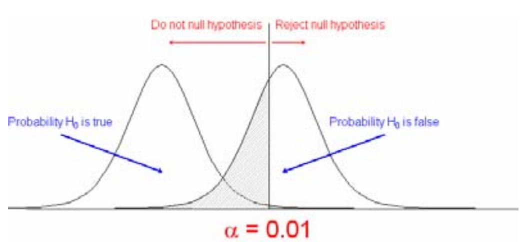
BIO4200 – stats fest 2007

Power analysis

murray.logan@sci.monash.edu.au

Statistical decisions

- **Why 0.05?**



Statistical decisions

	Statistical conclusion	
	Reject H_0 (conclude is effect)	**Retain H_0 (conclude no effect)**
Really is an effect	Correct decision – effect detected ✔	**Type II error** – effect not detected ✘
Really is no effect	**Type I error** – effected detected, none exists ✘	Correct decision – no effect detected, none exists ✔

Statistical decisions

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

	Statistical conclusion	
	Reject H_0 (conclude is effect)	**Retain H_0 (conclude no effect)**
Really is an effect	Correct decision – effect detected	**Type II error** – effect not detected
Really is no effect	**Type I error** – effected detected, none exists	Correct decision – no effect detected, none exists

Statistical decisions

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

	Statistical conclusion	
	Reject H_0 (conclude is effect)	**Retain H_0 (conclude no effect)**
Really is an effect	Correct decision – effect detected	**Type II error** – effect not detected
Really is no effect	**Type I error** – effected detected, none exists	Correct decision – no effect detected, none exists

Power of a test

- **Probability of detecting an effect if it exists**
- **Probability of correctly rejecting a false H_0**
- **Power = 1 – β (probability of making a Type II error)**
- **Usually aim for power ≈ 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

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

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 H_0
  - As α decreases, β increases, power decreases
  - Usually set at 0.05

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

Exact formula depends on the statistical test (i.e. it is different for *t*-test, ANOVA, chi-square, etc)

A priori power analysis

- **Sample size determination**

$$n \propto \left(\frac{\text{power} \ s}{\text{ES} \ \alpha}\right)^2$$

- **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

10

- **Effects of predation on mudflat crabs**
- **Two treatments:**
  - Caged vs cage control
- **H_0: population mean grab numbers is the same for both caged and control treatments**
  - $\mu_{cage} = \mu_{control}$
- **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

A priori power analysis – example 1

11

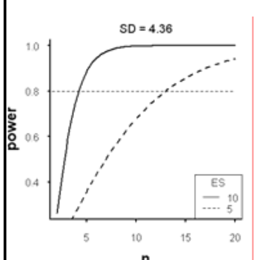
- **Aims:**
  - To detect a 50% increase in crab numbers due to caging (absence of fish predators)
    - Increase in mean from 20 to 30 → **ES=10**
  - **To be 80% sure of detecting such a difference if it occurred**
    - **power = 0.8**
- How many replicate plots per treatment required?
  - **What is required *n*?**

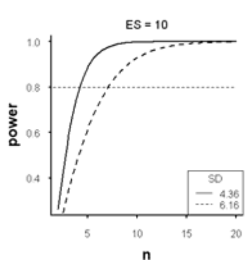
```
> power.t.test(power=0.8,sd=4.36,delta=10)
```

A priori power analysis – example 1

12

- **Altered ES**
  - Halved (ES=5)
- **Altered variability**
  - Doubled (s=6.16)



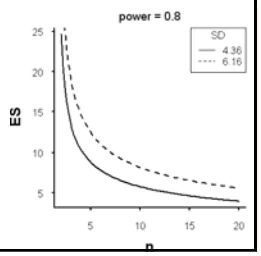


13

A priori power analysis – example 1

- **Minimum Detectable Effect Size (MDES)**
  - If ES cant be determined (no prior information)

$$ES \propto \frac{\text{power} \ \sigma}{\sqrt{n}}$$



14

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**
  - $\mu_{High} = \mu_{Medium} = \mu_{Low} = \mu_{Control}$
- **Pilot study**
  - Growth rate of 5 seedlings normal soil (control soil)
  - Mean growth rate = 20 (units)
  - Variance in growth rate between seedlings = 9
    - $s = 3.00$

15

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**
    - **power = 0.8**
- How many replicate plots per treatment required?
  - **What is required *n*?**
  - **Need to estimate between group variability**

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

A priori power analysis – example 2

16

- **Need to consider planned comparisons**
  - If only want to determine whether the addition of nitrogen effects growth
    - $\mu_{High} = \mu_{Medium} = \mu_{Low} \neq \mu_{Control}$
    - $\mu_{Control}$ expected be similar to pilot study (=10)
    - Others expected to be 50% greater
      - $\mu_{High} = \mu_{Medium} = \mu_{Low}$ = 10*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

17

- **Need to consider planned comparisons**
  - If want to determine whether High and Medium treatments are different to Low and Control
    - $\mu_{High} = \mu_{Medium} \neq \mu_{Low} = \mu_{Control}$
    - $\mu_{Control}$ and μ_{Low} expected be similar to pilot study (=10)
    - μ_{High} and μ_{Medium} expected to be 50% greater
      - $\mu_{High} = \mu_{Medium}$ = 10*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

18

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

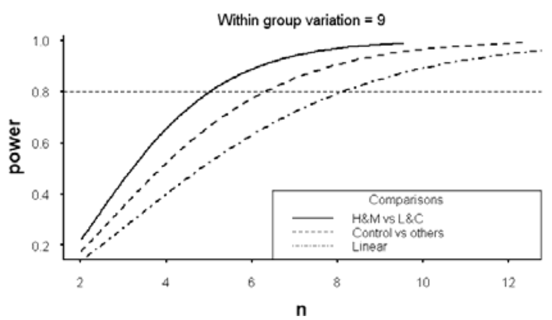
```
> power.anova.test(group=4,power=0.8,between.var=4.63,within.var=9)
```

- *n*=8.11 (9)

A priori power analysis – example 2

19

- **ANOVA**



A priori power analysis – example 3

20

- **Relationship between food consumption and tooth wear in possums**
- **X? possums ranging in tooth class from 1 (low) to 6 (high)**
- **H_0: 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

21

- **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

22

- **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

23

- **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
      - Grp1 = Grp2 = Grp3 < Grp4 (one different)
      - Grp1 = Grp2 = Grp3 < Grp4 (two different)
      - Grp1 < Grp2 < Grp3 < Grp4 (trend)

Estimation of variance

24

- **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

25

- **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

26

- **If statistically non-significant result**
  - Report power of test to detect relevant effect size

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

- **From output**
  - Effect size (ES) - Magnitude of difference(s)
    - *t*-test – difference between means
    - ANOVA – $\sqrt{MS_{Groups}}$ / $\sqrt{MS_{Residual}}$
    - Regression – $MS_{Regression}$ / $\sqrt{MS_{Residual}}$ or r (correlation coeficient)
  - Background variability (σ)
    - *t*-test – within group variation
    - ANOVA – $MS_{Residual}$
    - Regression – $MS_{Residual}$ or 1 (standardized)
  - Sample size (n)

A posteriori power analysis – example 4

27

- **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
  - $Mean_{Reduced}$ = 0.75, $Mean_{Control}$ = 0.5

```
> power.t.test(power=0.8,sd=0.5,n=31)
```

A posteriori power analysis – example 4

28

- **Sample effect size**
  - $\text{Mean}_{\text{Reduced}} - \text{Mean}_{\text{Control}} = 0.75 - 0.5$
  - ES = 0.25 (50% increase)
  - Power = 0.5 (50%)
    - 50% probability of detecting
  - Minimum Detectable Effect size (at power = 0.8)
    - ES=0.36 (72% increase)