Power analysis

Statistical decisions

Why 0.05?

\[ \alpha = 0.01 \]

Statistical decisions

<table>
<thead>
<tr>
<th>Statistical conclusion</th>
<th>Reject ( H_0 ) (conclude is effect)</th>
<th>Retain ( H_0 ) (conclude no effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really is an effect</td>
<td>Correct decision – effect detected</td>
<td>Type II error – effect not detected</td>
</tr>
<tr>
<td>Really is no effect</td>
<td>Type I error – effect not detected, none exists</td>
<td>Correct decision – no effect detected, none exists</td>
</tr>
</tbody>
</table>
**Statistical decisions**

- **Type I errors - when falsely (incorrectly) reject a null hypothesis**
  - Conclude that there is an effect, when there really is not
  - $\alpha$ - probability of a Type I error (0.05)
  - Minimize by setting $\alpha$ as low as possible

<table>
<thead>
<tr>
<th>Statistical conclusion</th>
<th>Reject $H_0$ (conclude is effect)</th>
<th>Retain $H_0$ (conclude no effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really is an effect</td>
<td>Correct decision – effect detected</td>
<td>Type II error – effect not detected</td>
</tr>
<tr>
<td>Really is no effect</td>
<td>Type I error – effect detected, none exists</td>
<td>Correct decision – no effect detected, none exists</td>
</tr>
</tbody>
</table>

- **Type II errors - when falsely (incorrectly) retain a null hypothesis**
  - Conclude that there is an no effect, when there really is an effect
  - $\beta$ - probability of a Type II error
  - Typically, approx 20%

<table>
<thead>
<tr>
<th>Statistical conclusion</th>
<th>Reject $H_0$ (conclude is effect)</th>
<th>Retain $H_0$ (conclude no effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really is an effect</td>
<td>Correct decision – effect detected</td>
<td>Type II error – effect not detected</td>
</tr>
<tr>
<td>Really is no effect</td>
<td>Type I error – effect detected, none exists</td>
<td>Correct decision – no effect detected, none exists</td>
</tr>
</tbody>
</table>

**Power of a test**

- Probability of detecting an effect if it exists
- Probability of correctly rejecting a false $H_0$
- $Power = 1 - \beta$ (probability of making a Type II error)
- Usually aim for power $\approx 0.8$
Statistical power depends on

- **Effect size (ES)**
  - Magnitude of the difference between treatments
  - Large differences (effect sizes) are easier to detect

- **Background variation (σ)**
  - Variation between sampling units
  - Estimated by sample standard deviation (s)
  - Greater background variability, less likely to detect effects

\[
\text{power}(1 - \beta) \propto \frac{ES}{\sigma}
\]

Statistical power depends on

- **Sample size (n) for each treatment group**
  - Increasing sample size makes effects easier to detect

- **Significance level (α)**
  - Type I error rate
  - Probability of falsely rejecting a H₀
  - As α decreases, β increases, power decreases
  - Usually set at 0.05

\[
\text{power}(1 - \beta) \propto \frac{ES \sqrt{n \alpha}}{\sigma}
\]

A priori power analysis

- **Sample size determination**
  - \[ n \propto \left( \frac{\text{power} \, s^2}{ES \, \alpha} \right) \]

- **Need to know**
  - Desired power (typically 0.8)
  - 80% probability of detecting an effect
  - Background variability (σ)
  - Estimated by s from pilot study or literature
  - Effect size (ES)
  - Magnitude of the effect that would be biologically significant
A priori power analysis – example 1

Effects of predation on mudflat crabs

Two treatments:
- Caged vs cage control

H₀: population mean crab numbers is the same for both caged and control treatments
- \( \mu_{\text{cage}} = \mu_{\text{control}} \)

Pilot study
- Number of crabs in 3 plots (no cages)
- Mean number of crabs in plots = 20
- Variance in crab numbers between plots = 19
  \( s = 4.36 \)

Aims:
- To detect a 50% increase in crab numbers due to caging (absence of fish predators)
  Increase in mean from 20 to 30 \( \rightarrow \) \( ES=10 \)
- To be 80% sure of detecting such a difference if it occurred
  \( \text{power} = 0.8 \)

How many replicate plots per treatment required?

What is required \( n \)?

\[ \texttt{> power.t.test(power=0.8,sd=4.36,delta=10)} \]

Altered ES
- Halved (ES=5)

Altered variability
- Doubled (s=6.16)
**A priori power analysis – example 1**

- Minimum Detectable Effect Size (MDES)
  - If ES can't be determined (no prior information)

\[
ES \propto \frac{\text{power}}{\sqrt{n}}
\]

**A priori power analysis – example 2**

- Effects of nitrogen on seedling growth
- Four treatments:
  - High, Medium, Low and Control (no) Nitrogen in potting soil
- \( H_0: \) population mean seedling growth rate is the same for all soil nitrogen treatments
  - \( \mu_{\text{High}} = \mu_{\text{Medium}} = \mu_{\text{Low}} = \mu_{\text{Control}} \)
- Pilot study
  - Growth rate of 5 seedlings normal soil (control soil)
  - Mean growth rate = 20 (units)
  - Variance in growth rate between seedlings = 9
  - \( \sigma = 3.00 \)

**A priori power analysis – example 2**

- Aims:
  - To detect a 50% increase growth rate due to soil nitrogen
    - Increase in mean from 10 to 15 → ES=5
  - To be 80% sure of detecting such a difference if it occurred
    - \( \text{power} = 0.8 \)
- How many replicate plots per treatment required?
  - What is required \( n \)?
  - Need to estimate between group variability

```r
> power.anova.test(group=4, power=0.8, between.var=9, within.var=9)
```
A priori power analysis – example 2

Need to consider planned comparisons
• If only want to determine whether the addition of nitrogen effects growth
  • $H_{High} = H_{Medium} = H_{Low} \neq H_{Control}$
  • $H_{Control}$ expected to be similar to pilot study ($=10$)
  • Others expected to be 50% greater
    • $H_{High} = H_{Medium} = H_{Low} = 10 \times 1.5 = 15$
  • Variation between treatment means (10, 15, 15, 15)
    • $s^2 = 6.25$

> power.anova.test(group=4, power=0.8, between.var=6.25, within.var=9)

\[ n = 6.3 \] (7)

A priori power analysis – example 2

Need to consider planned comparisons
• If want to determine whether High and Medium treatments are different to Low and Control
  • $H_{High} = H_{Medium} \neq H_{Low} = H_{Control}$
  • $H_{Control}$ and $H_{Low}$ expected to be similar to pilot study ($=10$)
  • $H_{High}$ and $H_{Medium}$ expected to be 50% greater
    • $H_{High} = H_{Medium} = 10 \times 1.5 = 15$
  • Variation between treatment means (10, 10, 15, 15)
    • $s^2 = 8.33$

> power.anova.test(group=4, power=0.8, between.var=8.33, within.var=9)

\[ n = 5.02 \] (6)

A priori power analysis – example 2

Need to consider planned comparisons
• If want to determine whether there is a linear trend in growth rate with increasing soil nitrogen
  • $H_{High} > H_{Medium} > H_{Low} > H_{Control}$
  • $H_{Control}$ is expected to be similar to pilot study ($=10$)
  • $H_{High}$ is expected to be 50% greater
    • $H_{High} = 10 \times 1.5 = 15$
  • $H_{Medium}$ and $H_{Low}$ are at even increments between
  • Variation between treatment means (10, 11.7, 13.3, 15)
    • $s^2 = 4.63$

> power.anova.test(group=4, power=0.8, between.var=4.63, within.var=9)

\[ n = 8.11 \] (9)
A priori power analysis – example 2

- ANOVA

![ANOVA graph](image)

With group variation = 9

- power

n

A priori power analysis – example 3

- Relationship between food consumption and tooth wear in possums
- X possums ranging in tooth class from 1 (low) to 6 (high)
- H0: population slope is equal to zero
  - \( \beta = 0 \)
  - Is the same as population correlation equals zero
  - \( r = 0 \)
- Previous study
  - 6 koalas of varying tooth wear
  - \( r^2 = 0.91 \) (r=0.95)

A priori power analysis – example 3

- Aims:
  - To detect a similar association between food consumption per unit of change in tooth wear
  - \( r = 0.95 \)
  - To be 80% sure of detecting such a relationship if it occurred
  - power = 0.8
- How many replicate plots per treatment required?
  - What is required \( n \)?

```
> pwr.r.test(power=0.8,r=0.95)
```

\( n = 5.18 \) (6)
Effect size

- How big?
  - What size of effect or trend is biologically important?
  - How big an effect or trend do we want to detect if it occurs?

- Where do we get suggested effect sizes from?
  - Biological knowledge/experience
  - Previous work/literature
  - Compliance requirements
    - E.g. water quality

Specification of effect size

- Depends on test
  - t-test – difference between means
  - Regression – $r^2$ or $r$
  - ANOVA – more complicated
    - Depends on hypothesis (e.g. four groups)
      - Difference between smallest and largest mean
        - $\text{Grp}_1 = \text{Grp}_2 = \text{Grp}_3 < \text{Grp}_4$ (one different)
        - $\text{Grp}_1 = \text{Grp}_2 = \text{Grp}_3 < \text{Grp}_4$ (two different)
        - $\text{Grp}_1 < \text{Grp}_2 < \text{Grp}_3 < \text{Grp}_4$ (trend)

Estimation of variance

- Where do we get suggested effect sizes from?
  - Biological knowledge/experience
  - Previous work/literature
    - Same systems
    - Similar systems
  - Pilot studies

- Estimated variance must be based on same sort of test
  - t-test – Paired vs independent two sample
  - ANOVA
  - Regression
Options for planning

- **Sample size determination (n)**
  - Desired power (0.8)
  - Effect size (EF)
  - Estimation of variance
  - Apply a “safety” factor to calculated $n$
  - Plot power vs $n$

- **Minimum Detectable Effect Size determination**
  - Desired power (0.8)
  - Estimate of variance
  - Possible sample size (or range)
  - Plot ES vs $n$

---

A posteriori power analysis

- **If statistically non-significant result**
  - Report power of test to detect relevant effect size
  
  $$\text{power}(1 - \beta) \propto \frac{\text{ES}}{\sigma} \sqrt{n}$$

- **From output**
  - Effect size (ES) - Magnitude of difference(s)
    - t-test – difference between means
    - ANOVA – $\sqrt{\text{MS}_{\text{between}} / \text{MS}_{\text{residual}}}$
    - Regression – $\sqrt{\text{MS}_{\text{regression}} / \text{MS}_{\text{residual}}}$ or $r$ (correlation coefficient)
  - Background variability ($\sigma$)
    - t-test – within group variation
    - ANOVA – $\text{MS}_{\text{residual}}$
    - Regression – $\text{MS}_{\text{residual}}$ or $r$ (standardized)
  - Sample size ($n$)

---

A posteriori power analysis – example 4

- **Plant growth in response to reduced herbivores**

- **Two treatments**
  - Reduced herbivore damage vs Normal herbivore damage (control)
  - $n=31$ plants in each treatment

- **Statistical outcome**
  - $t_{60} = 0.260$, $P = 0.48$ (not significant)
  - Within group variation = 0.5
  - Mean$_{\text{Reduced}} = 0.75$, Mean$_{\text{Control}} = 0.5$

```
> power.t.test(power=0.8, sd=0.5, n=31)
```
**A posteriori power analysis – example 4**

### Sample effect size
- \( \text{Mean}_{\text{reduced}} - \text{Mean}_{\text{Control}} = 0.75 - 0.5 \)
- \( \text{ES} = 0.25 \) (50% increase)

- Power = 0.5 (50%)
- 50% probability of detecting

- Minimum Detectable Effect size (at power = 0.8)
  - \( \text{ES} = 0.36 \) (72% increase)