

Comparison of Two Population Means

Design of Experiments - Montgomery
Section 2-4 and 2-5

3

Two-sample T-Test

- $H_0 : \mu_1 = \mu_2$ (Null Hypothesis)
- $H_1 : \mu_1 \begin{matrix} < \\ > \\ \neq \end{matrix} \mu_2$ (Alternative Hypothesis)
- Collect data - n_1 and n_2 observations

$$y_{11}, y_{12}, \dots, y_{1n_1}$$

$$y_{21}, y_{22}, \dots, y_{2n_2}$$

$$\bar{y}_1 = \frac{y_{11} + \dots + y_{1n_1}}{n_1} \quad \bar{y}_2 = \frac{y_{21} + \dots + y_{2n_2}}{n_2}$$

- Is observed difference $\bar{y}_1 - \bar{y}_2$ "unusual" if $\mu_1 = \mu_2$?

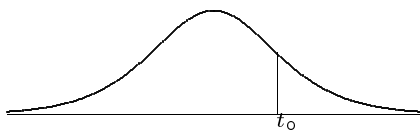
Use $t_o = (\bar{y}_1 - \bar{y}_2) / S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where

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

3-1

Assumptions

1. Independent observations
 2. Equal variances
 3. Normally distributed observations
- Assuming $H_0: \mu_1 = \mu_2$, these three assumptions define the distribution of t_o to be t-distributed with $n_1 + n_2 - 2$ degrees of freedom
 - "Unusual" then quantified by the probability that a randomly drawn t is more extreme than t_o (tail region of distribution)
 - Reject null hypothesis if this probability is "small". "Small" based on choice of significance level α



3-2

Example

(Samuels 7.36) In a study of lettuce growth, **ten seedlings were randomly allocated** to be grown in either a **standard nutrient solution** or in a **solution containing extra nitrogen**. After 22 days, the plants were harvested and weighed. The table below summarizes the results. Can we conclude that extra nitrogen enhances growth?

| Nutrient Solution | Leaf Dry Weight (gm) | | |
|-------------------|----------------------|------|------|
| | n | Mean | SD |
| Standard | 5 | 3.62 | 0.54 |
| Extra | 5 | 4.17 | 0.67 |

Solution: $S_p^2 = (4(.54)^2 + 4(.67)^2) / 8 = 0.37$. Our test statistic is then $t_o = (4.17 - 3.62) / \sqrt{2(.37)/5} = 1.43$. This question asks about a one-sided alternative. With 8 degrees of freedom, the P -value is between .05 and .10. From Table 2, $t_{.10} = 1.397$ and $t_{.05} = 1.860$. If α were greater than .10, we would reject the null and conclude that extra nitrogen enhances growth. If α were less than .05, we would **not reject the null** and conclude there is not sufficient evidence to state that the extra nitrogen enhances growth.

3-3

Statistical Model

Could also express the problem as

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \begin{cases} i = 1, 2 \\ j = 1, 2, \dots, n_i \end{cases}$$

where

μ = grand mean and τ_i is effect of treatment i

ϵ_{ij} is the random error component ($\epsilon_{ij} \sim N(0, \sigma^2)$)

Thus $\bar{y}_1, -\bar{y}_2 \sim N(\tau_1 - \tau_2, 2\sigma^2/n)$ when $n_1 = n_2 = n$

Can express Null in terms of treatment effects

$$H_0 : \tau_1 = \tau_2 = 0$$

H_1 : at least one τ_i different than 0

Will use this representation in the class

3-4

Type I and Type II errors

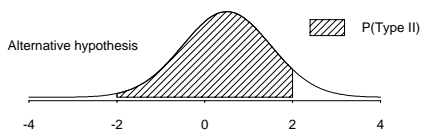
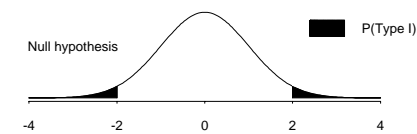
- In hypothesis testing, two types of errors

| | | TEST RESULT | |
|---------|-------|-------------|---|
| | | DNR | R |
| REALITY | H_0 | ♥ | I |
| | H_1 | II | ♥ |

- Type I error:** $\alpha = \Pr(\text{reject } H_0 | H_0 \text{ true})$
- Type II error:** $\beta = \Pr(\text{do not reject } H_0 | H_0 \text{ false})$
- Power of test (for specific H_1) is $1 - \beta$
- Significance level is α (this defines "unusual")

3-5

Choice of Sample Size/Computing Power



- Goal of test:** Detect diff of size δ with high prob

$$|\tau_1 - \tau_2| = \delta$$

- Choice of δ subjective (practical significance)
- Probability to detect difference is power
- Power depends on α , δ , σ , and n

3-6

Power/Sample Size Calculations

- Can form **Operating Characteristic Curve** (Power curve) for different levels of α , δ/σ and n
 - If σ known, use Normal distribution in calculations
 - If σ to be estimated, use non-central t (or table)

- Assume σ is known and $n_1 = n_2 = n$

$$H_0 : \bar{Y}_1 - \bar{Y}_2 \sim N(0, 2\sigma^2/n)$$

$$H_1 : \bar{Y}_1 - \bar{Y}_2 \sim N(\delta, 2\sigma^2/n)$$

Reject if (use H_0 dist)

$$\begin{aligned} \bar{Y}_1 - \bar{Y}_2 &> z_{\alpha/2} \sqrt{2\sigma^2/n} \\ \text{or} \\ \bar{Y}_1 - \bar{Y}_2 &< -z_{\alpha/2} \sqrt{2\sigma^2/n} \end{aligned}$$

Power: $\Pr(\text{Reject when } H_1 \text{ true})$ (use H_1 dist)

$$P(Z > z_{\alpha/2} - \delta/\sqrt{2\sigma^2/n}) + P(Z < -z_{\alpha/2} - \delta/\sqrt{2\sigma^2/n})$$

3-7

Example assuming known variance

Suppose $\alpha = .05$, $\sigma^2=12.5$, $n = 25$, and $\delta = 3.5$

- $z_{\alpha/2} = 1.96$ and $2\sigma^2/25 = 1$

$$\begin{aligned} \text{Power} &= \Pr(Z > 1.96 - 3.5) + \Pr(Z < -1.96 - 3.5) \\ &= .9382 + .0000 \\ &= .9382 \end{aligned}$$

- Can also use OCC (Figure 2-12 page 41)

This assumes σ is unknown so underestimates power

$$d = 3.5/2\sqrt{12.5} = .4950$$

$$n^* = 49$$

$$\text{Power} \approx 1 - .1 = .9$$

3-8

Power Calculations (σ estimated)

$$H_0 : |\tau_1 - \tau_2| = 0$$

$$H_1 : |\tau_1 - \tau_2| = \delta$$

Reject if:

$$\begin{aligned} \bar{Y}_1 - \bar{Y}_2 &> t_{2(n-1), 1-\alpha/2} \sqrt{2S_p^2/n} \\ \text{or} \\ \bar{Y}_1 - \bar{Y}_2 &< t_{2(n-1), \alpha/2} \sqrt{2S_p^2/n} \end{aligned}$$

Power: $\Pr(\text{reject} \mid H_1)$

$$\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{2S_p^2/n}} \sim t_{2(n-1)}(\delta/\sqrt{2\sigma^2/n})$$

Noncentral parameter $\delta/\sqrt{2\sigma^2/n}$

Compute probability of rejection given noncentral t

3-9

Using SAS - Example on Page 42-43

tpower.sas

```
/* Data set that contains variables common to both procedures */
data new; input alpha sigma;
cards;
.05 .25
;

/* Figure 1: Compute a power curve */

data new1; set new;
n=9; do delta = 0 to 1 by .10;
df = 2*(n-1); nc = delta/(sigma*sqrt(2/n));
rlow = tinv(alpha/2,df); rhigh = tinv(1-alpha/2,df);
p=1-probt(rhigh,df,nc)+probt(rlow,df,nc); output;
end;

symbol1 v=circle i=sm5; title1 'Power Curve I for t-test';
axis1 label=('prob'); axis2 label=('Difference in Means');
proc gplot; plot p*delta / haxis=axis2 vaxis=axis1; run;

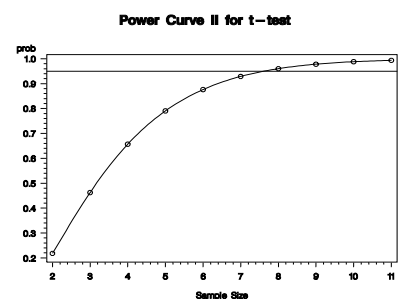
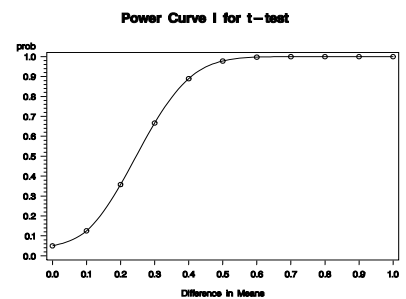
/* Figure 2: Find appropriate sample size */

data new2; set new;
delta=.5; do n=2 to 11 by 1;
df = 2*(n-1); nc = delta/(sigma*sqrt(2/n));
rlow = tinv(alpha/2,df); rhigh = tinv(1-alpha/2,df);
p=1-probt(rhigh,df,nc)+probt(rlow,df,nc); output;
end;

symbol1 v=circle i=sm5; title1 'Power Curve II for t-test';
axis1 label=('prob'); axis2 label=('Sample Size');
proc gplot; plot p*n / haxis=axis2 vaxis=axis1 vref=0.95; run;
```

3-10

Ouput



3-11

Confidence Intervals

- Besides $\delta = \bar{y}_1 - \bar{y}_2$, want statement of accuracy
- $\bar{\delta} \pm t_{\alpha/2} S_p \sqrt{1/n_1 + 1/n_2}$ (100(1- α)% confidence interval)
- In long run, true difference δ will be contained in 100(1- α)% of the intervals
- Are 100(1- α)% confident your single CI is one that contains the true difference δ
- Consider two-sided hypothesis test w/ level α
 - Reject if $|\bar{y}_1 - \bar{y}_2| > t_{\alpha/2} S_p \sqrt{1/n_1 + 1/n_2}$
- Consider 100(1 - α)% CI
 - Half-width of CI is $t_{\alpha/2} S_p \sqrt{1/n_1 + 1/n_2}$
 - 0 not in interval if $|\bar{y}_1 - \bar{y}_2| > t_{\alpha/2} S_p \sqrt{1/n_1 + 1/n_2}$
- Will reject H_0 if 0 not in confidence interval
- Can immediately test any $H_0 : \delta = \delta_0$ at level α

3-12

Paired Comparison

- Can often improve precision by pairing
- Removes explainable variation from the analysis
- Like material in each population
 - Twins for drug/health studies
 - Same specimen given both trts
 - Similar plots in a field
- Look at difference between **each** pair
- Changing 2n observations into n indep obs

$$d_i = y_{1i} - y_{2i}$$

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

$$t_0 = \bar{d} / (S_d / \sqrt{n})$$

$$t_0 \sim t_{n-1}$$

3-13

Statistical Model

- Pairing considered additive block effect

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \begin{cases} i = 1, 2 \\ j = 1, 2, \dots, n \end{cases}$$

where β_j is the pair j effect.

- $E(\bar{y}_1 - \bar{y}_2)$ still $\tau_1 - \tau_2$ because β cancels
- Two sample t-test vs Paired t-test
 - Trade off between df and variance reduction

| | | |
|-----------|------------------------------|--|
| | Two sample | Paired |
| Variance: | $\frac{2\sigma^2/n}{2(n-1)}$ | $\frac{2(\sigma^2 - \text{Cov}(Y_1, Y_2))/n}{(n-1)}$ |
| DF: | | |

Pairing advantageous when positive correlation. If correlation slight relative to σ^2 , then loss of dfs may result in blocking being a disadvantage.

3-14

Example

Paired T-test/Randomization Paired Test

In a study of egg cell maturation, the eggs from each of four female frogs were divided into two batches and one batch was exposed to progesterone. After two minutes, the cAMP content was measured. It is believed that cAMP is a substance that can mediate cellular response to hormones.

| FROG | cAMP Content | | Diff |
|------|--------------|--------------|------|
| | Control | Progesterone | |
| 1 | 6 | 4 | 2 |
| 2 | 4 | 5 | -1 |
| 3 | 5 | 2 | 3 |
| 4 | 4 | 2 | 2 |

- t-test:** $d = \{2, -1, 3, 2\} \rightarrow \bar{d} = 1.5$ and $s_d^2 = .866$. The test statistic is 1.732. Using Table II and 3 degrees of freedom, the P-value is between .05 and .10 (one-sided), .10 and .20 (two-sided). The actual two-sided P-value is close to 0.18.
- randomization:** The result of each pair does not depend on the allocation of treatments. Thus there are $2^4 = 16$ possible outcomes. The observed outcome is 2-1+3+2=6.

| | |
|------------|------------------|
| $ \sum d $ | # of occurrences |
| 8 | 2 |
| 6 | 2 |
| 4 | 4 |
| 2 | 6 |
| 0 | 2 |

From the table, there are four of sixteen outcomes as or more "unlikely" simply due to chance. Thus the P-value is 0.25.

3-15