Two-Sample Tests

Learning Objectives

In this chapter, you learn how to use hypothesis testing for comparing the difference between:

- The means of two independent populations

Two-Sample Tests Overview

Two Sample Tests

Independent Population Means	Means, Related Populations	Independent Population Proportions	Independent Population Variances

Examples

Group 1 vs. Group 2	Same group before vs. after treatment	Proportion 1vs. Proportion 2	Variance 1 vs. Variance 2

Two-Sample Tests

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

Goal: Test hypothesis or form a confidence interval for the difference between two population means, $\mu_1 - \mu_2$

The point estimate for the difference between sample means:

$$\overline{X}_1 - \overline{X}_2$$

Two-Sample Tests Independent Populations

Independent Population Means

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown

- Different data sources
  - Independent: Sample selected from one population has no effect on the sample selected from the other population
- Use the difference between 2 sample means
- Use Z test, pooled variance t test, or separate-variance t test

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known → Use a **Z** test statistic

σ_1 and σ_2 unknown → Use S to estimate unknown σ, use a **t** test statistic

Two-Sample Tests
Independent Populations

Independent Population Means

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown

Assumptions:

- Samples are randomly and independently drawn
- population distributions are normal

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

When σ_1 and σ_2 are known and both populations are normal, the test statistic is a Z-value and the standard error of $\bar{X}_1 - \bar{X}_2$ is

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Two-Sample Tests
Independent Populations

Independent
Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

The test statistic is:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}}}$$

Two-Sample Tests Independent Populations

Two Independent Populations, Comparing Means

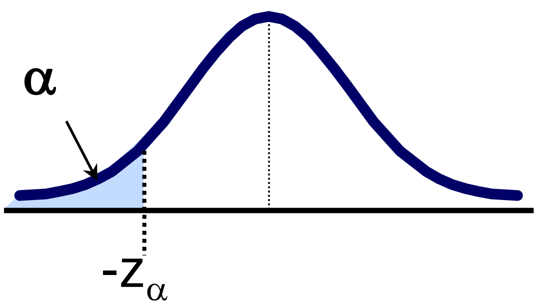
Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: \mu_1 \geq \mu_2$ $H_1: \mu_1 < \mu_2$	$H_0: \mu_1 \leq \mu_2$ $H_1: \mu_1 > \mu_2$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
i.e.,	i.e.,	i.e.,
$H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_1: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$

Two-Sample Tests Independent Populations

Two Independent Populations, Comparing Means

Lower-tail test:

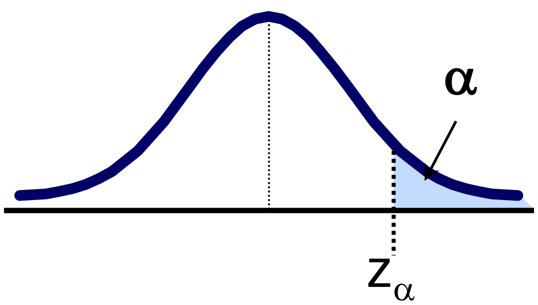
H_0: $\mu_1 - \mu_2 \geq 0$
H_1: $\mu_1 - \mu_2 < 0$



Reject H_0 if $Z < -Z_a$

Upper-tail test:

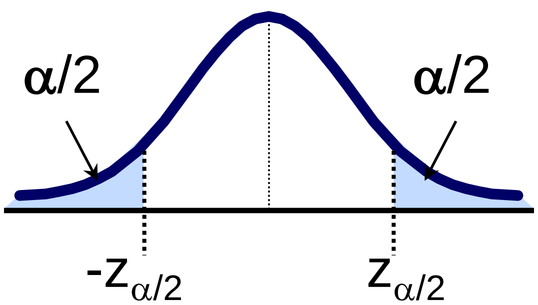
H_0: $\mu_1 - \mu_2 \leq 0$
H_1: $\mu_1 - \mu_2 > 0$



Reject H_0 if $Z > Z_a$

Two-tail test:

H_0: $\mu_1 - \mu_2 = 0$
H_1: $\mu_1 - \mu_2 \neq 0$



Reject H_0 if $Z < -Z_{a/2}$
or $Z > Z_{a/2}$

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown but assumed equal

Two-Sample Tests Independent Populations

Independent Population Means

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown

Forming interval estimates:

- The population variances are assumed equal, so use the two sample standard deviations and pool them to estimate σ
- the test statistic is a t value with $(n_1 + n_2 - 2)$ degrees of freedom

Two-Sample Tests Independent Populations

Independent Population Means

- σ_1 and σ_2 known
- σ_1 and σ_2 unknown

The pooled standard deviation is:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^{\,2} + (n_2 - 1)S_2^{\,2}}{(n_1 - 1) + (n_2 - 1)}}$$

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

The test statistic is:

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

Where t has $(n_1 + n_2 - 2)$ d.f., and

$$S_p^2 = \frac{(n_1 - 1)S_1^{\ 2} + (n_2 - 1)S_2^{\ 2}}{(n_1 - 1) + (n_2 - 1)}$$

Two-Sample Tests Independent Populations

- You are a financial analyst for a brokerage firm. Is there a difference in dividend yield between stocks listed on the NYSE & NASDAQ? You collect the following data:

	NYSE	NASDAQ
Number	**21**	**25**
Sample mean	**3.27**	**2.53**
Sample std dev	**1.30**	**1.16**

Assuming both populations are approximately normal with equal variances, is there a difference in average yield ($\alpha = 0.05$)?

Two-Sample Tests Independent Populations

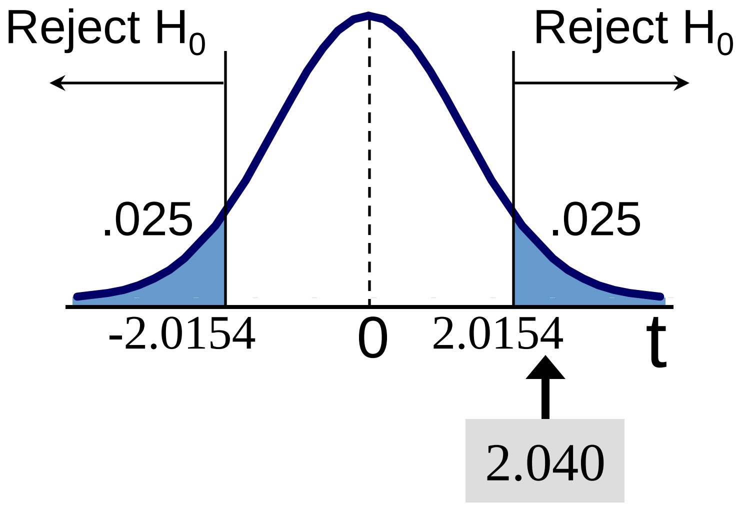
The test statistic is:

$$t=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{\left(3.27-2.53\right)-0}{\sqrt{1.5021\left(\frac{1}{21}+\frac{1}{25}\right)}}=2.040$$

$$S_p^2=\frac{\left(n_1-1\right)S_1{}^2+\left(n_2-1\right)S_2{}^2}{(n_1-1)+(n_2-1)}=\frac{\left(21-1\right)1.30^2+\left(25-1\right)1.16^2}{(21-1)+(25-1)}=1.5021$$

Two-Sample Tests
Independent Populations

- **H_0: $\mu_1 - \mu_2 = 0$ i.e. ($\mu_1 = \mu_2$)**
- **H_1: $\mu_1 - \mu_2 \neq 0$ i.e. ($\mu_1 \neq \mu_2$)**
- **$\alpha = 0.05$**
- **df = 21 + 25 - 2 = 44**
- **Critical Values: t = ± 2.0154**
- **Test Statistic: 2.040**



Decision: Reject H_0 at $\boldsymbol{\alpha} = 0.05$

Conclusion: There is evidence of a difference in the means.

Independent Populations Unequal Variance

- If you cannot assume population variances are equal, the pooled-variance t test is inappropriate
- Instead, use a separate-variance t test, which includes the two separate sample variances in the computation of the test statistic
- The computations are complicated

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

The confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm Z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

Two-Sample Tests Independent Populations

Independent Population Means

σ_1 and σ_2 known

σ_1 and σ_2 unknown

The confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{n_1+n_2-2} \sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Where

$$S_p^2 = \frac{(n_1-1)S_1^{\,2} + (n_2-1)S_2^{\,2}}{(n_1-1)+(n_2-1)}$$

Chapter Summary

In this chapter, we have

- Compared two independent samples
  - Performed Z test for the differences in two means
  - Performed pooled variance t test for the differences in two means
  - Formed confidence intervals for the differences between two means