Lecture 2: Simple Comparative Experiments

Montgomery: Section 2.4 and 2.5

Two-Sample t-Test

- $H_0 : \mu_1 = \mu_2$ (Null Hypothesis)

- $H_1 : \mu_1 \begin{array}{c} < \\ > \\ \neq \end{array} \mu_2$ (Alternative Hypothesis)

- Collect data: n_1 and n_2 observations

$$y_{11}, y_{12}, \ldots, y_{1n_1} \qquad y_{21}, y_{22}, \ldots, y_{2n_2}$$

$$\overline{y}_1 = \frac{y_{11} + \ldots + y_{1n_1}}{n_1} \qquad \overline{y}_2 = \frac{y_{21} + \ldots + y_{2n_2}}{n_2}$$

- Is observed difference $\overline{y}_1 - \overline{y}_2$ “unusual” if $\mu_1 = \mu_2$?

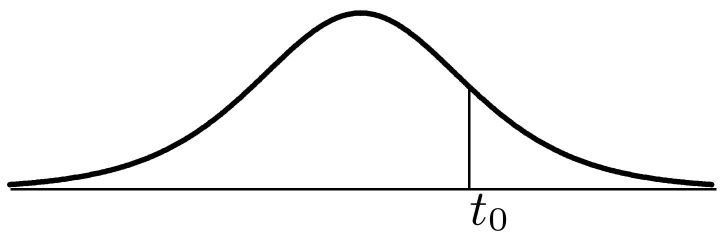
Use $$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ where

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

Assumptions

1. Independent observations
2. Equal variances ($\sigma_1^2 = \sigma_2^2$)
3. Normally distributed observations

- Assuming H_0: $\mu_1 = \mu_2$, these three assumptions define the distribution of t_0 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_0 (tail region of distribution)
- Reject null hypothesis if this probability is "small". "Small" based on choice of significance level α



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?

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

Solution: $S^2_{pool} = (4(.54)^2 + 4(.67)^2)/8 = 0.37$. Our test statistic is then $t_0 = (4.17 - 3.62)/\sqrt{2(.37)/5} = 1.43$. 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.

Statistical Model

Could also express the problem as

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

where

μ = grand mean (so $\tau_1 + \tau_2 = 0$) and τ_i is effect of treatment i

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

Thus $\overline{y}_{1.} - \overline{y}_{2.} \sim N(\tau_1 - \tau_2, 2\sigma^2/n)$ 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

Type I and Type II Errors

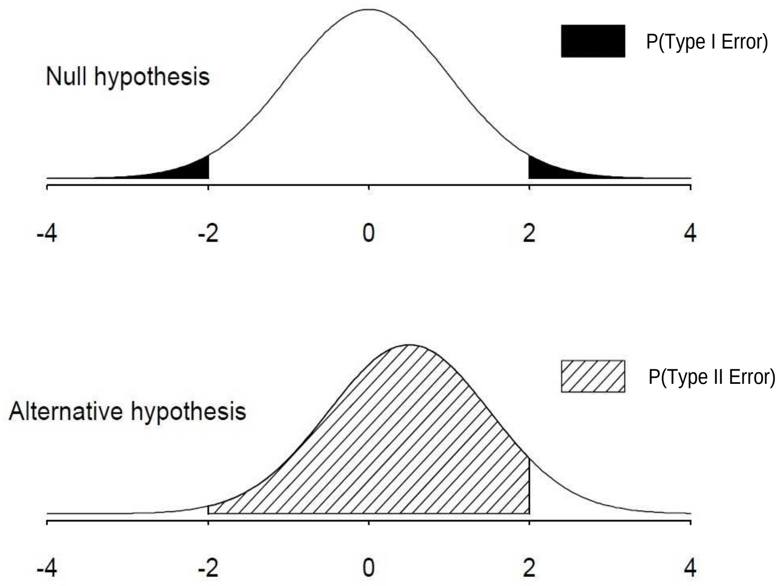
- In hypothesis testing, two types of errors

TEST RESULT

REALITY		DNR	R
	H_0	$\heartsuit$	I
	H_1	II	$\heartsuit$

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

Choice of Sample Size/Computing Power



- **Goal of test:** Detect diff of size $\delta = \tau_1 - \tau_2$ with high prob
- Choice of δ subjective (practical significance)
- Probability to detect difference is power
- Power depends on α, δ, σ, and n

Power/Sample Size Calculations

- Can form **O**perating **C**haracteristic **C**urve (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 $n_1 = n_2 = n$

  $H_0 : \overline{Y}_1 - \overline{Y}_2 \sim N(0, 2\sigma^2/n) \qquad H_1 : \overline{Y}_1 - \overline{Y}_2 \sim N(\delta, 2\sigma^2/n)$

  Reject H_0 if $|\bar{Y}_1 - \bar{Y}_2|/\sqrt{2\sigma^2/n} > z_{\alpha/2}$.

  Power: Pr(Reject $H_0|H_1$ is true), i.e.,

$$
\begin{aligned}
& P(|\bar{Y}_1 - \bar{Y}_2|/\sqrt{2\sigma^2/n} > z_{\alpha/2}|H_1) \\
= \quad & P(|(\bar{Y}_1 - \bar{Y}_2 - \delta)/\sqrt{2\sigma^2/n} + \delta/\sqrt{2\sigma^2/n}| > z_{\alpha/2}|H_1) \\
= \quad & P(|Z + \delta/\sqrt{2\sigma^2/n}| > z_{\alpha/2}) \\
= \quad & P(Z > z_{\alpha/2} - \delta/\sqrt{2\sigma^2/n}) + P(Z < -z_{\alpha/2} - \delta/\sqrt{2\sigma^2/n})
\end{aligned}
$$

Example (Assuming Known Variance)

Suppose $\alpha = .05$, σ^2=12.5, and $\delta = 3.5$

- For $n = 25$, $z_{\alpha/2} = 1.96$ and $2\sigma^2/n = 1$. The power is

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

- Find the smallest n such that the power $1 - \beta \geq 95\%$
  - Need $P(|Z + \delta/\sqrt{2\sigma^2/n}| > z_{\alpha/2}) \geq 0.95$.
  - Equivalently, $P(|Z + 0.7\sqrt{n}| \geq 1.96) \geq 0.95$
  - When $n = 26$, the power is 0.9462;
  - When $n = 27$, the power is 0.9533;
  - So, the sample size is chosen to be $n = 27$.

Power Calculations (σ Estimated)

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

$$H_1 : \tau_1 - \tau_2 = \delta \neq 0$$

- Reject H_0, if:

$$\overline{Y}_1 - \overline{Y}_2 > t_{\alpha/2,2(n-1)} \sqrt{2S^2_{pool}/n}$$

$$\text{or}$$

$$\overline{Y}_1 - \overline{Y}_2 < -t_{\alpha/2,2(n-1)} \sqrt{2S^2_{pool}/n}$$

- Power: Pr(reject $H_0 \mid H_1$), compute probability of rejection given noncentral t

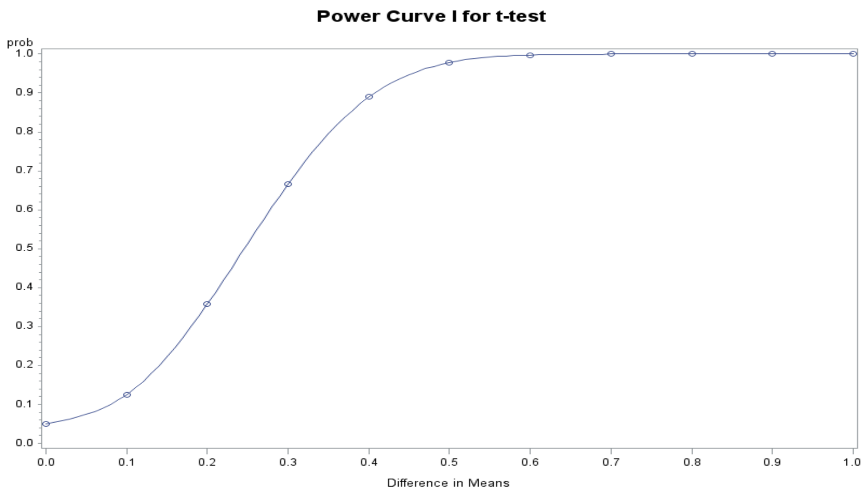
$$\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{2S^2_{pool}/n}} \sim t_{2(n-1)}(\delta/\sqrt{2\sigma^2/n})$$

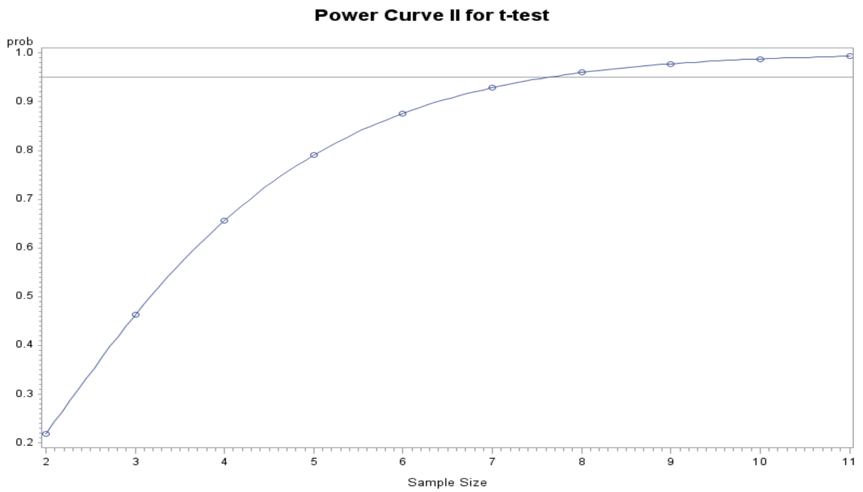
Noncentral parameter $\delta/\sqrt{2\sigma^2/n}$

Using SAS - Example on Page 46-47 (tpower.sas)

```
/* Figure 1: Compute a power curve */
data new1; alpha=.05; sigma=.25; n=9; df=2*(n-1);
  do delta = 0 to 1 by .10;
     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; alpha=.05; sigma=.25; delta=.5;
  do n=2 to 11 by 1;
     nc=delta/(sigma*sqrt(2/n)); df = 2*(n-1);
     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;
```

Ouput


Power Curve I for t-test
prob
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Difference in Means


Power Curve II for t-test
prob
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
2 3 4 5 6 7 8 9 10 11
Sample Size

Confidence Intervals

- Besides $\hat{\delta} = \overline{y}_1 - \overline{y}_2$, want statement of accuracy
- $\hat{\delta} \pm t_{\alpha/2}\mathrm{S}_{pool}\sqrt{1/n_1 + 1/n_2}$ (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 H_0 if $|\overline{y}_1 - \overline{y}_2| > t_{\alpha/2}S_{pool}\sqrt{1/n_1 + 1/n_2}$
- Consider $100(1-\alpha)$% CI
  - Half-width of CI is $t_{\alpha/2}S_{pool}\sqrt{1/n_1 + 1/n_2}$
  - 0 not in interval if $|\overline{y}_1 - \overline{y}_2| > t_{\alpha/2}S_{pool}\sqrt{1/n_1 + 1/n_2}$
- Will reject H_0 if 0 not in confidence interval
- Can immediately test any $H_0 : \delta = \delta_0$ at level α

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 independent observations

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

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

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

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

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.

- $E(\overline{y}_{1.} - \overline{y}_{2.})$ 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:	$2\sigma^2/n$	$2(\sigma^2 - \mathrm{Cov}(Y_1, Y_2))/n$
DF:	$2(n-1)$	$(n-1)$

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

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		
	Control	Progesterone	Diff
1	6	4	2
2	4	5	-1
3	5	2	3
4	4	2	2

- **t-test:** $d = \{2, -1, 3, 2\} \rightarrow \overline{d} = 1.5$ and $s_{\overline{d}} = .866$. 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 Test:** 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.

$\lvert \sum_{i=1}^{4} d_i \rvert$	# 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.