Hypothesis tests and confidence intervals

The 95% confidence interval for μ is the set of values, μ_0, such that the null hypothesis $H_0 : \mu = \mu_0$ would not be rejected (by a two-sided test with α = 5%).

The 95% CI for μ is the set of plausible values of μ.

If a value of μ is plausible, then as a null hypothesis, it would not be rejected.

For example: 9.98 9.87 10.05 10.08 9.99 9.90

(assumed iid normal(μ,σ).)

$\bar{X}$ = 9.98; s = 0.082; n = 6 `qt(0.975,5)` = 2.57

95% CI for μ = 9.98 $\pm$ 2.57 $\cdot$ 0.082 / $\sqrt{6}$

= 9.98 $\pm$ 0.086 = (9.89,10.06)

Sample size calculations

$$n = \frac{\text{\$ available}}{\text{\$ per sample}}$$

Too little data $\longrightarrow$ A total waste

Too much data $\longrightarrow$ A partial waste

Power

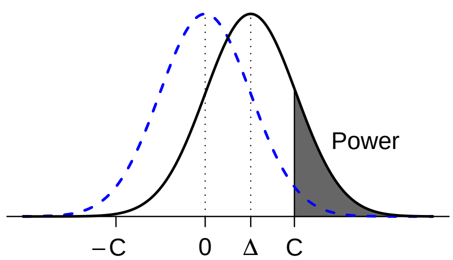
$X_1, \ldots, X_n$ iid normal(μ_A, σ_A) $Y_1, \ldots, Y_m$ iid normal(μ_B, σ_B)

Test $H_0 : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$ at $\alpha = 0.05$.

Test statistic: $T = \dfrac{\bar{X} - \bar{Y}}{\widehat{SD}(\bar{X} - \bar{Y})}$.

Critical value: C such that $\Pr(|T| > C \mid \mu_A = \mu_B) = \alpha$.

Power: $\Pr(|T| > C \mid \mu_A \neq \mu_B)$



Power depends on...

- The design of your experiment
- What test you're doing
- Chosen significance level, α
- Sample size
- True difference, $\mu_A - \mu_B$
- Population SD's, σ_A and σ_B.

The case of known population SDs

Suppose σ_A and σ_B are known.

Then $\bar{X} - \bar{Y} \sim$ normal($\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}$)

Test statistic: $Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}$

If H_0 is true ($\mu_A = \mu_B$), $Z \sim$ normal(0,1)

$\implies C = z_{\alpha/2}$ so that $\Pr(|Z| > C \mid \mu_A = \mu_B) = \alpha$

For example, for α = 0.05, C = `qnorm(0.975)` = 1.96.

Power when the population SDs are known

If $\mu_A - \mu_B = \Delta$, then $Z = \dfrac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \sim$ normal(0,1)

$$\Pr\left(\frac{|\bar{X}-\bar{Y}|}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} > 1.96\right) = \Pr\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} > 1.96\right) + \Pr\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} < -1.96\right)$$

$$= \Pr\left(\frac{\bar{X}-\bar{Y}-\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right) + \Pr\left(\frac{\bar{X}-\bar{Y}-\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}} < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right)$$

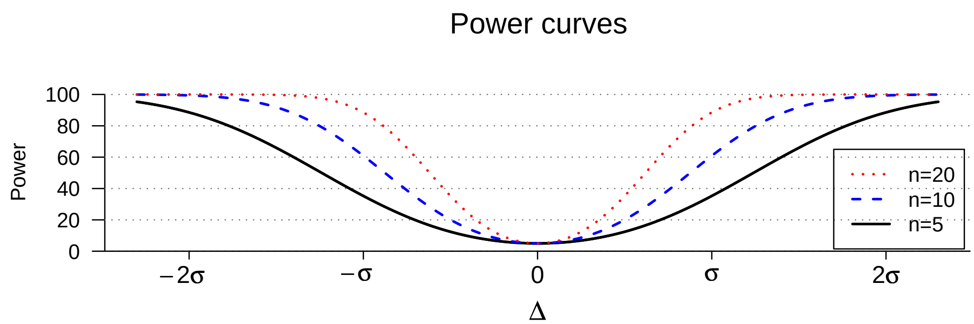
$$= \Pr\left(Z > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right) + \Pr\left(Z < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right)$$

Calculations in R

$$\text{Power} = \Pr\left(Z > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right) + \Pr\left(Z < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right)$$

```
C <- qnorm(0.975)
se <- sqrt( sigmaA^2/n + sigmaB^2/m )

power <- 1 - pnorm(C - delta/se) +
           pnorm(-C - delta/se)
```



Power depends on ...

$$\text{Power} = \Pr\left(Z > C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right) + \Pr\left(Z < -C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right)$$

- Choice of α (which affects C)

  Larger $\alpha \rightarrow$ less stringent $\rightarrow$ greater power.

- $\Delta = \mu_A - \mu_B =$ the true "effect."

  Larger $\Delta \rightarrow$ greater power.

- Population SDs, σ_A and σ_B

  Smaller σ's $\rightarrow$ greater power.

- Sample sizes, n and m

  Larger n, m $\rightarrow$ greater power.

Choice of sample size

We mostly influence power via n and m.

Power is greatest when $\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}$ is as small as possible.

Suppose the total sample size N = n + m is fixed.

$$\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} \text{ is minimized when } n = \frac{\sigma_A}{\sigma_A+\sigma_B}N \text{ and } m = \frac{\sigma_B}{\sigma_A+\sigma_B}N$$

For example:

If $\sigma_A = \sigma_B$, we should choose n = m.

If $\sigma_A = 2\ \sigma_B$, we should choose n = 2 m.

(e.g., if σ_A = 4 and σ_B = 2, we might use n=20 and m=10)

Calculating the sample size

Suppose we seek 80% power to detect a particular value of $\mu_A - \mu_B = \Delta$, in the case that σ_A and σ_B are known.

(For convenience here, let's pretend that $\sigma_A = \sigma_B$ and that we plan to have equal sample sizes for the two groups.)

$$\text{Power} \approx \Pr\left(Z > C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n}+\frac{\sigma_B^2}{m}}}\right) = \Pr\left(Z > 1.96 - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}\right)$$

$\longrightarrow$ Find n such that $\Pr\left(Z > 1.96 - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}\right) = 80\%$.

Thus $1.96 - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}} = \texttt{qnorm(0.2)}$ = –0.842.

$\Longrightarrow \sqrt{n} = \frac{\sigma}{\Delta}\,[1.96 - (-0.842)]\,\sqrt{2}$ $\Longrightarrow n = 15.7 \times (\frac{\sigma}{\Delta})^2$

Equal but unknown population SDs

$X_1, \ldots, X_n$ iid normal(μ_A, σ) $Y_1, \ldots, Y_m$ iid normal(μ_B, σ)

Test $H_0 : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$ at $\alpha = 0.05$.

$$\hat{\sigma}_p = \sqrt{\frac{s_A^2(n-1)+s_B^2(m-1)}{n+m-2}} \qquad \widehat{SD}(\bar{X} - \bar{Y}) = \hat{\sigma}_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

Test statistic: $T = \dfrac{\bar{X} - \bar{Y}}{\widehat{SD}(\bar{X} - \bar{Y})}$.

In the case $\mu_A = \mu_B$, T follows a t distribution with n + m – 2 d.f.

Critical value: C = `qt(0.975, n+m-2)`

Power: equal but unknown pop'n SDs

Power = $\Pr\left(\frac{|\bar{X}-\bar{Y}|}{\hat{\sigma}_p\sqrt{\frac{1}{n}+\frac{1}{m}}} > C\right)$

In the case $\mu_A - \mu_B = \Delta$,

the statistic $\frac{\bar{X}-\bar{Y}}{\hat{\sigma}_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$ follows a non-central t distribution.

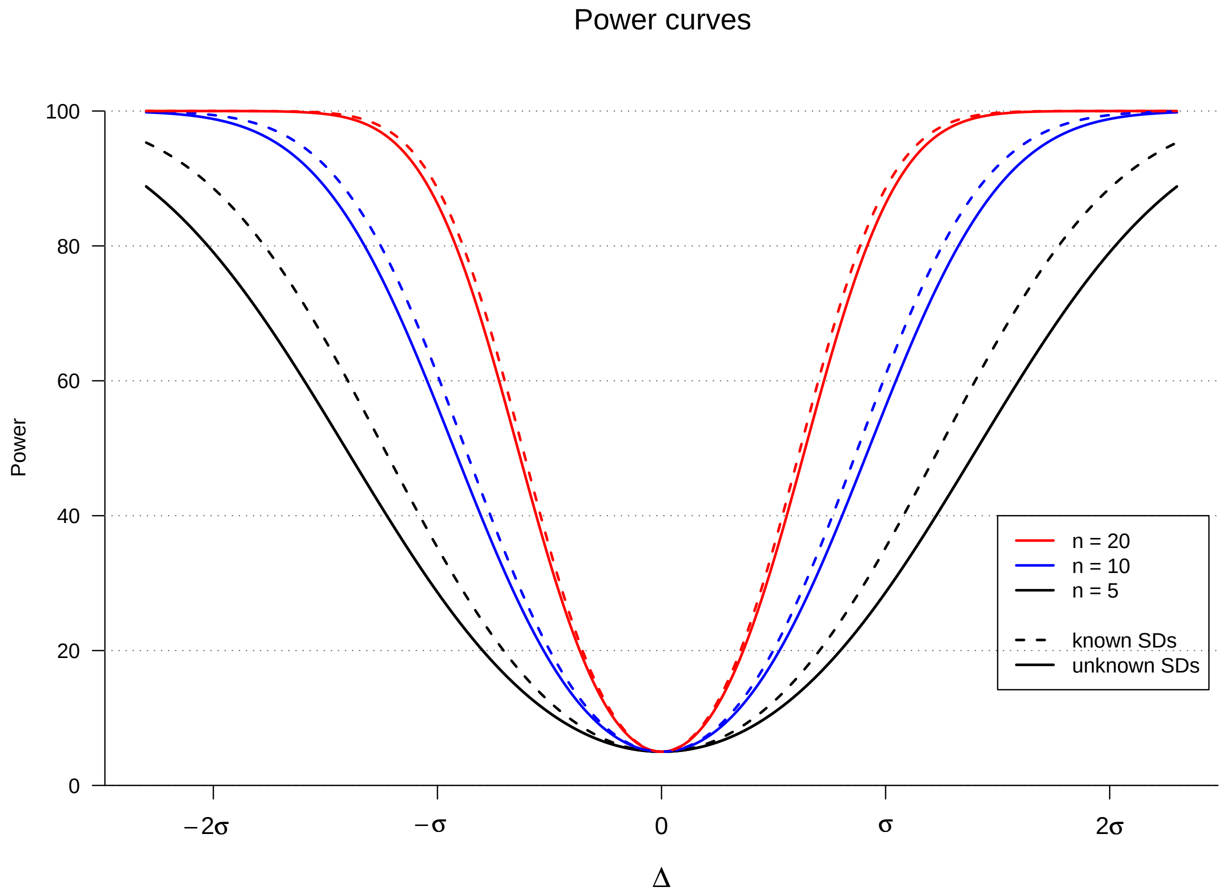
This distribution has two parameters:

degrees of freedom (as before)

the non-centrality parameter, $\frac{\Delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$

```
C <- qt(0.975, n + m - 2)
se <- sigma * sqrt( 1/n + 1/m )

power <- 1 - pt(C, n+m-2, ncp=delta/se) +
             pt(-C, n+m-2, ncp=delta/se)
```



A built-in function: power.t.test()

Calculate power (or determine the sample size) for the t-test when:

- Sample sizes equal
- Population SDs equal

Arguments:

- `n` = sample size
- `delta` = Δ = $\mu_2 - \mu_1$
- `sd` = σ = population SD
- `sig.level` = α = significance level
- `power` = the power
- `type` = type of data (two-sample, one-sample, paired)
- `alternative` = two-sided or one-sided test

Examples

A. n = 10 for each group; effect = Δ = 5; pop'n SD = σ = 10

```
power.t.test(n=10, delta=5, sd=10)
```

$\Longrightarrow$ 18%

B. power = 80%; effect = Δ = 5; pop'n SD = σ = 10

```
power.t.test(delta=5, sd=10, power=0.8)
```

$\Longrightarrow$ n = 63.8 $\Longrightarrow$ 64 for each group

C. power = 80%; effect = Δ = 5; pop'n SD = σ = 10; one-sided

```
power.t.test(delta=5, sd=10, power=0.8,
             alternative="one.sided")
```

$\Longrightarrow$ n = 50.2 $\Longrightarrow$ 51 for each group

Unknown and different pop'n SDs

$X_1, \dots, X_n$ iid normal(μ_A, σ_A) $Y_1, \dots, Y_m$ iid normal(μ_B, σ_B)

Test $H_0 : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$ at $\alpha = 0.05$.

Test statistic: $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{m}}}$

To calculate the critical value for the test, we need the null distribution of T (that is, the distribution of T if $\mu_A = \mu_B$).

To calculate the power, we need the distribution of T given the value of $\Delta = \mu_A - \mu_B$.

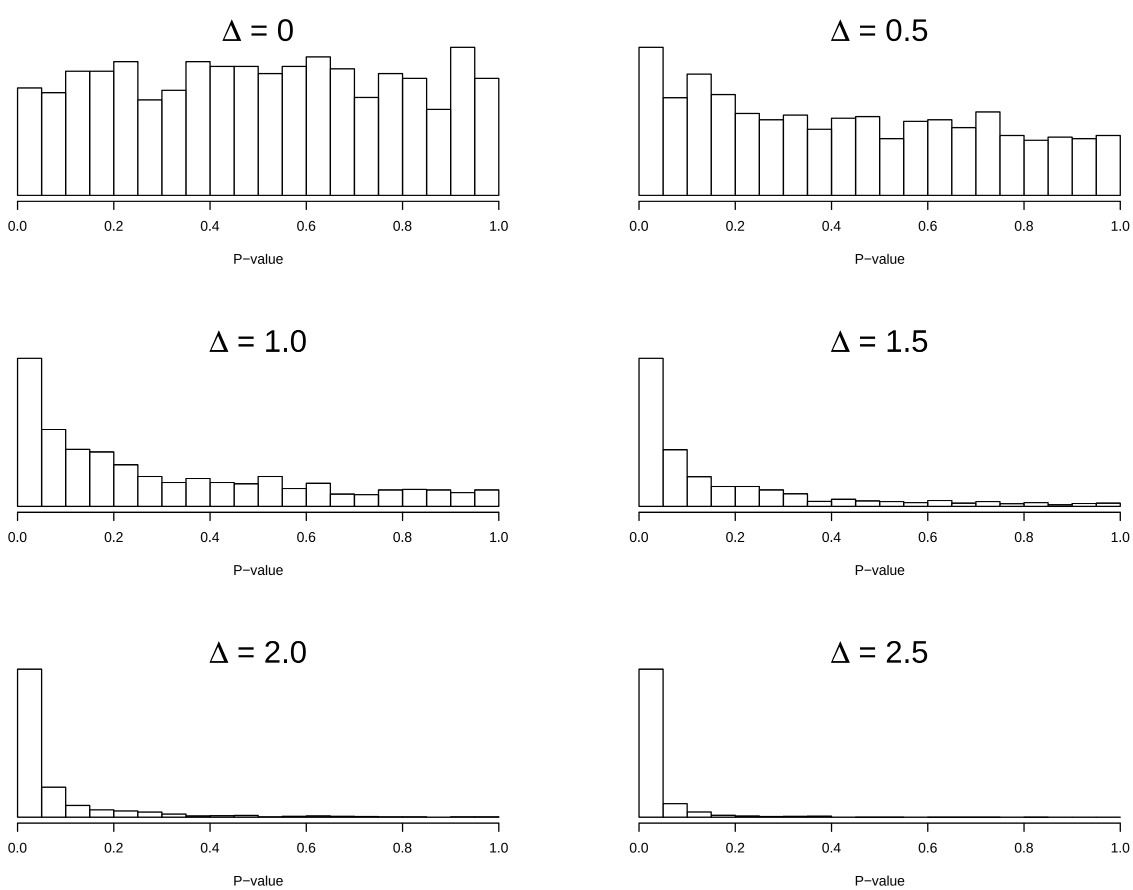
We don't really know either of these.

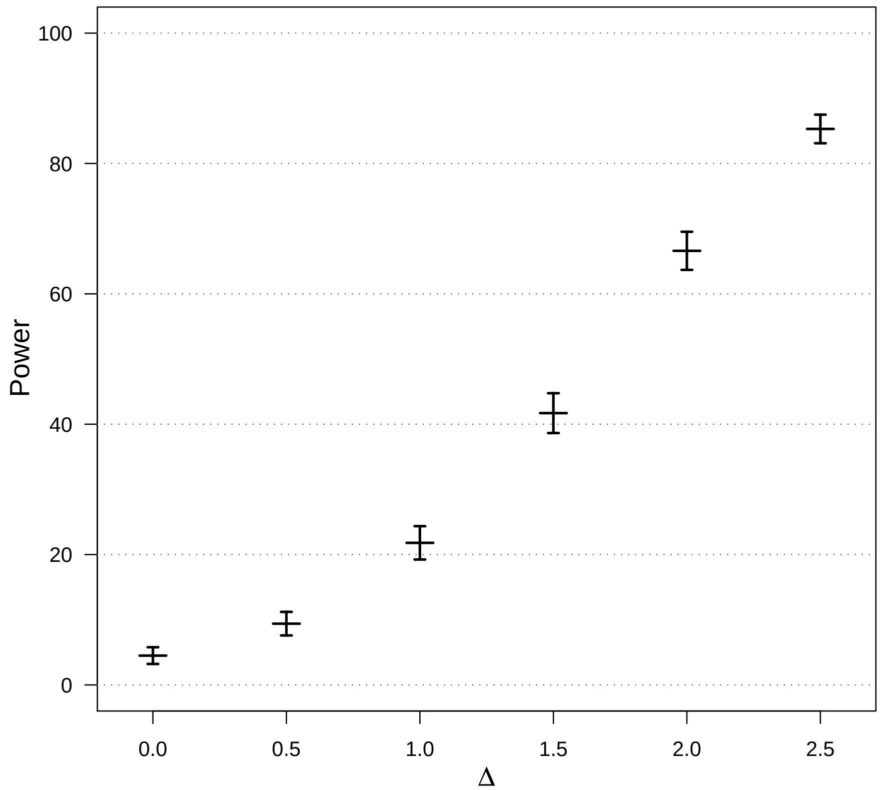
Power by computer simulation

- Specify n, m, σ_A, σ_B, and $\Delta = \mu_A - \mu_B$, and the significance level, α.
- Simulate data under the model.
- Perform the proposed test and calculate the P-value.
- Repeat many times.

Example: n = 5, m = 10, σ_A = 1, σ_B = 2,

Δ = 0.0, 0.5, 1.0, 1.5, 2.0 or 2.5.





Determining sample size

The things you need to know:

- Structure of the experiment
- Method for analysis
- Chosen significance level, α (usually 5%)
- Desired power (usually 80%)

- Variability in the measurements
  - If necessary, perform a pilot study, or use data from prior experiments or publications

- The smallest meaningful effect

Reducing sample size

- Reduce the number of treatment groups being compared.
- Find a more precise measurement (e.g., average survival time rather than proportion dead).
- Decrease the variability in the measurements.
  - Make subjects more homogenous.
  - Use stratification.
  - Control for other variables (e.g., weight).
  - Average multiple measurements on each subject.

Tests to compare two means

1. Assume $\sigma_1 \equiv \sigma_2$
   (a) Calculate pooled estimate of population SD
   (b) $\hat{\text{SE}} = \hat{\sigma}_{\text{pooled}} \sqrt{\frac{1}{n} + \frac{1}{m}}$
   (c) Compare to t(df = n + m – 2)

   In R: `t.test` with `var.equal=TRUE`

2. Allow $\sigma_1 \neq \sigma_2$
   (a) $\hat{\text{SE}} = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$
   (b) Compare to t with df from nasty formula.

   In R: `t.test` with `var.equal=FALSE` (the default)

Estimated type I error rates

$X_1, \ldots, X_4$ iid normal(μ, σ) $Y_1, \ldots, Y_4$ iid normal(μ, $\sigma \times \tau$)

10,000 simulations

$\tau = 1$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.948	0.000	0.948
Reject H_0	0.009	0.043	0.052
	0.957	0.043	

$\tau = 2$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.940	0.000	0.940
Reject H_0	0.012	0.048	0.060
	0.952	0.048	

$\tau = 1.5$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.944	0.000	0.944
Reject H_0	0.009	0.047	0.056
	0.953	0.047	

$\tau = 4$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.924	0.000	0.924
Reject H_0	0.023	0.054	0.076
	0.946	0.054	

Estimated power

$X_1, \ldots, X_4$ iid normal(μ, σ) $Y_1, \ldots, Y_4$ iid normal($\mu+2$, $\sigma \times \tau$)

10,000 simulations

$\tau = 1$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.344	0.000	0.344
Reject H_0	0.046	0.611	0.656
	0.389	0.611	

$\tau = 2$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.658	0.000	0.658
Reject H_0	0.060	0.282	0.342
	0.718	0.282	

$\tau = 1.5$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.532	0.000	0.532
Reject H_0	0.057	0.411	0.468
	0.589	0.411	

$\tau = 4$

Assume $\sigma_1 \equiv \sigma_2$	Allow $\sigma_1 \neq \sigma_2$: FTR H_0	Allow $\sigma_1 \neq \sigma_2$: Reject H_0	
FTR H_0	0.836	0.000	0.836
Reject H_0	0.047	0.117	0.164
	0.883	0.117	