

Sample size calculations

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

Too few animals \rightarrow A total waste

Too many animals \rightarrow A partial waste

Power

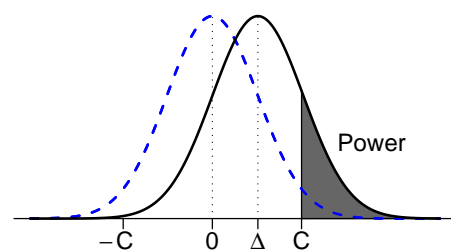
X_1, \dots, X_n iid normal(μ_A, σ_A) Y_1, \dots, Y_m iid normal(μ_B, σ_B)

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

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

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

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

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

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

For example, for $\alpha = 0.05$, $C = \text{qnorm}(0.975) = 1.96$.

Power when the population SDs are known

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

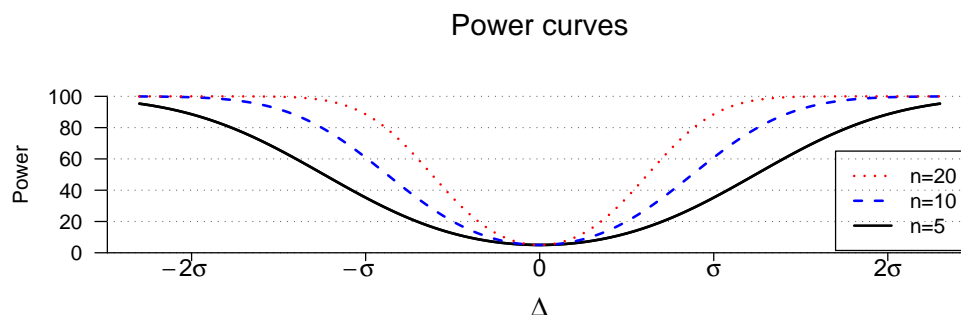
$$\begin{aligned} \Pr\left(\frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} > 1.96\right) &= \Pr\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} > 1.96\right) + \Pr\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}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

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 $\sigma_A = 4$ and $\sigma_B = 2$, 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.)

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

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

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

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

Equal but unknown population SDs

X_1, \dots, X_n iid normal(μ_A, σ)

Y_1, \dots, Y_m iid normal(μ_B, σ)

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

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

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

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

Critical value: $C = \text{qt}(0.975, n+m-2)$

Power: equal but unknown pop'n SDs

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

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.

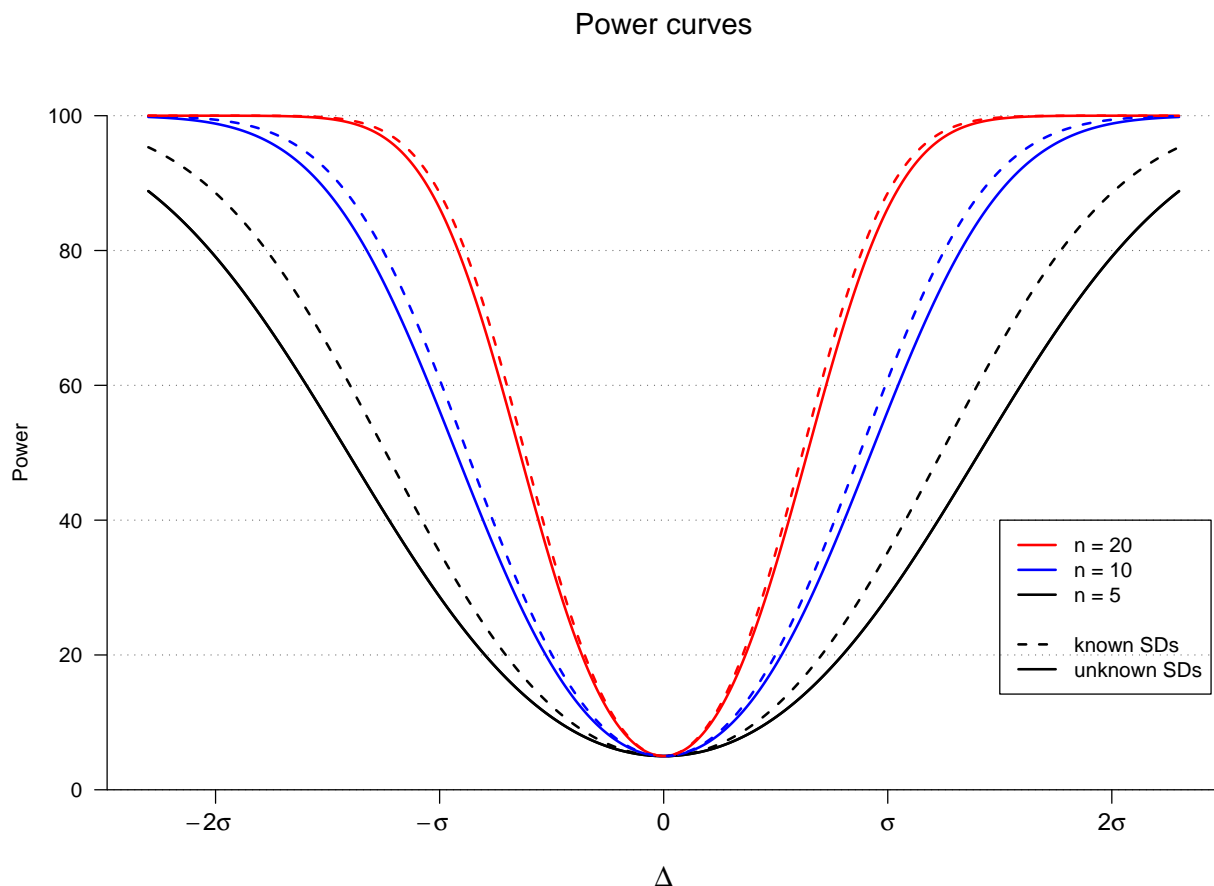
This distribution has two parameters:

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` = $\Delta = \mu_2 - \mu_1$
- `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 = $\Delta = 5$; pop'n SD = $\sigma = 10$

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

⇒ 18%

B. power = 80%; effect = $\Delta = 5$; pop'n SD = $\sigma = 10$

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

⇒ `n` = 63.8 ⇒ 64 for each group

C. power = 80%; effect = $\Delta = 5$; pop'n SD = $\sigma = 10$; one-sided

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

⇒ `n` = 50.2 ⇒ 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 $\Delta = \mu_A - \mu_B$.

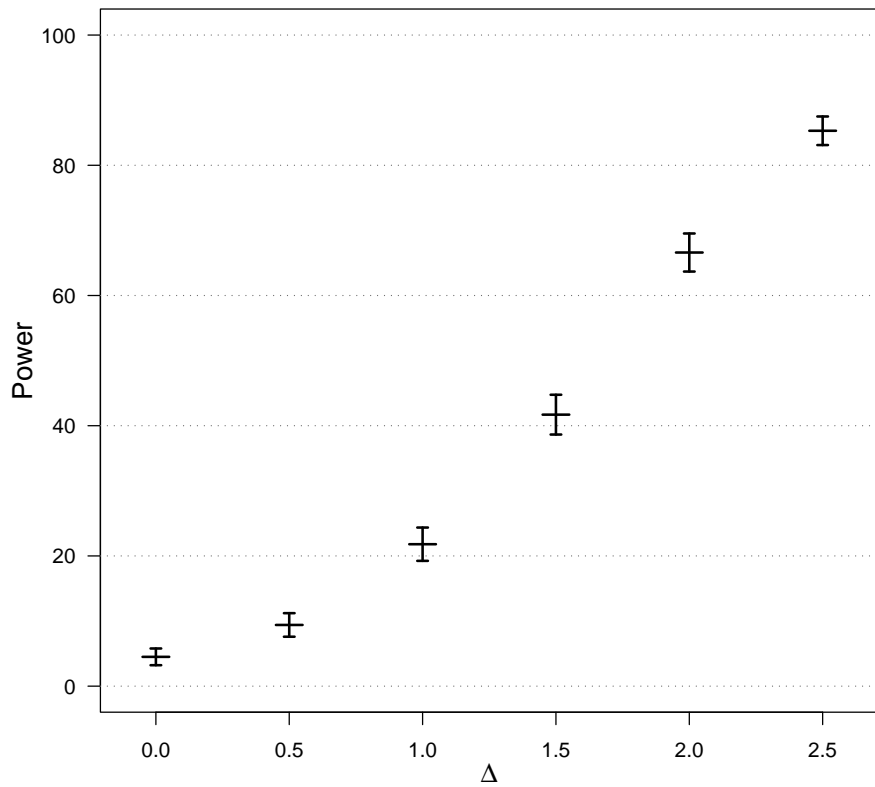
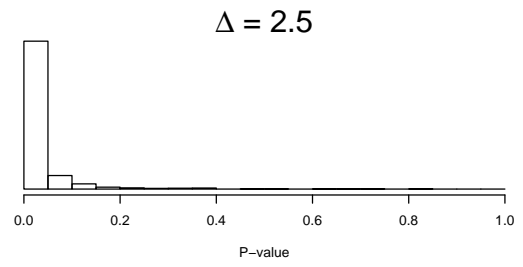
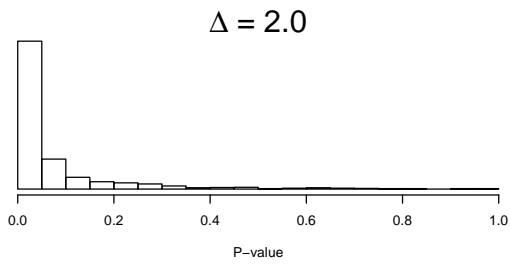
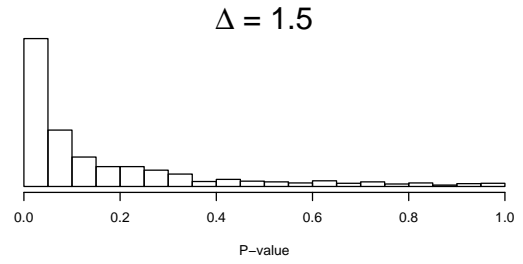
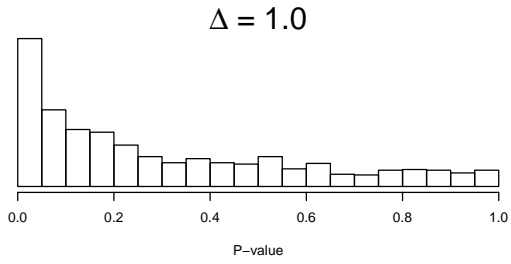
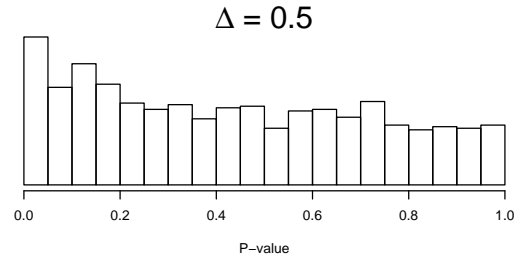
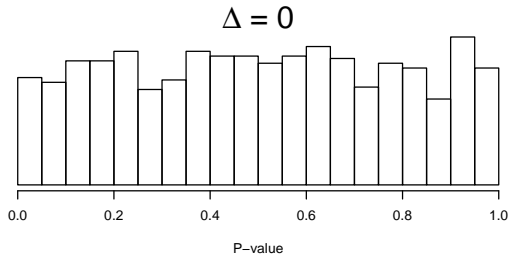
We don't **really** know either of these.

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: $n = 5, m = 10, \sigma_A = 1, \sigma_B = 2,$

$\Delta = 0.0, 0.5, 1.0, 1.5, 2.0$ 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.