Statistics 514: Design of Experiments

Topic 3

Topic Overview

This topic will cover

- t-tests (Montgomery 2-4, 2-5)
- Power calculations (Montgomery 2-4.2)

Two-sample t-test

- $H_0 : \mu_1 = \mu_2$ *(Null Hypothesis)*
- $H_1 : \mu_1 \neq \mu_2$ *(Alternative Hypothesis)* or $H_1 : \mu_1 < \mu_2$ or $H_1 : \mu_1 > \mu_2$
- Collect data – n_1 and n_2 observations

$$y_{1,1}, y_{1,2}, \cdots, y_{1,n_1}$$
$$y_{2,1}, y_{2,2}, \cdots, y_{2,n_2}$$

$$\bar{y}_1 = \frac{y_{1,1} + \ldots + y_{1,n_1}}{n_1} \quad \bar{y}_2 = \frac{y_{2,1} + \ldots + y_{2,n_2}}{n_2}$$

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

  Use $t_0 = (\bar{y}_1 - \bar{y}_2)/S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, where

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

Statistical Model

Could also express the problem as

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

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 class

Assumptions

1. Independent observations
2. Equal variances
3. Normally distributed observations

Mechanics

- 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.

Normality

- Helped by CLT
- Skewness mitigated by differencing $\bar{Y}_i$'s
- Not much help by using t over Normal.

Common variance

Can be more important than normality. Choosing a balanced design $n_1 = n_2$ mitigates.

Alternatives

- **Analysis:** Nonparametric alternatives: bootstrap; empirical likelihood; Kruskal-Wallis; assume different variances:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- **Design:** *Paired t-test*
  - Can often improve precision by pairing (form of blocking)
  - Removes explainable variation from the analysis

– Changing $2n$ observations into n independent observations

$$\begin{aligned} d_i &= y_{i,1} - y_{i,2} \\ S_d^2 &= \frac{1}{n-1}\sum(d_i - \bar{d})^2 \\ t_0 &= \bar{d}/(S_d/\sqrt{n}) \\ t_0 &\sim t_{n-1} \end{aligned}$$

Statistical Model for Paired Comparisons

- Pairing considered additive block effect

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

- Note: Equal variances not necessary.
- $\mathrm{E}(\bar{y}_{1.} - \bar{y}_{2.})$ still equals $\tau_1 - \tau_2$ 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 samples are positively correlated. If correlation slight relative to σ^2, then loss of df's may result in block being a disadvantage.

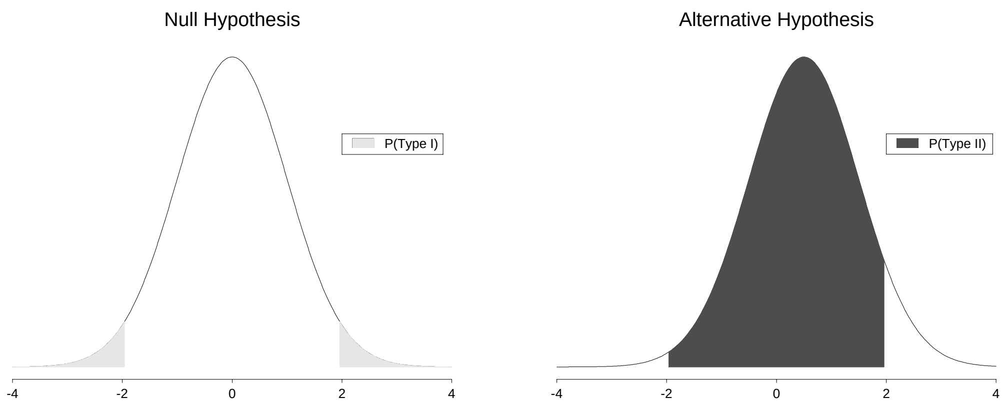
Type I and Type II Errors

- In hypothesis testing, two types of errors:

REALITY

TEST RESULT	H_0	H_1
DNR	♡	II
R	I	♡

- *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")



Choice of Sample Size/Computing Power

- **Goal of test**: Detect differences of size δ with high probability

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

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

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 $n_1 = n_2 = n$.

$$\begin{aligned} \mathrm{H}_0 : \bar{Y}_1 - \bar{Y}_2 &\sim N(0, 2\sigma^2/n) \\ \mathrm{H}_1 : \bar{Y}_1 - \bar{Y}_2 &\sim N(\delta, 2\sigma^2/n) \end{aligned}$$

Reject if (using H_0 distribution)

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

Power: Pr(Reject when H_1 true) (use H_1 distribution)

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

Example assuming known variance

Suppose $\alpha = 0.05$, $\sigma^2 = 12.5$, $n = 25$, and $\delta = 3.5$.

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

$$\begin{aligned}\text{Power} &= \Pr(Z > 1.96 - 3.5) + \Pr(Z < -1.96 - 3.5)\\ &= 0.9382 + 0.0000\\ &= 0.9382\end{aligned}$$

- Can also use OCC (Figure 2-12 page 41)
  This assumes σ is unknown so underestimates power.
  $d = 3.5/2\sqrt{12.5} = 0.4950$
  $n^* = 49$
  Power $\approx 1 - 0.1 = 0.9$

Power Calculations (σ estimated)

$$\text{H}_0 : \tau_1 - \tau_2 = 0$$
$$\text{H}_1 : |\tau_1 - \tau_2| = \delta$$

Reject if

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

Power: $\Pr(\text{reject} \mid \text{H}_1)$

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

$$\text{Noncentrality parameter: } \delta/\sqrt{2\sigma^2/n}$$

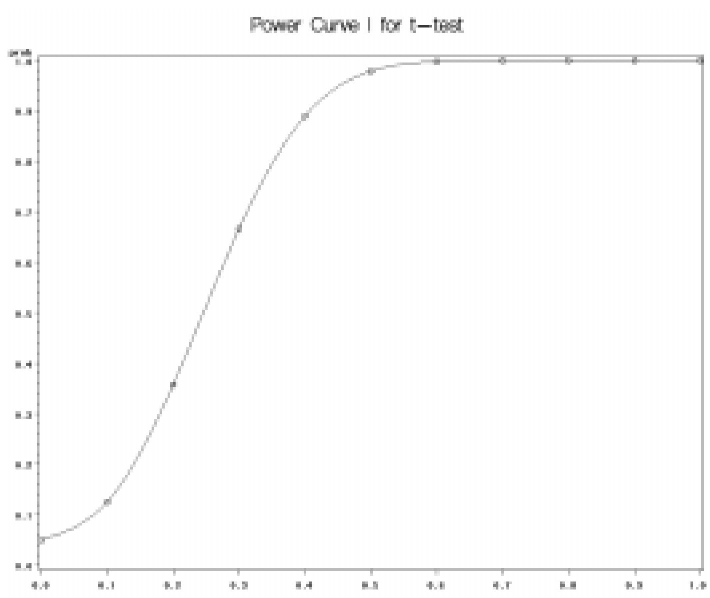
Compute probability of rejection given noncentral t.

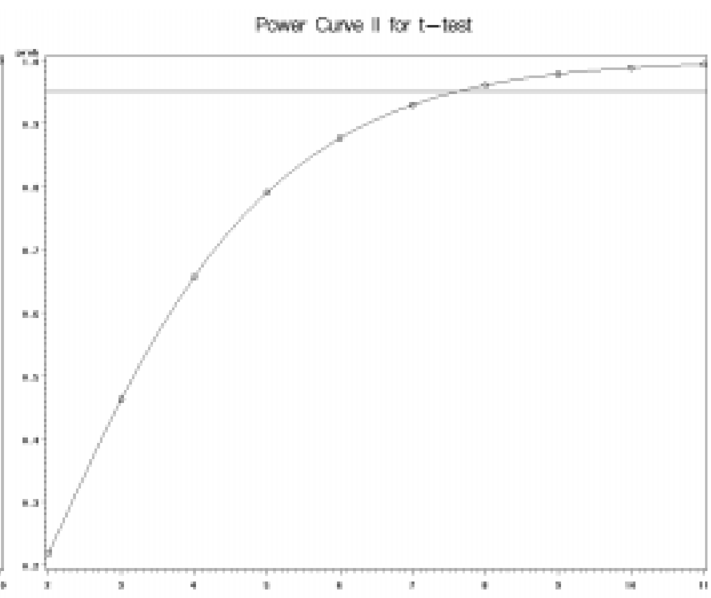
Using SAS - Example on Page 42-43 (`tpower.sas`)

```
/* Data set that contains variables common to both procedures */
data new; input alpha sigma;
 cards;
   0.05  0.25
 ;

/* Figure 1: Compute a power curve */
```

```
data new1; set new;
 n=9; do delta = 0 to 1 by 0.10;
            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 = 0.5; 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;
```

Output





Using R

```
# Input variables to both procedures
alpha = 0.05
sigma = 0.25

# Figure 1: Compute a power curve
n = 9
p = rep(0, 11)                  ## Initialize p
delta = seq(0, 1, 0.1)          ## Initialize delta
for(i in 1:11) {
  del = delta[i]
  df = 2*(n-1)
  nc = del/(sigma*sqrt(2/n))    ## non-centrality parameter
  rlow = qt(alpha/2, df)        ## lower endpoint of acceptance region
  rhigh = qt(1-alpha/2, df)     ## upper endpoint of acceptance region
  p[i] = 1-pt(rhigh - nc, df) + pt(rlow - nc, df)
}

# Plot power curve
plot(delta, p,
     xlab = "Delta", ylab = "Power")


# Figure 2: Find appropriate sample size
delta = 0.5            ## Initialize delta
p = rep(0, 10)         ## Initialize p
for(n in 2:11) {
  df = 2*(n-1)
  nc = delta/(sigma*sqrt(2/n))
  rlow = qt(alpha/2, df)
  rhigh = qt(1-alpha/2, df)
  p[n-1] = 1 - pt(rhigh - nc, df) + pt(rlow - nc, df)
}

# Plot the power curve
plot(2:11, p,
     xlab = "Sample Size", ylab = "Power")
abline(h = 0.95)
```

Confidence Intervals

- Besides $\hat{\delta} = \bar{y}_1 - \bar{y}_2$, want statement of accuracy
- $\hat{\delta} \pm t_{\alpha/2} S_p \sqrt{1/n_1 + 1/n_2}$ ($100(1-\alpha)\%$ confidence interval)
- In long run, true difference δ will be contained in $100(1-\alpha)\%$ of the intervals.

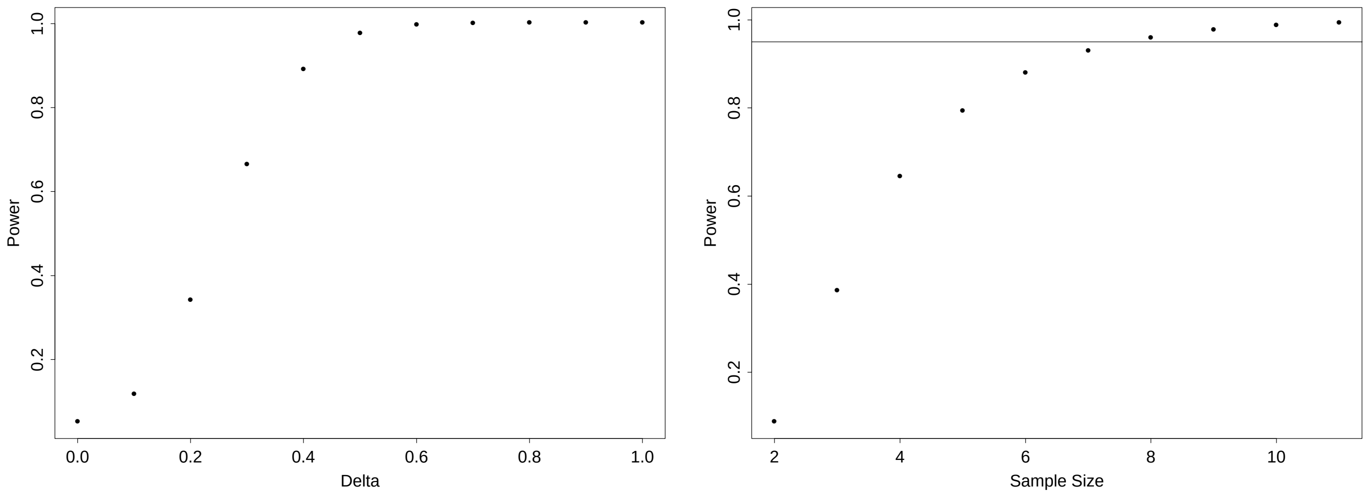


Figure 1: Output of R code

- Are $100(1-\alpha)\%$ confident your single CI is one that contains the true difference δ.
- Consider two-sided hypothesis test with level α.
  - 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.