Biological VS statistical significance

date

Readings for effect sizes

- See the PDFs in the ANGEL folder

Readings for Thursday

- We start our field guide to probability.
- For review, Dalgaard Chapters 3-4

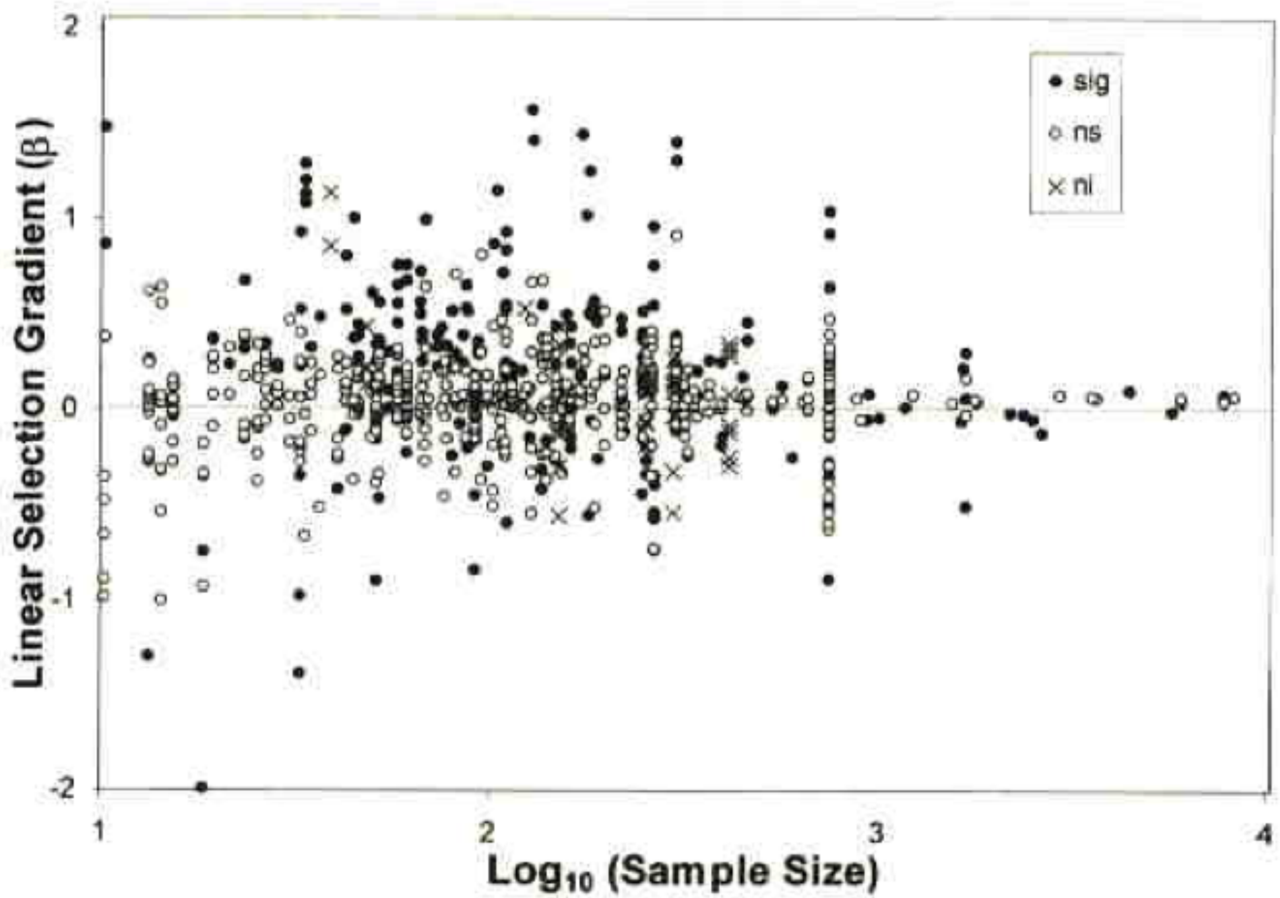
- Primary readings Bolker Chapter 4. Gelman and Gill Chapter 2 (pages 13-26).

- For more advanced readings in probability see syllabus

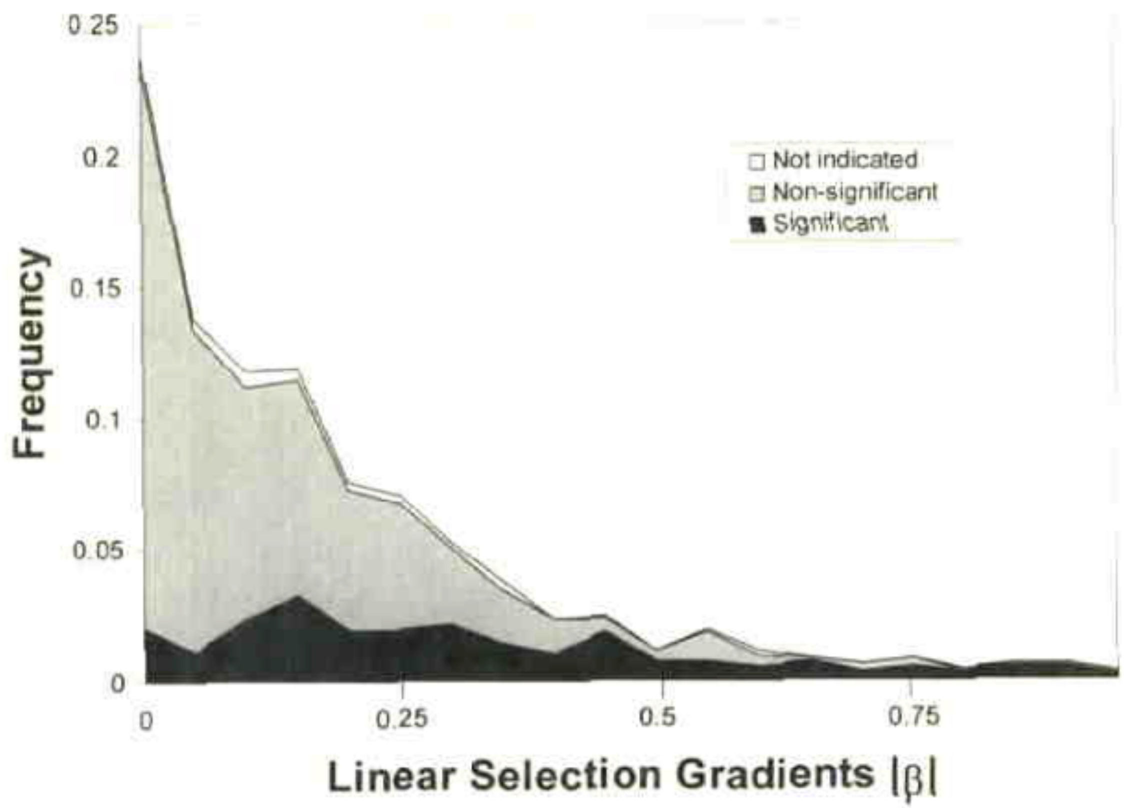
Goals for the day

Discuss the idea of effect sizes, and why (with CIs) they are central to any statistical and biological inference

Motivating example: The strength of natural selection in the wild



Is directional selection strong in nature?



Kingsolver et al 2001

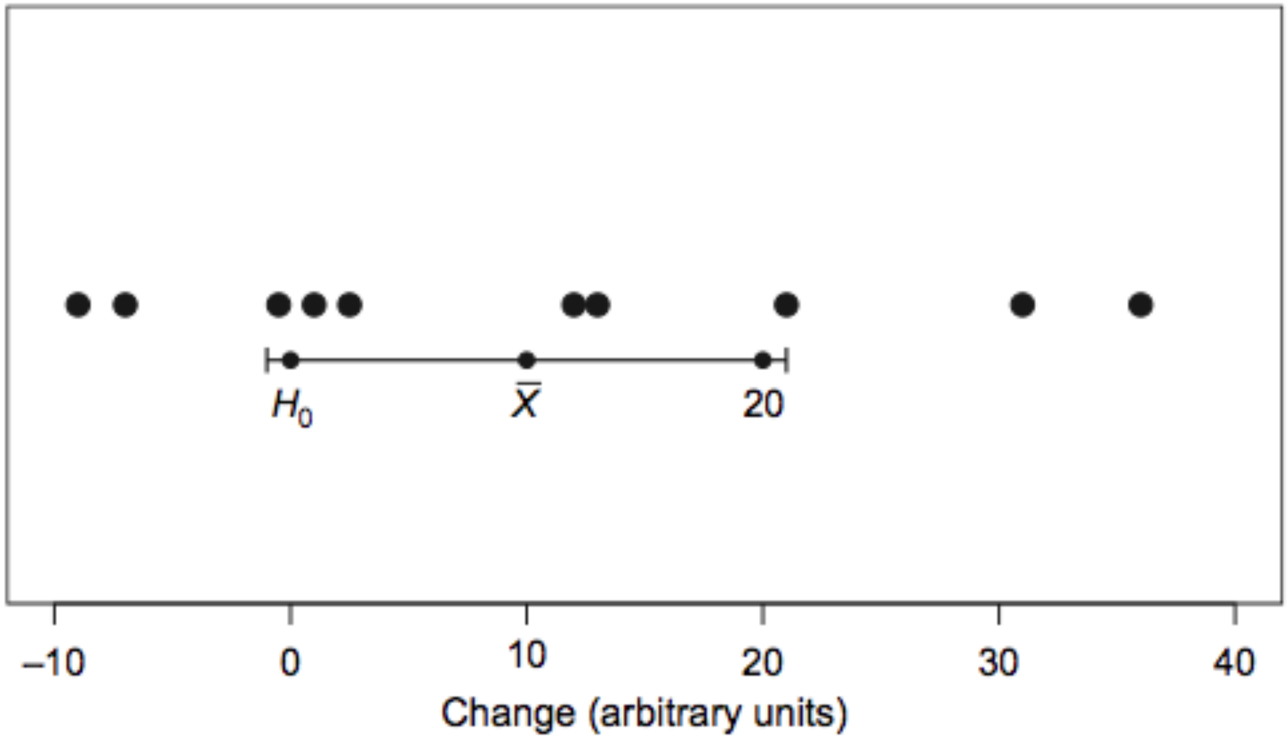
The Big Picture

- The coefficients from our models are not simply estimates to be examined along with p-values, but are probably the most important aspect of the model with respect to your ability to assess the importance of particular variables.

Salient points of the material

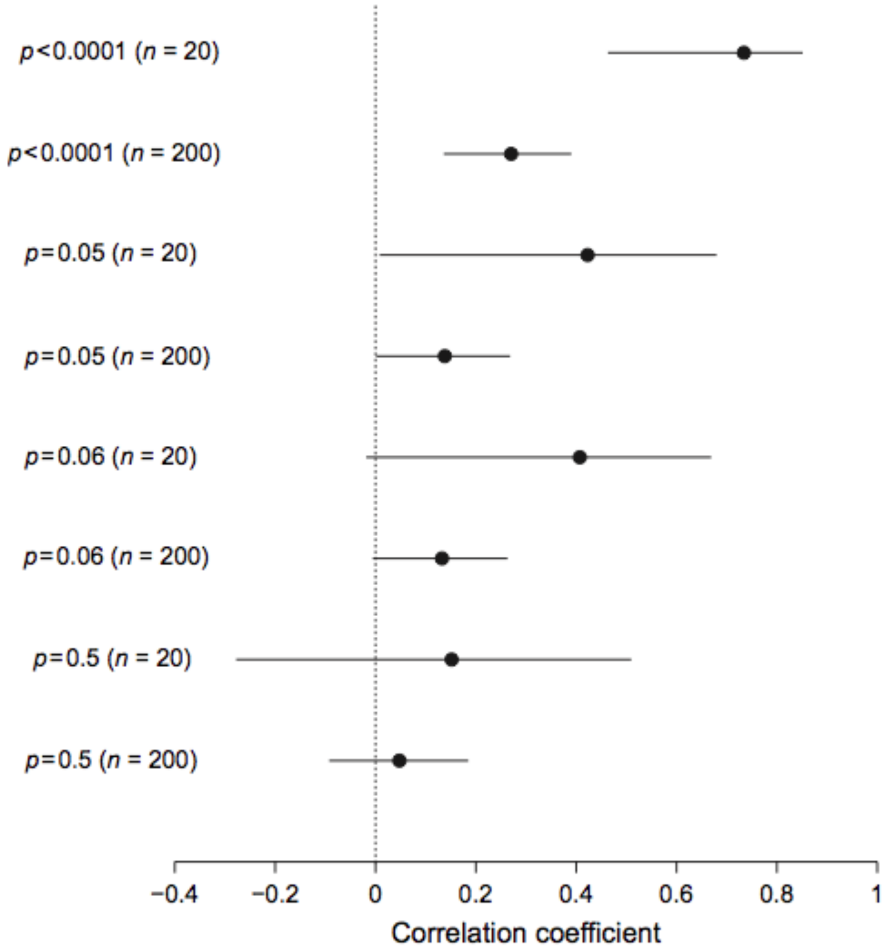
- There are several classes of effect sizes (unstandardized, scaled by pooled sd, scaled by mean, variance accounted for, odds ratios).
- Deciding which one to use may depend a fair bit on the question at hand, and what you plan to compare your results to.
- This can take a considerable amount of thought.

The Counter-null



Nakagawa & Cuthill 2007

Practical VS statistical significance



It may be significant, but is it important?

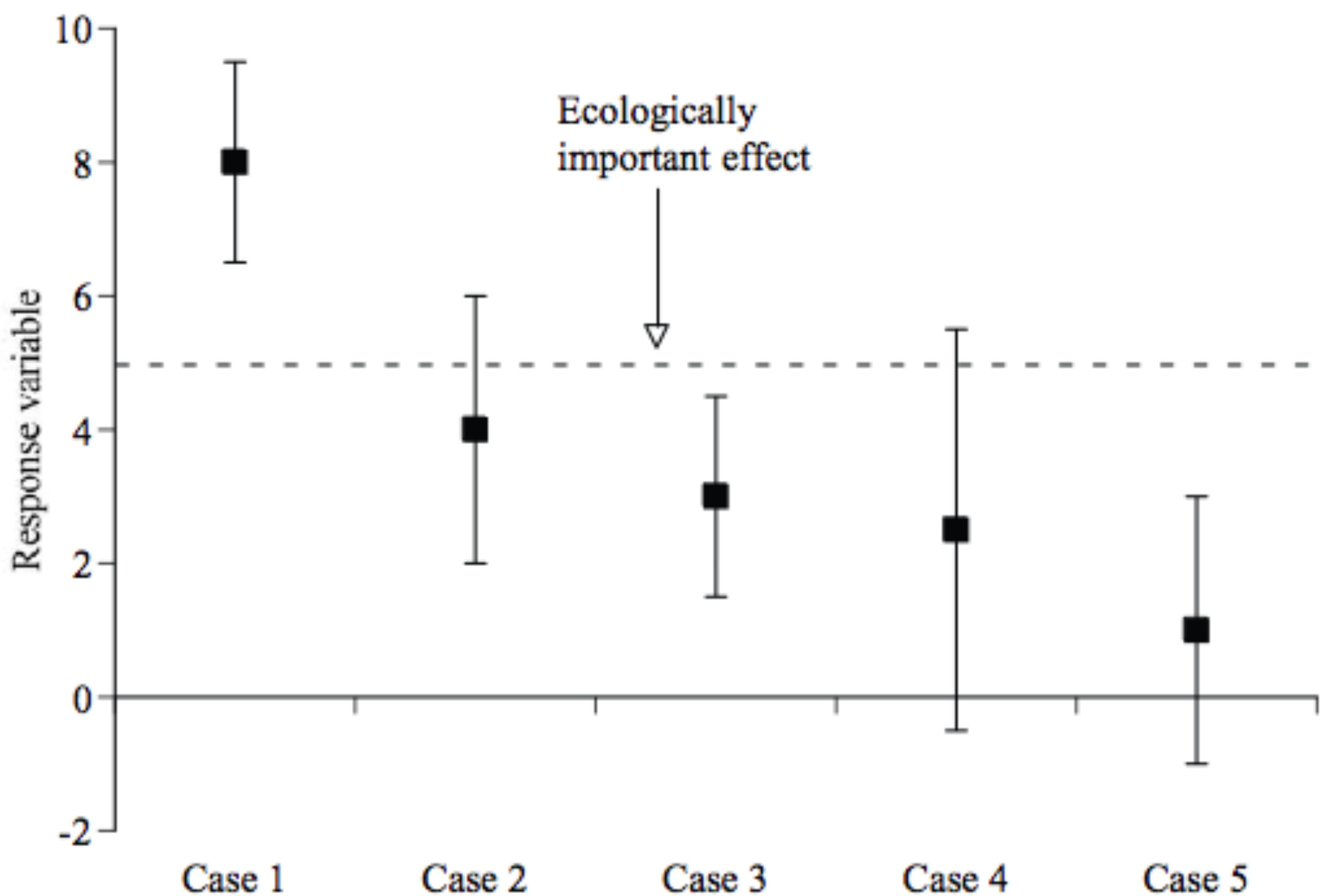


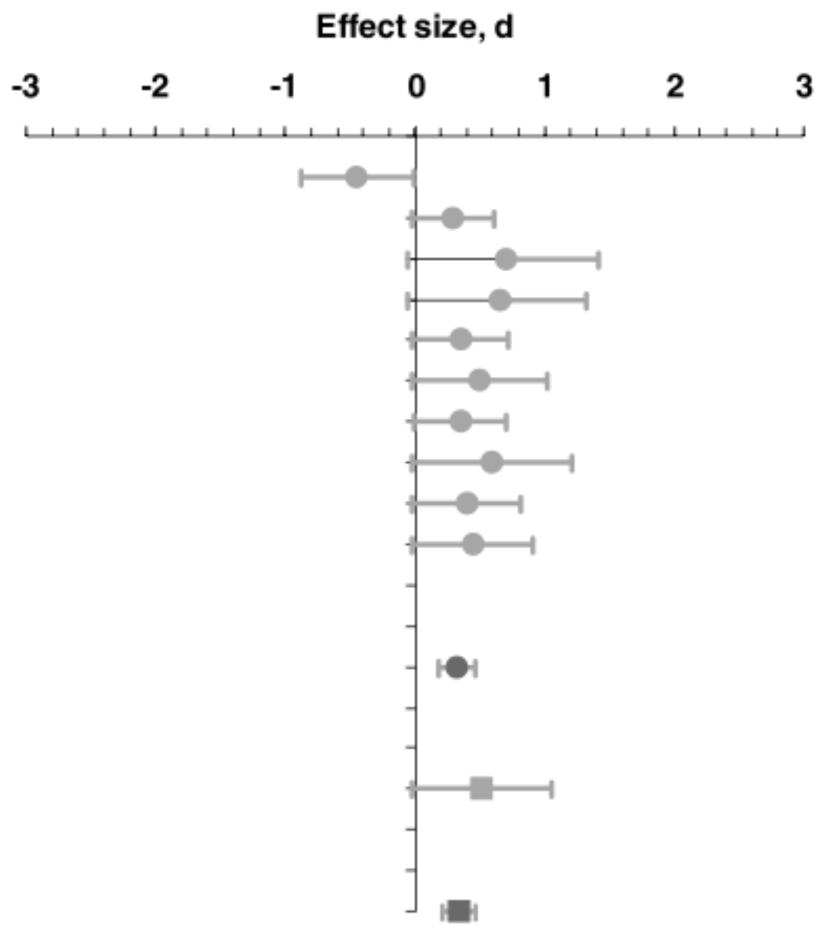
Table 1. Equations for calculating d statistics

Case	Equation		Description	References
Comparing two independent or dependent groups (i.e. both paired and unpaired t-test cases)	$d = \frac{m_2 - m_1}{s_{\text{pooled}}}$ $s_{\text{pooled}} = \sqrt{\frac{(n_2 - 1)s_2^2 + (n_1 - 1)s_1^2}{n_1 + n_2 - 2}}$	(1) (2)	m_1 and m_2 are means of two groups or treatments, s_{pooled} is pooled standard deviation, n is sample size (in the case of dependent design, the number of data points), s^2 is variance.	Cohen (1988); Hedges (1981)
Comparing two independent groups (i.e. unpaired t-test case)	$d = t_{\text{unpaired}}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	(3)	Alternatively, t values can be used to calculate d values; t_{unpaired} is the t value from the unpaired t-test (compare with Equation 10 in the text)	Rosenthal (1994)
Comparing two dependent groups (i.e. paired, or repeated-measure t-test case)	$d = t_{\text{paired}}\sqrt{\frac{2(1 - r_{12})}{n}}$	(4)	t_{paired} is the t score from the paired t-test, r_{12} is correlation coefficient between two groups, and note that $n = n_1 = n_2$ not $n = n_1 + n_2$	Dunlap *et al.* (1996)

Free software by David B. Wilson to calculate these effect statistics is downloadable (see Table 4). Strictly speaking, Equations 1 to 4 are for Hedges's g but in the literature these formulae are often referred to as d or Cohen's d while Equation 10 is Cohen's d (see Kline, 2004, p.102 for more details; see also Rosenthal, 1994; Cortina & Nouri, 2000).

Nakagawa & Cuthill 2007

Effect size in context



Thompson 2007

Can you think of other ways to scale the measures of effect sizes?

What might we do if we have many levels to a given categorical predictor?

R^2: The co-efficient of determination

- R^2 is probably the most commonly used quantity for model fit.

- Often described as the proportion of variation *explained* by the model.

- I prefer: proportion of variation *accounted* for by the model.

- I really prefer thinking about 1 -R^2: proportion of variation for unaccounted for (how much are you missing the mark).

R^2: The co-efficient of determination

- SS.total = SS.model + SS.residual
- (un-adjusted) R^2 = 1 - (SS.residual/SS.Total)

  = SS.model/SS.Total
- $0 \leq R^2 \leq 1$
- However, when you add more parameters to a model, at worse they do not increase SS.model (they will never decrease it).
- Effectively unadjusted R^2 will always increase with more parameters added to the model.
- It does not penalize more complex models (violating our parsimony principal).

Adjusted R^2

- Adjust for parsimony principle

- Adjusted $R^2 = 1 - (n-1)/(n-p)(1-R^2)$
  $= 1 -$ residual MS/total MS

- Adj. R^2 can decrease with increasing numbers of parameters (p).

- Information theoretic approaches are still far better ways of comparing different models.

Generalized R^2

$$\text{Generalized R}^2 = \frac{1-\left(\dfrac{L(\text{null})}{L(\hat{\theta})}\right)^{2/n}}{1-L(\text{null})^{2/n}}$$

Where L(null) is the log Likelihood for the null model, L(theta hat) is the log likelihood for the MLE of the model, and n is sample size.

This does not adjust for parameters.

Is R^2 useful?

- Yes. It is very useful in making a statement about **overall** model fit (% variation accounted for).
- But, it is ***not useful*** in the comparison between models.

Model vs predictor specific R^2

- While in a glm with multiple predictors, the coefficients are adjusted for the presence of one another, this is not the case for R^2.
- Most statistical software provides the R^2 for the full model.
- So how do we assess variance accounted for at a predictor level?

Model vs predictor specific R^2

- Can we just fit individual models for each predictor to calculate the R^2?

Coefficient of Partial Determination
Partial R^2

•We can instead adjust the R2 in a manner analogous to adjusting coefficients for other predictor variables.

•These are called partial R^2 (named to provide similar meaning to partial regression coefficients.).

•These allow you to adjust the R^2 for a given predictor, given all of the other predictors in the model.

•You can do this in R using the partial.R2 function in the asbio library.

```
partial.R2(model.without.predictor, model.with.predictor)
```

Coefficient of Partial Determination
Partial R^2

The partial R^2 for X_1, given that X_2 is already in the model is calculated as

$$R^2_{Y1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSM(X_1 \mid X_2)}{SSE(X_2)}$$

This extends more generally.

$$R^2_{Y4|123} = \frac{SSM(X_4 \mid X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)}$$

SSE = Error (residual) Sum of Squares
SSM = Model (regression) sum of squares (fitted)

Salient points of the material

- There are several classes of effect sizes (unstandardized, scaled by pooled sd, scaled by mean, variance accounted for, odds ratios).
- Deciding which one to use may depend a fair bit on the question at hand, and what you plan to compare your results to.
- This can take a considerable amount of thought.