Chapter 9
Comparing Two Population Means

9.1 Introduction

9.2 Analysis of Paired Samples

9.3 Analysis of Independent Samples

9.1 Introduction

9.1.1 Two Sample Problems

- A set of data observations $x_1, \ldots, x_n$ from a population A
  with a cumulative dist. $F_A(x)$,
  a set of data observations $y_1, \ldots, y_m$ from another population B
  with a cumulative dist. $F_B(x)$.

- How to compare the means of the two populations, μ_A and μ_B ?
- What if the variances are not the same between the two populations ?

9.1.1 Two Sample Problems

- Example 51. **Acrophobia Treatments**
  - In an experiment to investigate whether the new treatment is effective or not, a group of 30 patients suffering from acrophobia are randomly assigned to one of the two treatment methods.

  - 15 patients undergo the standard treatment, say A, and 15 patients undergo the proposed new treatment B.

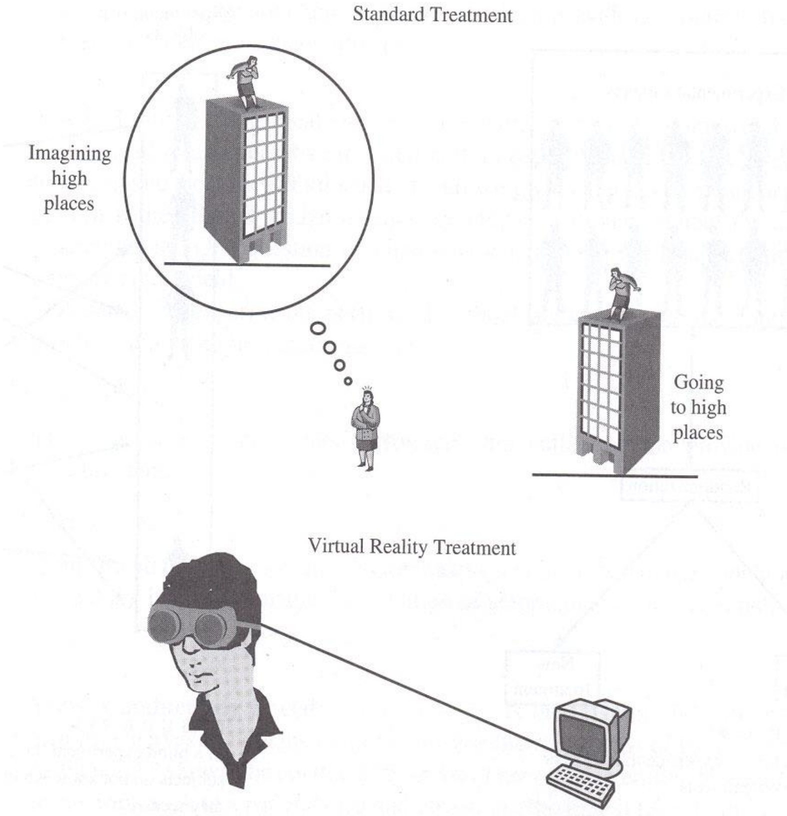

Fig.9.3 Treating acrophobia.

- observations $x_1, \dots , x_{15}$ ~ A (standard treatment),
  observations $y_1, \dots , y_{15}$ ~ B (new treatment).

- For this example, a comparison of the population means μ_A and μ_B, provides an indication of whether the new treatment is any better or any worse than the standard treatment.

- It is good experimental practice to **randomize** the allocation of subjects or experimental objects between standard treatment and the new treatment, as shown in Figure 9.4.

- Randomization helps to eliminate any bias that may otherwise arise if certain kinds of subject are “favored” and given a particular treatment.

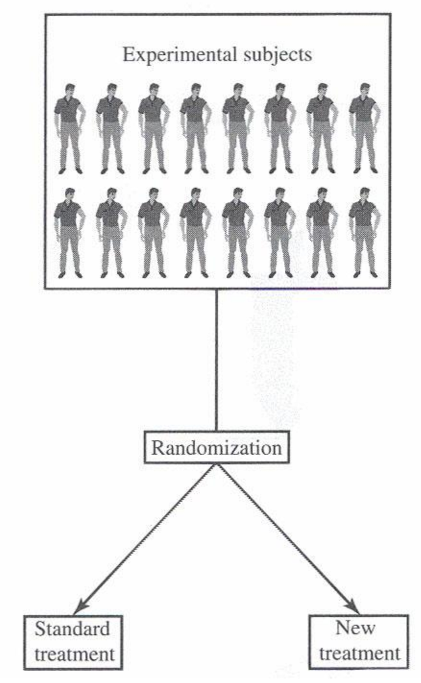

Fig.9.4 Randomization of experimental subjects between two treatment

• Some more words: placebo, blind experiment, double-blind experiment

- Example 45. **Fabric Absorption Properties**

  - If the rollers rotate at 24 revolutions per minute, how does changing the pressure from 10 pounds per square inch (type A) to 20 pounds per square inch (type B) influence the water pickup of the fabric?

  - data observations x_i of the fabric water pickup with type A pressure and observations y_i with type B pressure.

  - A comparison of the population means μ_A and μ_B shows how the average fabric water pickup is influenced by the change in pressure.

- Consider testing $\mathrm{H}_0\colon \mu_A = \mu_B$

- The p-value can be obtained just in the same way as for one-sample problems

9.1.2 Paired Samples Versus Independent Samples

- Example 55. **Heart Rate Reduction**

  - A new drug for inducing a temporary reduction in a patient's heart rate is to be compared with a standard drug.

  - Since the drug efficacy is expected to depend heavily on the particular patient involved, a ***paired experiment*** is run whereby each of 40 patients is administered one drug on one day and the other drug on the following day.

  - **blocking**: it is important to block out unwanted sources of variation that otherwise might cloud the comparisons of real interest

- Data from paired samples are of the form $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ which arise from each of n experimental subjects being subjected to both "treatments"

- The comparison between the two treatments is then based upon the pairwise differences $z_i = x_i - y_i$, $1 \le i \le n$

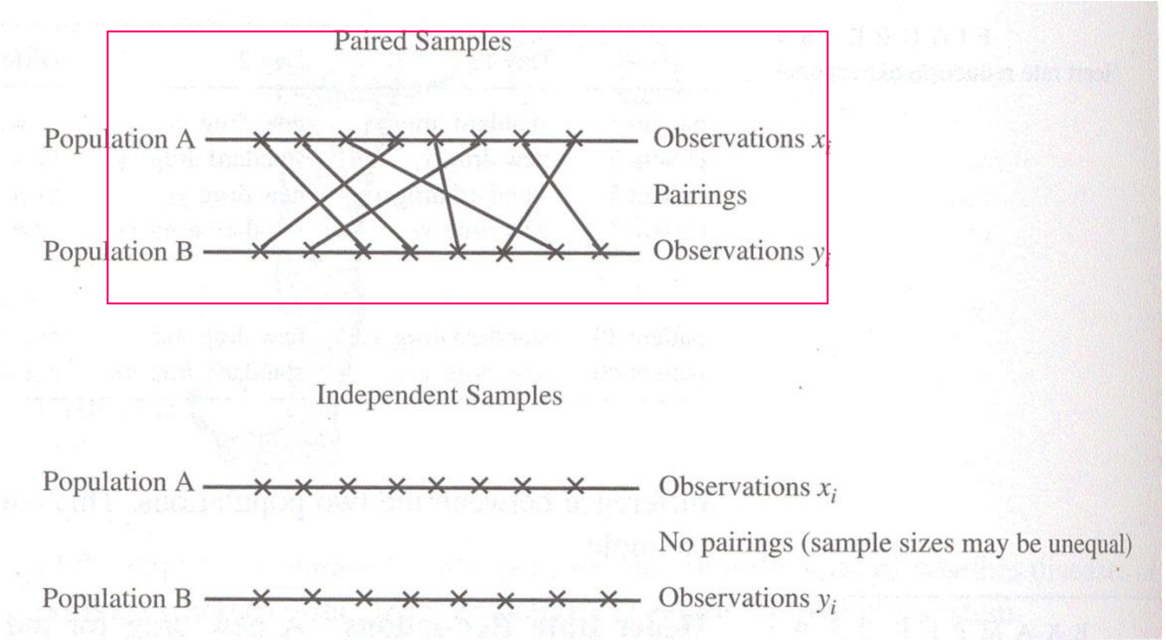

Fig.9.10 Paired and independent samples

9.2 Analysis of Paired Samples

9.2.1 Methodology

- Data observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

  One sample technique can be applied to the data set

$$z_i = x_i - y_i \, , \; 1 \le i \le n,$$

  in order to make inferences about the unknown mean μ (average difference).

- $\mathrm{x_i} = \mu_A + \gamma_i + \epsilon_I^A,$

  $\mathrm{y_i} = \mu_B + \gamma_i + \epsilon_I^B,$

where μ_A (or μ_B) is the effect of treatment A (or B), γ_i the effect by subject I, and ϵ_i^A (or ϵ_i^B) the measurement error for subject I under treatment A (or B).

- $z_i = \mu_A - \mu_B + \epsilon_i^{AB}$

z_i may be regarded as observations from a distribution with mean

$$\mu = \mu_A - \mu_B.$$

9.2.2 Examples

- Example 55. **Heart Rate Reduction**

  -The sample mean $\overline{z} = -2.655$, the sample standard deviation $s = 3.730$, so that

$$t = \frac{\sqrt{n}(\overline{z} - \mu)}{s} = -4.50.$$

  -For testing H_0: $\mu = 0$ vs $H_A: \mu \neq 0$,

  p-value $= 2 \times P(T > 4.50) \approx 0.0001$ where $T \sim t_{39}$.

Patient	Standard drug x_i	New drug y_i	$z_i = x_i - y_i$
1	28.5	34.8	−6.3
2	26.6	37.3	−10.7
3	28.6	31.3	−2.7
4	22.1	24.4	−2.3
5	32.4	39.5	−7.1
6	33.2	34.0	−0.8
7	32.9	33.4	−0.5
8	27.9	27.4	0.5
9	26.8	35.4	−8.6
10	30.7	35.7	−5.0
11	39.6	40.4	−0.8
12	34.9	41.6	−6.7
13	31.1	30.8	0.3
14	21.6	30.5	−8.9
15	40.2	40.7	−0.5
16	38.9	39.9	−1.0
17	31.6	30.2	1.4
18	36.0	34.5	1.5
19	25.4	31.2	−5.8
20	35.6	35.5	0.1
21	27.0	25.3	1.7
22	33.1	34.5	−1.4
23	28.7	30.9	−2.2
24	33.7	31.9	1.8
25	33.7	36.9	−3.2
26	34.3	27.8	6.5
27	32.6	35.7	−3.1
28	34.5	38.4	−3.9
29	32.9	36.7	−3.8
30	29.3	36.3	−7.0
31	35.2	38.1	−2.9
32	29.8	32.1	−2.3
33	26.1	29.1	−3.0
34	25.6	33.5	−7.9
35	27.6	28.7	−1.1
36	25.1	31.4	−6.3
37	23.7	22.4	1.3
38	36.3	43.7	−7.4
39	33.4	30.8	2.6
40	40.1	40.8	−0.7

- A 99% two-sided confidence interval of $\mu = \mu_A - \mu_B$:

$$\left(\bar{z} - \frac{t_{0.005,39}s}{\sqrt{n}}, \bar{z} + \frac{t_{0.005,39}s}{\sqrt{n}}\right) = (-4.252, -1.058)$$

- Consequently, the new drug provides a reduction in a patient's heart rate of somewhere between 1% and 4.25% more on average than the standard drug.

• This confidence interval also suggests rejection of H_0: $\mu = 0$ against H_A: $\mu \neq 0$ at the significance level 0.01.

9.3 Analysis of Independent Samples

	Samples	size	mean	standard deviation
Population A	$x_1, x_2, \cdots, x_n$	n	$\overline{x}$	s_x
Population B	$y_1, y_2, \cdots, y_m$	m	$\overline{y}$	s_y

The point estimate of $\mu_A - \mu_B$ is $\overline{x} - \overline{y}$.

The standard error of $\overline{x} - \overline{y}$ is $se(\overline{x} - \overline{y}) = \sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}$

- Assume that σ_A^2 and σ_B^2 are unknown.

  The s.e.'s for two-sample t-tests are as follows:
  - The general procedure:

$$s.e.(\overline{x}-\overline{y})=\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}.$$

  - The pooled variance procedure:

$$s.e.(\overline{x}-\overline{y})=s_p\sqrt{\frac{1}{n}+\frac{1}{m}}$$

    where $s_p^2=\frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2}$ (called the pooled sample variance).

- When the variances are known, we use a two-sample z-test.

9.3.1 General Procedure (Smith-Satterthwaite test)

- We use the statistic

$$T = \frac{\bar{x}-\bar{y}-(\mu_A-\mu_B)}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}}.$$

This statistic follows approximately t-distribution with the d.f., v, as the largest integer not larger than

$$v^* = \frac{\left(s_x^2/n + s_y^2/m\right)^2}{s_x^4/n^2(n-1) + s_y^4/m^2(m-1)}$$

- A two-sided $1-\alpha$ level confidence interval for $\mu_A - \mu_B$ is given by the end points

$$\bar{x} - \bar{y} \pm t_{\alpha/2,v}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

- For testing $H_0: \mu_A - \mu_B = \delta$ vs $H_A: \mu_A - \mu_B \neq \delta$, the appropriate t-statistic is

$$T = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

9.3.2 Pooled Variance Procedure

- Assume $\sigma_A^2 = \sigma_B^2 = \sigma^2$.
- The unbiased estimate $\hat{\sigma}^2$ of σ^2 is given by

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}.$$

- The statistic

$$T = \frac{\overline{x} - \overline{y} - (\mu_A - \mu_B)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \ t_{n+m-2}$$

- A two-sided $1-\alpha$ level confidence interval for $\mu_A - \mu_B$ is given by the end-points

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

- For testing $H_0: \mu_A - \mu_B = \delta$ vs $H_A: \mu_A - \mu_B \neq \delta$, the appropriate t-statistic is

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

which follows $t_{\mathrm{n+m-2}}$ distribution under H_0.

Example of two samples

- Suppose we have two samples, $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$, from Normal distributions $N(\mu_A, \sigma_A^2)$ and $N(\mu_B, \sigma_B^2)$ respectively. The observed results from the samples are

$$n = 24,\ \bar{x} = 9.005,\ \ s_x = 3.438$$

$$\text{and } \mathrm{m} = 34,\ \bar{y} = 11.864,\ \ s_y = 3.305.$$

- If we apply the general procedure:

$$\mathrm{t} = \frac{9.005 - 11.864}{\sqrt{3.438^2/24 + 3.305^2/34}} = -3.169$$

$$v^* = \frac{\left(3.438^2/24 + 3.305^2/34\right)^2}{3.438^4/(24^2 \times 23) + 3.305^4/(34^2 \times 33)} = 48.43$$

- If we apply the pooled-variance procedure:

$$t = \frac{9.005-11.864}{s_p\sqrt{1/24+1/34}} = -3.192$$

$$s_p = \sqrt{\frac{23\times3.438^2+33\times3.305^2}{24+34-2}} = 3.360.$$

- For testing $H_0: \mu_A - \mu_B = 0$ vs $H_A: \mu_A - \mu_B \neq 0$
  - Using the general procedure

    P-value$=2 \times P(T \geq |t|) = 2 \times P(T \geq 3.169) = 0.0027$
  - Using the pooled-variance procedure

    P-value$=2 \times P(T \geq |t|) = 2 \times P(T \geq 3.192) = 0.0023$

9.3.3. z-Procedure

- When the population variances are known for the two samples, we can use a z-statistic instead of a t-statistic.
- A two-sided $1-\alpha$ level confidence interval for $\mu_A - \mu_B$ is given by the end-points

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}$$

- For testing $H_0: \mu_A - \mu_B = \delta$ vs $H_A: \mu_A - \mu_B \neq \delta$:

  We use $Z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}$ which follows $N(0,1)$ under H_0.

9.3.4. Examples

Example 51. **Acrophobia Treatments**

Test $H_0: \mu_A \geq \mu_B$ vs $H_A: \mu_A < \mu_B$

From data

$n = m = 15, \quad \bar{x} = 47.47, \quad \bar{y} = 51.20,$

$s_x = 11.40, \quad s_y = 10.09.$

- Using the general procedure:

$$v = \frac{\left(11.40^2/15 + 10.09^2/15\right)^2}{11.40^4/(15^2 \times 14) + 10.09^4/(15^2 \times 14)} = 27.59.$$

$$t = \frac{47.47 - 51.20}{\sqrt{11.40^2/15 + 10.09^2/15}} = -0.949.$$

$$P(T < -0.949) = 0.175.$$

Fig.9.20 Acrophobia treatment data set (improvement scores)

Standard treatment x_i	New treatment y_i
33	65
54	61
62	37
46	47
52	45
42	53
34	53
51	69
26	49
68	42
47	40
40	67
46	46
51	43
60	51

- Using the pooled variance procedure

$$s_p^2 = \frac{14\times 11.40^2 + 14\times 10.09^2}{28} = 115.88 \quad \text{or} \quad s_p = 10.8.$$

$$t = \frac{47.47 - 51.20}{10.8\sqrt{1/15 + 1/15}} = -0.946.$$

$$v = n + m - 2 = 28.$$

Almost the same as in the general procedure.

9.3.5. Sample Size Calculations

- Problem : determination of appropriate sample sizes n and m, or
  an assessment of the precision afforded by given sample sizes

- Suppose we use the pooled variance procedure.
  Then the interval length is

$$L = 2 \times t_{\alpha/2,v} s_p \sqrt{1/n + 1/m}$$

Example 45. **Fabric Absorption Properties**

$n = m = 15.\ s_x = 4.943, \quad s_y = 4.991.\ \ s_p^2 = 24.67.$

A 99% confidence two-sided confidence interval for $\mu_A - \mu_B$ was obtained as $(6.24, 16.26)$. The interval length is over 10%.

Find the desired sample sizes, $n = m = n_0$, so that the interval length is not larger than $L_0 = 4\%$.

$$L = 2 \times t_{\alpha/2,v} s_p \sqrt{1/n_0 + 1/n_0} \leq L_0 = 4$$

So n_0 must satisfy

$$\sqrt{n_0} \geq 2 \times t_{\alpha/2,v} s_p \sqrt{2}/4$$

- We can find the desired sample sizes analogously when we use the general procedure.

Python codes for two-sample tests of means

```
import numpy as np
import pandas as pd
import statsmodels.stats.weightstats as sms
data=pd.read_csv(“data/taxi.txt",sep='₩t',index_col=0)
dat=data/1000
print(dat.describe())
dat_A= dat['BrandA']
dat_B= dat['BrandB']
cm=sms.CompareMeans(sms.DescrStatsW(dat_A),
sms.DescrStatsW(dat_B))
```

```
print("General Two Sample t-test")
print("alternative hypothesis: true difference in means is not equal to 0")
print("t, p-value, df: %.4f, %.4f, %.1f" %cm.ttest_ind('two-sided', 'unequal'))
print("mean difference:", np.mean(dat_A) - np.mean(dat_B))

print("alternative hypothesis: true difference in means is greater than 0")
print("t, p-value, df: %.4f, %.4f, %.1f" %cm.ttest_ind('larger', 'unequal'))

print("Common-variance Two Sample t-test")
        # When sigma_1= sigma_2
print("alternative hypothesis: true difference in means is not equal to 0")
print("t, p-value, df: %.4f, %.4f, %.1f" %cm.ttest_ind('two-sided', 'pooled'))
```

Python output

	BrandA	BrandB
count	8.000000	8.000000
mean	37.250000	38.362500
std	6.546755	6.181063
min	30.100000	31.100000
25%	32.600000	35.175000
50%	35.550000	37.200000
75%	39.950000	40.875000
max	48.400000	47.800000

General Two Sample t-test

alternative hypothesis: true difference in means is not equal to 0

t, p-value, df: -0.3495, 0.7319, 14.0

mean difference: -1.1124999999999972

alternative hypothesis: true difference in means is greater than 0

t, p-value, df: -0.3495, 0.6340, 14.0

Common-variance Two Sample t-test

alternative hypothesis: true difference in means is not equal to 0

t, p-value, df: -0.3495, 0.7319, 14.0

Python codes for confidence intervals

```
cm = sms.CompareMeans(sms.DescrStatsW(data['BrandA']),
sms.DescrStatsW(data['BrandB']))

print("confidence interval:", cm.tconfint_diff(0.05, 'two-sided', 'unequal'))
                    # Using Satterthwaite approximation
    confidence interval: (-7942.041914420463, 5717.041914420463)

print("confidence interval:", cm.tconfint_diff(0.05, 'larger', 'unequal'))
    confidence interval: (-6720.5276208856, inf)

print("confidence interval:", cm.tconfint_diff(0.05, 'two-sided', 'pooled'))
                    # Assuming common variances
    confidence interval: (-7939.929753944845, 5714.929753944845)
```

Chapter Summary

9.1 Introduction

9.2 Analysis of Paired Samples

9.3 Analysis of Independent Samples

- **General procedure**
- **Pooled-variance procedure**