# Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 13, 2021

2021 Spring

# Chapter 9. Two-Sample Inferences

§ 9.1 Introduction

§ 9.2 Testing $H_0 : \mu_X = \mu_Y$

§ 9.3 Testing $H_0 : \sigma_X^2 = \sigma_Y^2$

§ 9.4 Binomial Data: Testing $H_0 : p_X = p_Y$

§ 9.5 Confidence Intervals for the Two-Sample Problem

§ 9.1 Introduction

§ 9.2 Testing $H_0 : \mu_X = \mu_Y$

§ 9.3 Testing $H_0 : \sigma_X^2 = \sigma_Y^2$

§ 9.4 Binomial Data: Testing $H_0 : p_X = p_Y$

§ 9.5 Confidence Intervals for the Two-Sample Problem

- Let $X_1, \cdots, X_n$ be a random sample of size $n$ from $N(\mu_X, \sigma_X^2)$.
- Let $Y_1, \cdots, Y_m$ be a random sample of size $m$ from $N(\mu_Y, \sigma_Y^2)$.

**Prob. 1** Testing $H_0 : \mu_X = \mu_Y$ if $\sigma_X^2 = \sigma_Y^2$.

**Prob. 2** Testing $H_0 : \mu_X = \mu_Y$ if $\sigma_X^2 \neq \sigma_Y^2$.

---

- True means: $\mu_X$, $\mu_Y$
- True std. dev.'s: $\sigma_X$, $\sigma_Y$
- True variances: $\sigma_X^2$, $\sigma_Y^2$

- Sample means: $\overline{X}$, $\overline{Y}$
- Sample std. dev.'s: $S_X$, $S_Y$
- Sample variances: $S_X^2$, $S_Y^2$

## When $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

**Def.** The **pooled variance**: $S_p^2 = \dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

$$= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{j=1}^n (Y_j - \overline{Y})^2}{n+m-2}$$

**Thm.** $T_{n+m-2} = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim$ Student t distr. of $n+m-2$ dgs of fd.

**Proof.** (See slides on Section 9.1) $\square$

# When $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

Testing $H_0 : \mu_X = \mu_Y$ v.s.

(at the $\alpha$ level of significance)

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

| $H_1 : \mu_X < \mu_Y$: | $H_1 : \mu_X \neq \mu_Y$: | $H_1 : \mu_X > \mu_Y$: |
|---|---|---|
| Reject $H_0$ if | Reject $H_0$ if | Reject $H_0$ if |
| $t \leq -t_{\alpha,n+m-2}$ | $\lvert t \rvert \geq t_{\alpha/2,n+m-2}$ | $t \geq t_{\alpha,n+m-2}$ |

E.g. Test whether Mark Twain and Snodgrass are the same person by checking the proportion of three-letter words at the 99% level of significance.

**Table 9.2.1** Proportion of Three-Letter Words

| Twain | Proportion | QCS | Proportion |
|---|---|---|---|
| Sergeant Fathom letter | 0.225 | Letter I | 0.209 |
| Madame Caprell letter | 0.262 | Letter II | 0.205 |
| Mark Twain letters in | | Letter III | 0.196 |
| *Territorial Enterprise* | | Letter IV | 0.210 |
| First letter | 0.217 | Letter V | 0.202 |
| Second letter | 0.240 | Letter VI | 0.207 |
| Third letter | 0.230 | Letter VII | 0.224 |
| Fourth letter | 0.229 | Letter VIII | 0.223 |
| First *Innocents Abroad* letter | | Letter IX | 0.220 |
| First half | 0.235 | Letter X | 0.201 |
| Second half | 0.217 | | |

Sol. We need to test

$$H_0 : \mu_X = \mu_Y \quad v.s. \quad H_1 : \mu_X \neq \mu_Y.$$

Since we are tesing whether they are the same person, one can assume that $\sigma_X^2 = \sigma_Y^2$.

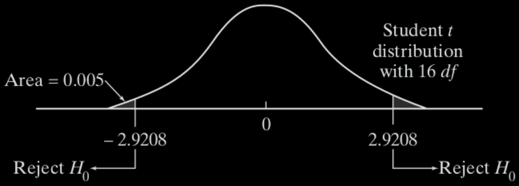
1. $n = 8,\ m = 10,$

$$\sum_{i=1}^{n} x_i = 1.855, \quad \sum_{i=1}^{n} x_i^2 = 0.4316$$
$$\sum_{i=1}^{m} y_i = 2.097, \quad \sum_{i=1}^{m} y_i^2 = 0.4406$$

2. Hence,

$$\bar{x} = 1.855/8 = 02319 \quad \bar{y} = 2.097/10 = 0.2097$$
$$s_X^2 = \frac{8 \times 0.4316 - 1.855^2}{8 \times 7} = 0.0002103$$
$$s_Y^2 = \frac{10 \times 0.4406 - 2.097^2}{10 \times 9} = 0.0000955$$
$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = ... = 0.0001457$$
$$t = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = ... = 3.88$$

3. Critical region: $|t| \geq t_{0.005,n+m-2} = t_{0.005,16} = 2.9208.$




4. Conclusion: Rejection! □

E.g. Comparing large-scales and small-scales companies:

Based on the data below, can we say that the return o equity differs between the two types of companies?

**Table 9.2.4**

| Large-Sales Companies | Return on Equity (%) | Small-Sales Companies | Return on Equity (%) |
|---|---|---|---|
| Deckers Outdoor | 21 | NVE | 21 |
| Jos. A. Bank Clothiers | 23 | Hi-Shear Technology | 21 |
| National Instruments | 13 | Bovie Medical | 14 |
| Dolby Laboratories | 22 | Rocky Mountain Chocolate Factory | 31 |
| Quest Software | 7 | Rochester Medical | 19 |
| Green Mountain Coffee Roasters | 17 | Anika Therapeutics | 19 |
| Lufkin Industries | 19 | Nathan's Famous | 11 |
| Red Hat | 11 | Somanetics | 29 |
| Matrix Service | 2 | Bolt Technology | 20 |
| DXP Enterprises | 30 | Energy Recovery | 27 |
| Franklin Electric | 15 | Transcend Services | 27 |
| LSB Industries | 43 | IEC Electronics | 24 |

Sol. Let $\mu_X$ and $\mu_Y$ be the average returns. We are asked to test

$$H_0 : \mu_X = \mu_Y \qquad v.s. \qquad H_1 : \mu_X \neq \mu_Y.$$

1.

$$n = 12, \qquad \sum_{i=1}^{n} x_i = 223 \qquad \sum_{i=1}^{n} x_i^2 = 5421$$

$$m = 12, \qquad \sum_{i=1}^{m} y_i = 263 \qquad \sum_{i=1}^{m} y_i^2 = 6157$$

2.

$$\bar{x} = 18.5833, \qquad s_X^2 = 116.0833$$

$$\bar{y} = 21.9167, \qquad s_Y^2 = 35.7197$$

$$w = \frac{18.5833 - 21.9167}{\sqrt{\frac{116.0833}{12} + \frac{35.7197}{12}}} = -0.9371932.$$

$$\hat{\theta} = \frac{116.0833}{35.7179} = 3.250 \quad \Rightarrow \quad \nu = \left[ \frac{(3.250 + 1)^2}{\frac{1}{11} 3.250^2 + \frac{1}{11} 1^2} \right] = [17.18403] = 17.$$

3. The critical region is $|w| \geq t_{\alpha/2,17} = 2.1098$.

4. Conclusion:

   Since $w = -0.94$ is not in the critical region, we fail to reject $H_0$. $\square$