100
100	$100 - 60 = 40$	1600
-20	$-20 - 60 = -80$	6400
70	$70 - 60 = 10$	100
$\Sigma d = 300$		$\Sigma(d - \bar{d})^2 = 9800$

- $\bar{d} = \frac{300}{5} = \mathbf{60}$

- $s_d = \sqrt{\frac{9800}{5-1}} = \sqrt{\frac{9800}{4}} = \sqrt{2450} = \mathbf{49.50}$



SAT Prep Course: The Final Decision

Step 5: Make a Decision (Part 2 - The t -statistic).

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$



SAT Prep Course: The Final Decision

Step 5: Make a Decision (Part 2 - The t -statistic).

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

$$t = \frac{60 - 0}{49.50 / \sqrt{5}}$$



SAT Prep Course: The Final Decision

Step 5: Make a Decision (Part 2 - The t -statistic).

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

$$t = \frac{60 - 0}{49.50 / \sqrt{5}}$$

$$t = \frac{60}{22.136} = \mathbf{2.71}$$



SAT Prep Course: The Final Decision

Step 5: Make a Decision (Part 2 - The t -statistic).

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

$$t = \frac{60 - 0}{49.50 / \sqrt{5}}$$

$$t = \frac{60}{22.136} = \mathbf{2.71}$$

Step 6: Interpret the Result.

Because $2.71 > 2.132$, we **Reject** H_0 . The data provides statistical evidence that the summer workshop significantly increased the students' SAT scores.



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.

What happened when we assumed they were independent?



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.

What happened when we assumed they were independent?

- We calculated a complex pooled variance ($s_p^2 = 206.5$).



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.

What happened when we assumed they were independent?

- We calculated a complex pooled variance ($s_p^2 = 206.5$).
- We calculated our t -statistic to be **0.716**.



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.

What happened when we assumed they were independent?

- We calculated a complex pooled variance ($s_p^2 = 206.5$).
- We calculated our t -statistic to be **0.716**.
- Our critical limits were ± 2.101 .



Example 2: The Schadek vs. Bowyer Revelation

Discussion

Do you remember this problem from Lecture 6?

We compared the real estate appraisals of two agents (Schadek and Bowyer) on a sample of 10 homes. Last week, we assumed they were **Independent Samples**.

What happened when we assumed they were independent?

- We calculated a complex pooled variance ($s_p^2 = 206.5$).
- We calculated our t -statistic to be **0.716**.
- Our critical limits were ± 2.101 .
- **Result:** We failed to reject H_0 . We concluded there was no difference between the two appraisers.



The Schadek vs. Bowyer Revelation

But wait! Look at the raw data from the textbook (Page 320):

Home	Schadek (\$000s)	Bowyer (\$000s)	Difference (d)
A	235	228	7
B	210	205	5
C	231	219	12
D	242	240	2
E	205	198	7



The Schadek vs. Bowyer Revelation

But wait! Look at the raw data from the textbook (Page 320):

Home	Schadek (\$000s)	Bowyer (\$000s)	Difference (d)
A	235	228	7
B	210	205	5
C	231	219	12
D	242	240	2
E	205	198	7

Schadek and Bowyer evaluated the **exact same homes**. These are **Dependent Samples!**
Let's recalculate this using the Paired t -test trick to see why pairing matters.



The Schadek vs. Bowyer Revelation (Paired Math)

If we evaluate all 10 pairs correctly, the book shows:

- $\Sigma d = 46 \rightarrow \bar{d} = 4.6$
- $\Sigma(d - \bar{d})^2 = 174.40$



The Schadek vs. Bowyer Revelation (Paired Math)

If we evaluate all 10 pairs correctly, the book shows:

- $\Sigma d = 46 \rightarrow \bar{d} = 4.6$
- $\Sigma(d - \bar{d})^2 = 174.40$

1. Find s_d :

$$s_d = \sqrt{\frac{174.40}{9}} = \sqrt{19.377} = 4.402$$



The Schadek vs. Bowyer Revelation (Paired Math)

If we evaluate all 10 pairs correctly, the book shows:

- $\Sigma d = 46 \rightarrow \bar{d} = 4.6$
- $\Sigma(d - \bar{d})^2 = 174.40$

1. Find s_d :

$$s_d = \sqrt{\frac{174.40}{9}} = \sqrt{19.377} = 4.402$$

2. Find the Paired t -statistic:

$$t = \frac{4.6 - 0}{4.402/\sqrt{10}} = \frac{4.6}{1.392} = 3.305$$



Why Dependent Samples are Superior

3. The Paired Decision:

- $df = 10 - 1 = 9$.
- Critical value ($\alpha = 0.05$, Two-Tailed) = ± 2.262 .
- Our computed $t = 3.305$.



Why Dependent Samples are Superior

3. The Paired Decision:

- $df = 10 - 1 = 9$.
- Critical value ($\alpha = 0.05$, Two-Tailed) = ± 2.262 .
- Our computed $t = 3.305$.

Conclusion: Because $3.305 > 2.262$, we **Reject** H_0 ! There *is* a significant difference between the two appraisers!



Why Dependent Samples are Superior

3. The Paired Decision:

- $df = 10 - 1 = 9$.
- Critical value ($\alpha = 0.05$, Two-Tailed) = ± 2.262 .
- Our computed $t = 3.305$.

Conclusion: Because $3.305 > 2.262$, we **Reject** H_0 ! There *is* a significant difference between the two appraisers!

The Lesson: When we treated them as independent, the natural price differences between huge mansions and small condos created massive variance ($s_p^2 = 206.5$), hiding the truth.



Why Dependent Samples are Superior

3. The Paired Decision:

- $df = 10 - 1 = 9$.
- Critical value ($\alpha = 0.05$, Two-Tailed) = ± 2.262 .
- Our computed $t = 3.305$.

Conclusion: Because $3.305 > 2.262$, we **Reject** H_0 ! There *is* a significant difference between the two appraisers!

The Lesson: When we treated them as independent, the natural price differences between huge mansions and small condos created massive variance ($s_p^2 = 206.5$), hiding the truth. By pairing them, we eliminated the house-to-house variation, allowing us to accurately see that Schadek systematically appraises homes higher than Bowyer!



15 Minute Break

Two-Sample Tests about Proportions

Transitioning back to Nominal Data

For the first half of class, we dealt with continuous data (Means). Now, we shift back to nominal data (Success/Failure) from Chapter 15.



Transitioning back to Nominal Data

For the first half of class, we dealt with continuous data (Means). Now, we shift back to nominal data (Success/Failure) from Chapter 15.

- We want to investigate whether two populations have the same proportion of "successes".



Transitioning back to Nominal Data

For the first half of class, we dealt with continuous data (Means). Now, we shift back to nominal data (Success/Failure) from Chapter 15.

- We want to investigate whether two populations have the same proportion of "successes".
- Hypotheses generally look like: $H_0 : \pi_1 = \pi_2$ vs. $H_1 : \pi_1 \neq \pi_2$



Transitioning back to Nominal Data

For the first half of class, we dealt with continuous data (Means). Now, we shift back to nominal data (Success/Failure) from Chapter 15.

- We want to investigate whether two populations have the same proportion of "successes".
- Hypotheses generally look like: $H_0 : \pi_1 = \pi_2$ vs. $H_1 : \pi_1 \neq \pi_2$

Because our null hypothesis assumes the two populations are identical ($\pi_1 = \pi_2$), we must combine the data from both samples to create a single, best estimate of that shared proportion. This is called the **Pooled Proportion** (p_c).



The Pooled Proportion & z-Statistic

1. The Pooled Proportion (p_c):

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}$$

Where:

- x_1, x_2 = Successes in each sample.
- n_1, n_2 = Size of each sample.



The Pooled Proportion & z-Statistic

1. The Pooled Proportion (p_c):

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}$$

Where:

- x_1, x_2 = Successes in each sample.
- n_1, n_2 = Size of each sample.

2. The z-Statistic:

We use the standard normal (z) distribution to test proportions:

$$z = \frac{p_1 - p_2}{\sqrt{p_c(1 - p_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$



The Pooled Proportion & z-Statistic

1. The Pooled Proportion (p_c):

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}$$

Where:

- x_1, x_2 = Successes in each sample.
- n_1, n_2 = Size of each sample.

2. The z-Statistic:

We use the standard normal (z) distribution to test proportions:

$$z = \frac{p_1 - p_2}{\sqrt{p_c(1 - p_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Rule of Thumb: The same rules apply! Both $n_1 p_c$, $n_1(1 - p_c)$, $n_2 p_c$, and $n_2(1 - p_c)$ must be ≥ 5 to safely use the z-distribution.



Example 3: Fear of Flying

Textbook Example: Fear of Flying

A consultant to the airline industry is investigating the fear of flying among adults. Specifically, the consultant wishes to know whether there is a difference in the proportion of men versus women who are fearful of flying.

In a sample of **100 men** (n_1), **19** reported a fear of flying (x_1).

In a sample of **200 women** (n_2), **62** reported a fear of flying (x_2).

Task: At the **0.05 significance level**, is there a difference between the genders?



Fear of Flying: Steps 1 to 4

1. Hypotheses (Two-Tailed):

- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$



Fear of Flying: Steps 1 to 4

1. Hypotheses (Two-Tailed):

- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$

2. Significance Level: $\alpha = 0.05$.



Fear of Flying: Steps 1 to 4

1. Hypotheses (Two-Tailed):

- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$

2. Significance Level: $\alpha = 0.05$.

3. Test Statistic: z-distribution.



Fear of Flying: Steps 1 to 4

1. Hypotheses (Two-Tailed):

- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$

2. Significance Level: $\alpha = 0.05$.

3. Test Statistic: z-distribution.

4. Decision Rule:

- Two-Tailed, $\alpha = 0.05$.
- Split $\alpha \rightarrow 0.025$ per tail.
- Critical z-values = ± 1.96 .



Fear of Flying: Steps 1 to 4

1. Hypotheses (Two-Tailed):

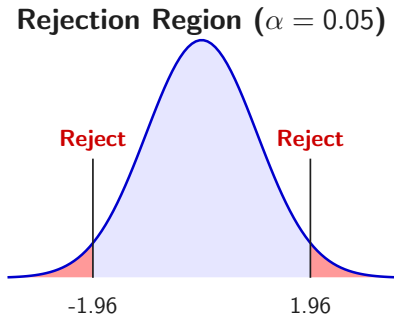
- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$

2. Significance Level: $\alpha = 0.05$.

3. Test Statistic: z-distribution.

4. Decision Rule:

- Two-Tailed, $\alpha = 0.05$.
- Split $\alpha \rightarrow 0.025$ per tail.
- Critical z-values = ± 1.96 .



Fear of Flying: Calculating Proportions

Step 5: Make a Decision (Part 1 - Proportions).

- Sample 1 (Men): $p_1 = \frac{19}{100} = \mathbf{0.19}$
- Sample 2 (Women): $p_2 = \frac{62}{200} = \mathbf{0.31}$



Fear of Flying: Calculating Proportions

Step 5: Make a Decision (Part 1 - Proportions).

- Sample 1 (Men): $p_1 = \frac{19}{100} = 0.19$
- Sample 2 (Women): $p_2 = \frac{62}{200} = 0.31$

Calculate the Pooled Proportion (p_c):

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}$$



Fear of Flying: Calculating Proportions

Step 5: Make a Decision (Part 1 - Proportions).

- Sample 1 (Men): $p_1 = \frac{19}{100} = \mathbf{0.19}$
- Sample 2 (Women): $p_2 = \frac{62}{200} = \mathbf{0.31}$

Calculate the Pooled Proportion (p_c):

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}$$

$$p_c = \frac{19 + 62}{100 + 200} = \frac{81}{300} = \mathbf{0.27}$$

Overall, 27% of the combined 300 respondents fear flying.



Fear of Flying: Calculating z-Statistic

Step 5: Make a Decision (Part 2 - The z-Statistic).

$$z = \frac{0.19 - 0.31}{\sqrt{0.27(1 - 0.27) \left(\frac{1}{100} + \frac{1}{200} \right)}}$$



Fear of Flying: Calculating z-Statistic

Step 5: Make a Decision (Part 2 - The z-Statistic).

$$z = \frac{0.19 - 0.31}{\sqrt{0.27(1 - 0.27) \left(\frac{1}{100} + \frac{1}{200} \right)}}$$

$$z = \frac{-0.12}{\sqrt{0.27(0.73) (0.01 + 0.005)}}$$



Fear of Flying: Calculating z-Statistic

Step 5: Make a Decision (Part 2 - The z-Statistic).

$$z = \frac{0.19 - 0.31}{\sqrt{0.27(1 - 0.27) \left(\frac{1}{100} + \frac{1}{200} \right)}}$$

$$z = \frac{-0.12}{\sqrt{0.27(0.73) (0.01 + 0.005)}}$$

$$z = \frac{-0.12}{\sqrt{0.1971(0.015)}} = \frac{-0.12}{\sqrt{0.0029565}}$$



Fear of Flying: Calculating z-Statistic

Step 5: Make a Decision (Part 2 - The z-Statistic).

$$z = \frac{0.19 - 0.31}{\sqrt{0.27(1 - 0.27) \left(\frac{1}{100} + \frac{1}{200} \right)}}$$

$$z = \frac{-0.12}{\sqrt{0.27(0.73) (0.01 + 0.005)}}$$

$$z = \frac{-0.12}{\sqrt{0.1971(0.015)}} = \frac{-0.12}{\sqrt{0.0029565}}$$

$$z = \frac{-0.12}{0.05437} = -2.207$$



Fear of Flying: Final Decision

Step 6: Interpret the Result.



Fear of Flying: Final Decision

Step 6: Interpret the Result.

- Our computed z-statistic is -2.207 .



Fear of Flying: Final Decision

Step 6: Interpret the Result.

- Our computed z-statistic is -2.207 .
- Our left-tail critical limit is -1.96 .



Fear of Flying: Final Decision

Step 6: Interpret the Result.

- Our computed z-statistic is -2.207 .
- Our left-tail critical limit is -1.96 .
- Because -2.207 is less than -1.96 , it falls into the red rejection region.



Fear of Flying: Final Decision

Step 6: Interpret the Result.

- Our computed z-statistic is -2.207 .
- Our left-tail critical limit is -1.96 .
- Because -2.207 is less than -1.96 , it falls into the red rejection region.

Conclusion: We **Reject** H_0 . The data provides significant evidence that there is a difference in the fear of flying between men and women. Specifically, a higher proportion of women fear flying compared to men.



Pop-Up Question 2: Which Test?

Discussion

Pop-Up Question!

Read the scenario below and decide which test we should use:

(A) Independent t -test, (B) Dependent t -test, or (C) Two-Proportion z -test.

"A marketing firm wants to know if their new website design leads to more purchases than the old design. They show the old design to 500 visitors and record that 45 bought an item. They show the new design to 500 different visitors and record that 60 bought an item."



Pop-Up Question 2: Which Test?

Discussion

Pop-Up Question!

Read the scenario below and decide which test we should use:

(A) Independent t -test, (B) Dependent t -test, or (C) Two-Proportion z -test.

"A marketing firm wants to know if their new website design leads to more purchases than the old design. They show the old design to 500 visitors and record that 45 bought an item. They show the new design to 500 different visitors and record that 60 bought an item."

Answer: (C) Two-Proportion z -test.

Why? Because the data is nominal ("Bought" or "Did Not Buy"). We are comparing the *proportion* of successes (45/500 vs. 60/500), not an average amount spent.



Summary & Next Steps

Summary & Preparation for Next Lecture

Summary:

- We explored **Dependent Samples**, realizing that matching data points (before/after) allows us to eliminate individual variation, creating a much more accurate test.
- We used the mathematical trick of analyzing the **Difference (d)** to run a one-sample t -test.
- We shifted to **Two-Sample Proportions**, learning to calculate a **Pooled Proportion (p_c)** to run a standard normal z -test.



Summary & Preparation for Next Lecture

Summary:

- We explored **Dependent Samples**, realizing that matching data points (before/after) allows us to eliminate individual variation, creating a much more accurate test.
- We used the mathematical trick of analyzing the **Difference (d)** to run a one-sample t -test.
- We shifted to **Two-Sample Proportions**, learning to calculate a **Pooled Proportion (p_c)** to run a standard normal z -test.

Prep for Next Lecture (Lecture 8):

- Next week is a **Software Lab and Class Activity!**
- We will be applying all of the concepts from Weeks 1 through 4 using Excel and other statistical software tools to prepare for the Midterm.

