Statistical Power Analysis to Define Reference Criteria for Montana

Brett D. Marshall [1,2] and David C. Richards [1,2]

[1] EcoAnalysts, Inc.

[2]Montana State University, Department of Ecology

Power Analysis and Multimetric Bioassessment

- Metric selection based on environmental sensitivity

  Δ/δ

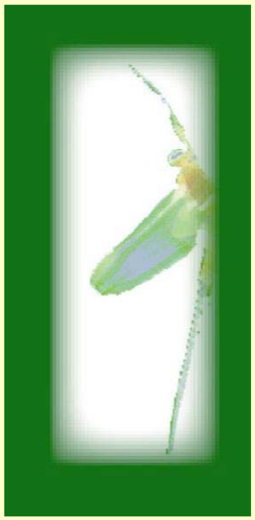
- Metric scaling and scoring.

Power Analysis and Multimetric Bioassessment

- BUT WHY?

POWER is a critical element of ecological investigation

$\neq$



Statistical Hypothesis Testing

Research question phrased as a pair of complementary hypotheses:

Null (H_o) and alternative (H_a) hypotheses

Ex. H_o: Impaired value = $\mu_{reference}$

H_a: Impaired value $\neq$ $\mu_{reference}$

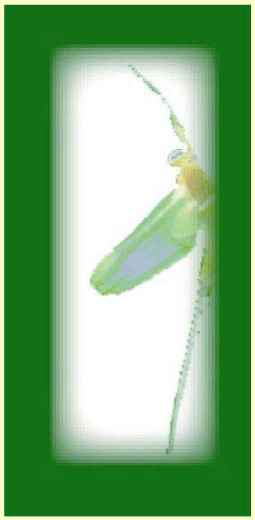
Type I error α

Example: *t*-test

$t > t_{crit}$: reject Ho

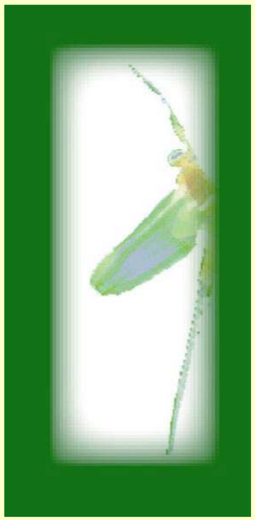
t_{crit} is based on previously determined P-value, quite often $\alpha = 0.05$

Interpreted as: evidence provided by the data suggests that attaining a t-value as extreme as the one observed is unlikely to occur by chance



But.....

There is always a chance (probability) that H_o is true

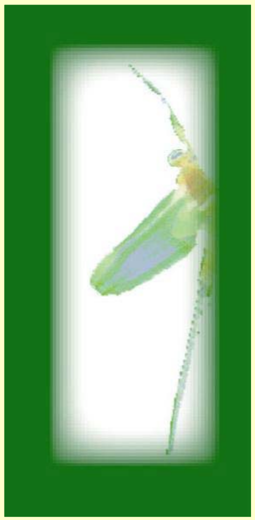


Type I and II errors

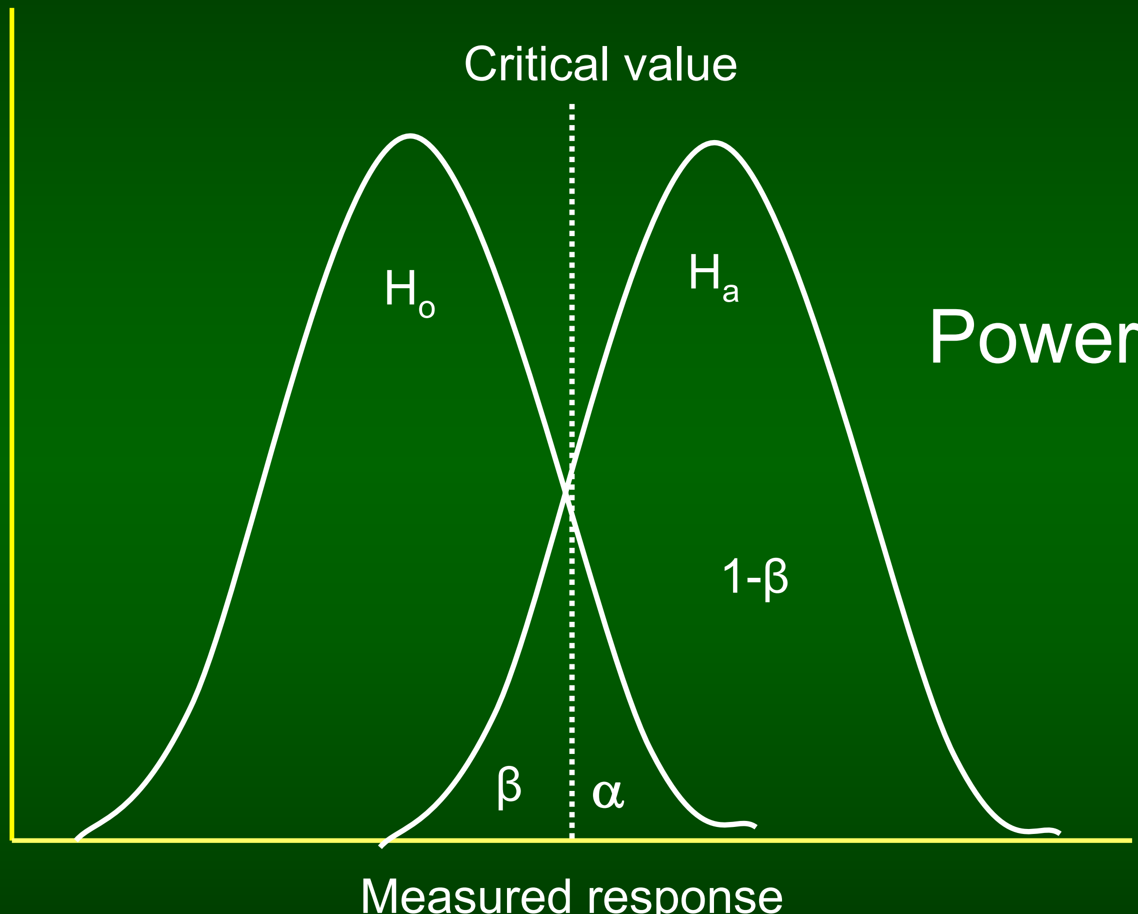
REALITY

INFERENCE (decision)	H_0: true	H_A: true
H_0:	———	Type-II error ← β
H_A:	Type-I error ↑ α	———

Example Type II error: observed metric was truly different than reference metric but researcher inferred no difference: Type II errors probably more important in our line of work!



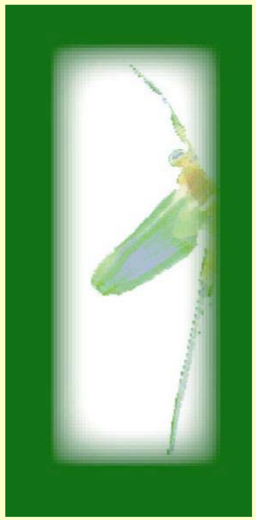
Type II error



Power = 1-β

Power is determined by 4 factors

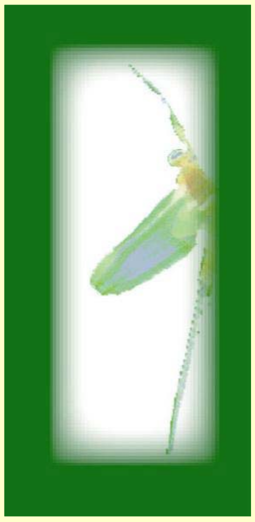
1. Power increases as:
    1. sample size, N increases
    2. α –level increases
    3. effect size, δ increases
2. Power decreases as:
    1. variance, σ increases



Power Analysis

For this design, statistical power is a function of

- μ: population mean
- σ: population variance
- n: number of replicates (reference sites)
- α: probability of type-1 statistical error
- β: probability of type-2 statistical error
- δ: effect size



Power Analysis

For this design, statistical power is a function of

- μ: population mean
- σ: population variance
- n: number of replicates
- α: probability of type-1 statistical error
- β: probability of type-2 statistical error
- δ: effect size

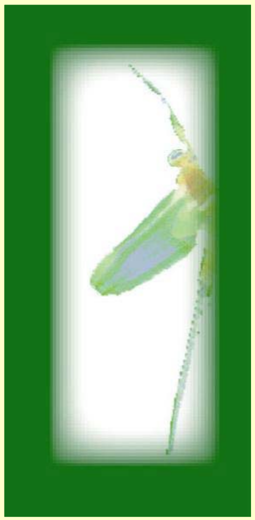
Effect size, δ

δ Least familiar to most researchers

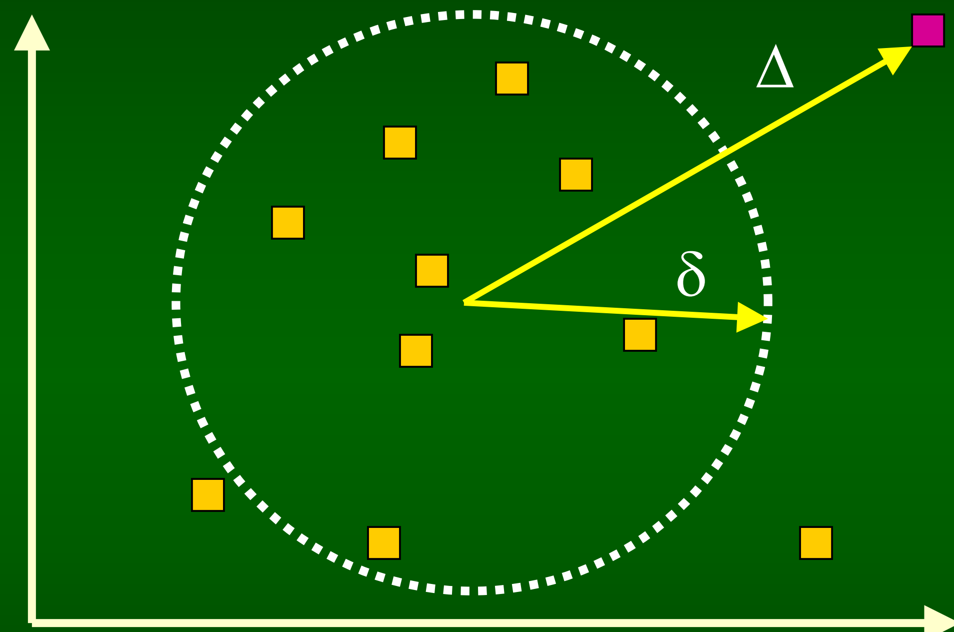
Broadly defined as:
difference between H_o and specific H_a

Choosing meaningful δ or suite of δ 's
is usually biggest challenge

<u>For example</u>
Is 1, 10, 25% change in EPT richness important?



Power Analysis and Metric Sensitivity: Δ/δ

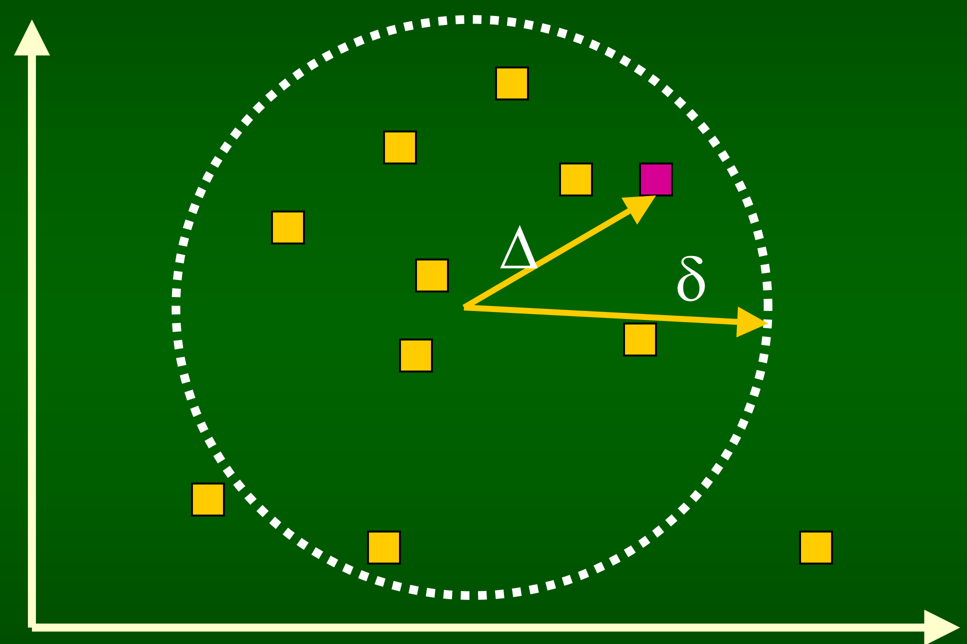


Δ/δ > 1

Δ = amount metric changes in samples from **impaired site**

δ = amount metric changes in samples from **reference** condition

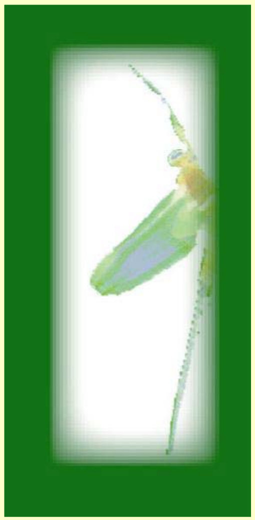
Power Analysis and Metric Sensitivity: Δ/δ



$\Delta/\delta < 1$

Δ = amount metric changes in samples from **impaired site**

δ = amount metric changes in samples from **reference** condition



Power Analysis and Metric Sensitivity: Δ/δ

<u>Delta/ delta can be used to:</u>

(1) Evaluate / screen metrics for use in bioassessment indices to ensure sensitive metrics are used.

(2) Screen metric response for different types of Impairments.

(3) Evaluate the cost-benefits of different field & lab procedures

Since we are considering change in Δ/δ, the tool for this is

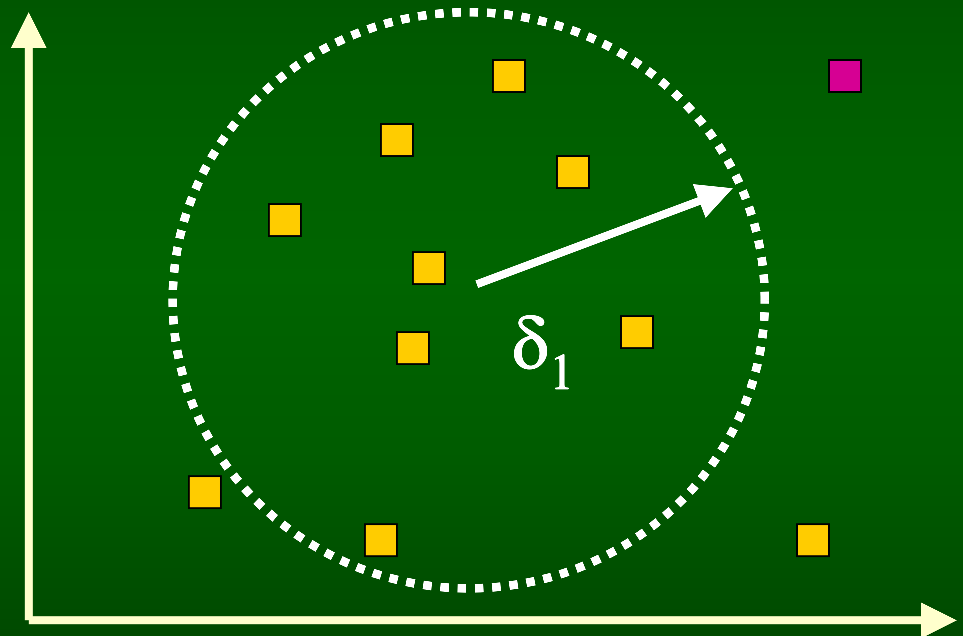
$$\Delta(\Delta/\delta)$$

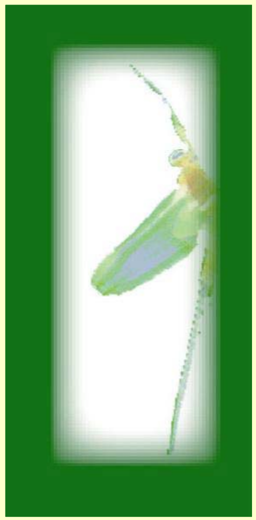
Another Use of Power analysis

We also used power analysis to Scale and Score metrics for a Multi-metric index that is about to be implemented for the State of Montana.

Defining Deviation Thresholds δ_1

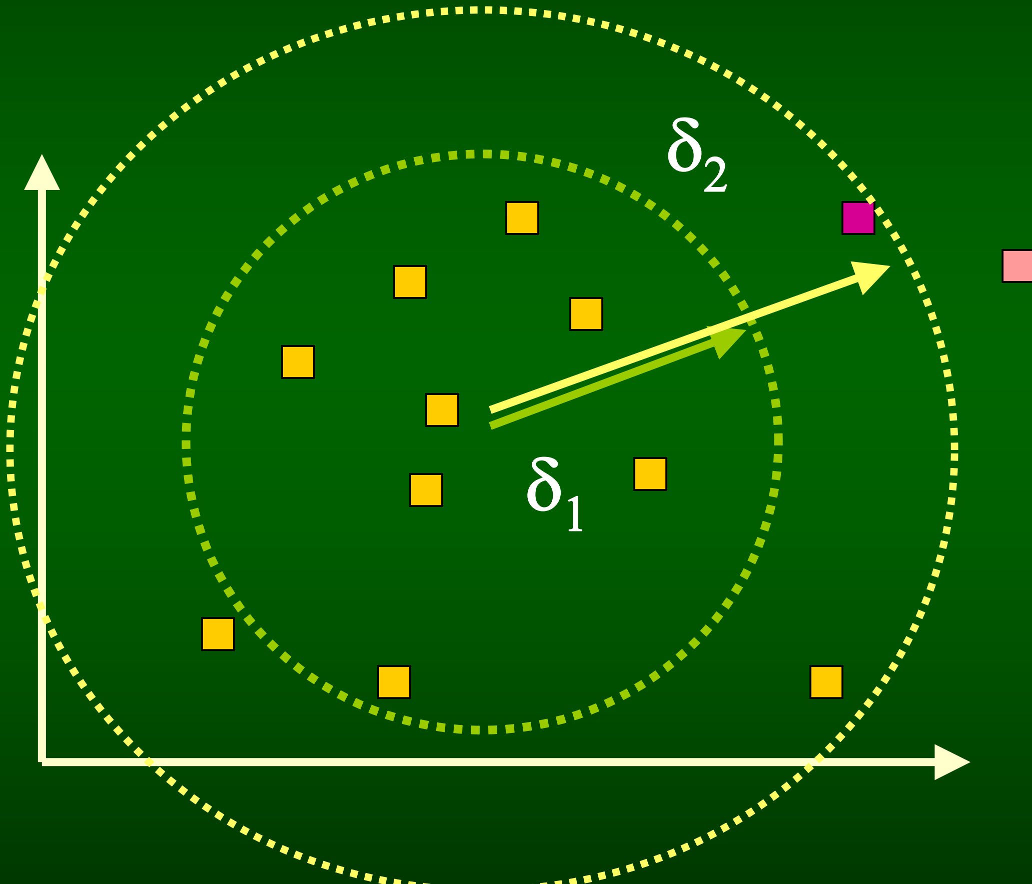
$\alpha = \beta = 0.15$



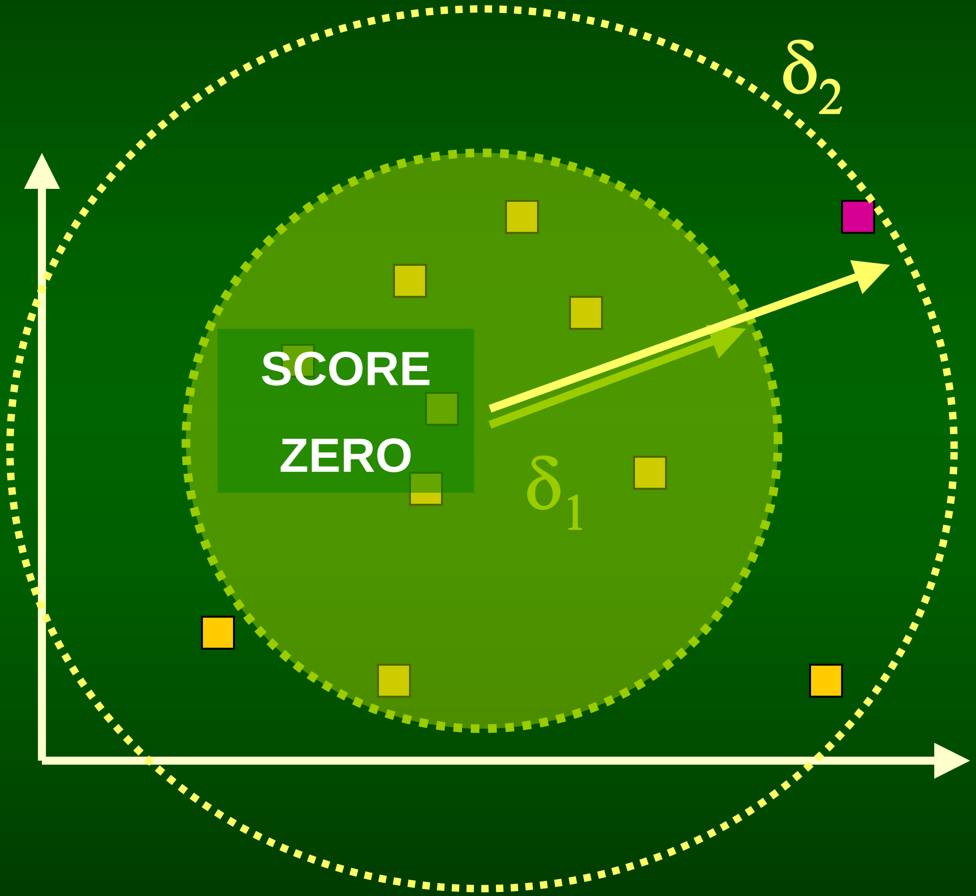


Defining Deviation Thresholds δ_2

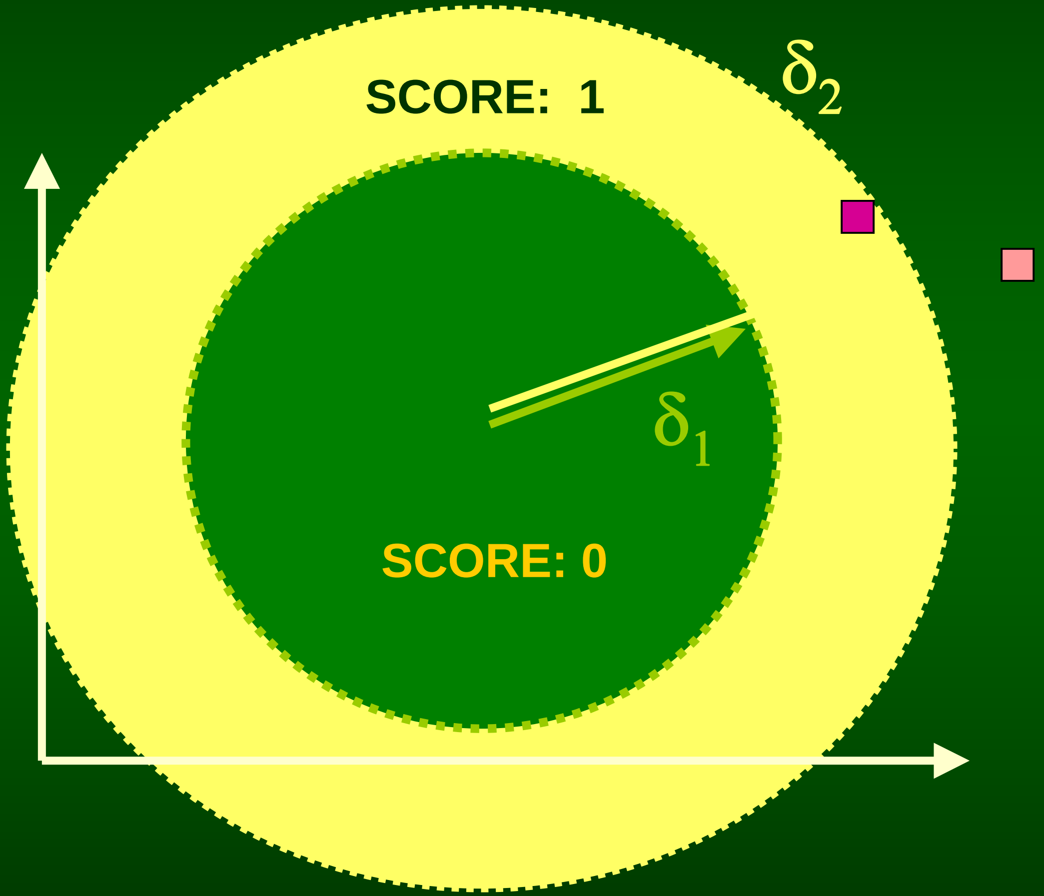
$\alpha = \beta = 0.01$



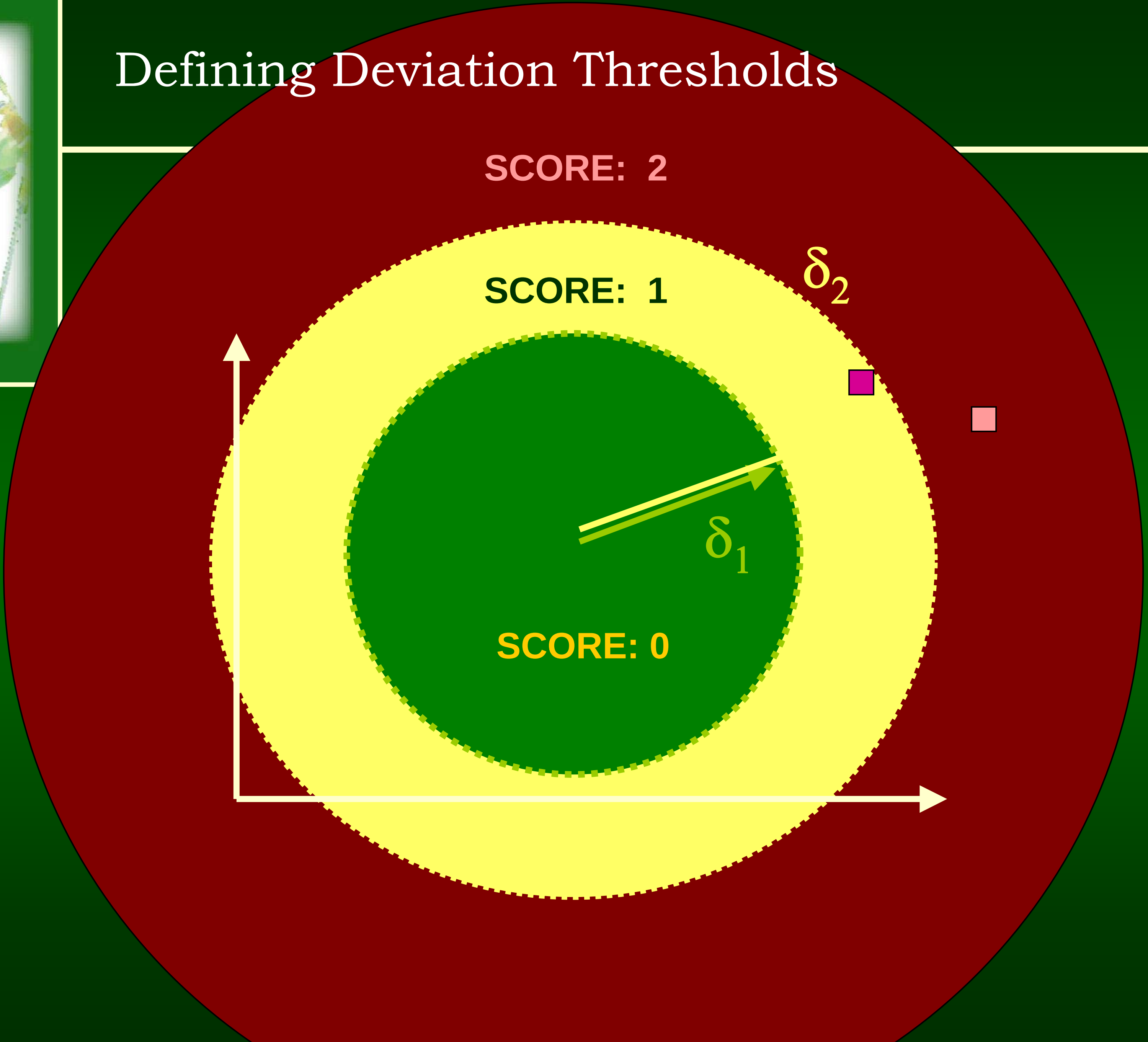
Defining Deviation Thresholds



Defining Deviation Thresholds

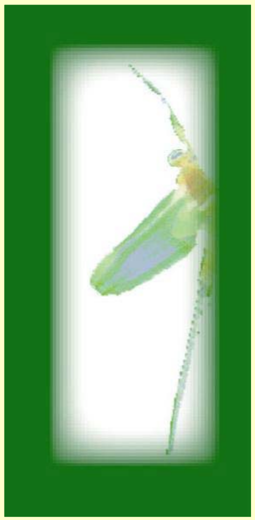


Defining Deviation Thresholds



Scoring the multimetric index

Since metrics are scored, “zero,” “1,” or “2,” depending on the probability of deviation from reference conditions, the sum of multimetric index score increases as the probability of a site deviating from the reference conditions increases

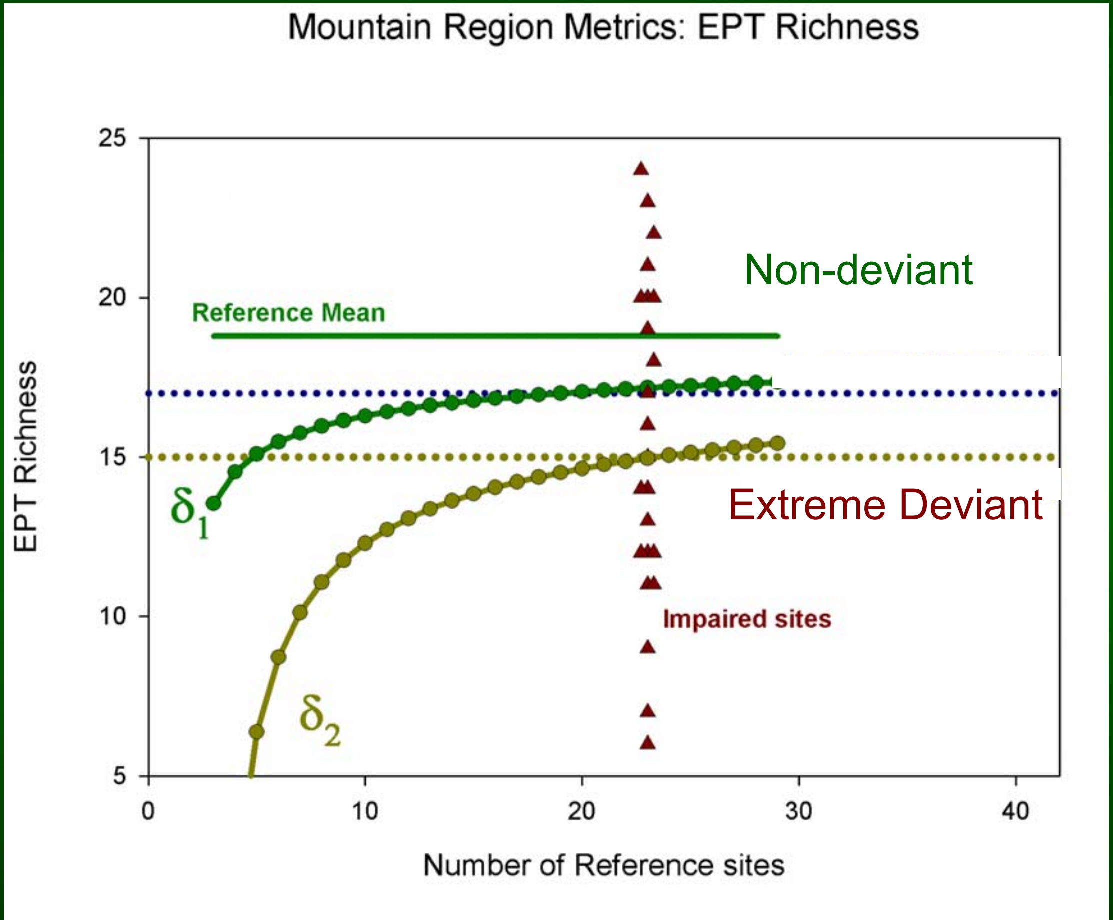


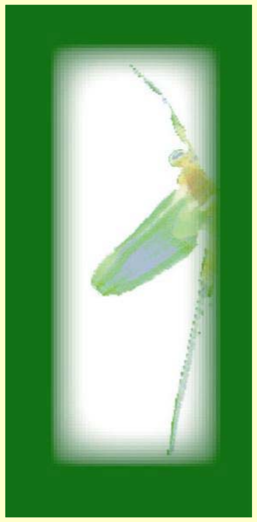
An example from the mountains of Montana

Ex. 6 metrics

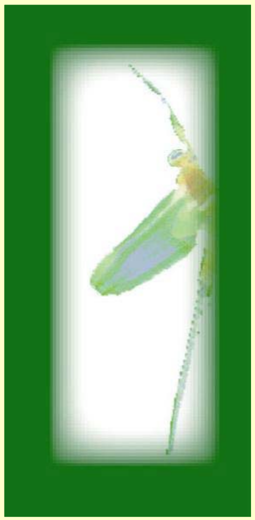
Total Score	
0	Non-Deviant
1	
2	
3	
4	Moderate Deviant
5	
6	
7	
8	Extreme Deviant
9	
10	
11	
12	Maximum Deviant

An example from the mountains of Montana





Moderate Deviant



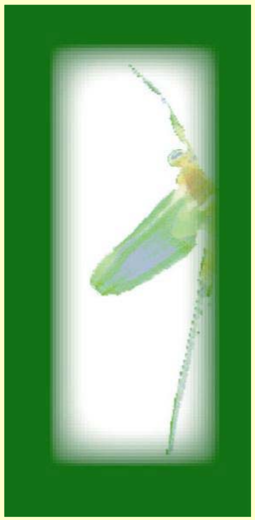
An example from the mountains of Montana

METRIC	δ_1	δ_2
Taxa Richness	< 25	< 22
Simpson’s Index	> 0.172	> 0.206
EPT-Richness	< 17	< 15
B:E	> 0.213	> 0.283
% Chironomidae	> 0.128	> 0.18
% Diptera	> 0.168	> 0.232
% Sprawlers & Burrowers	> 0.213	> 0.217
% Collectors	>0.422	> 0.555
% Gatherers	> 0.340	> 0.451

An example from the mountains of Montana

Mountain Reference Streams comparing 8 suites of metrics

Total Score	I	II	III	IV	V	VI	VII	VIII
Non-Deviant	83.3 %	83.3 %	78.9 %	73.7 %	73.7 %	84.2 %	78.9 %	84.2 %
Moderate Deviant	16.7%	16.7%	21.1%	26.3 %	26.3 %	15.8 %	21.1 %	15.8 %
Extreme Deviant	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Max. Deviant	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%



An example from the mountains of Montana

Mountain Reference Streams

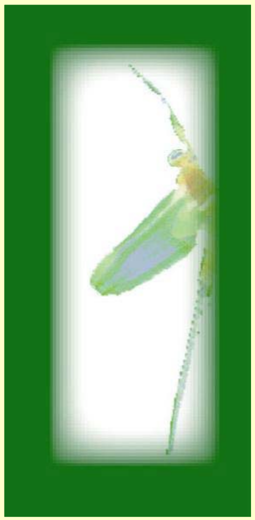
Total Score	I	II	III	IV	V	VI	VII	VIII
Non-Deviant	83.3 %	83.3 %	78.9 %	73.7 %	73.7 %	84.2 %	78.9 %	84.2 %
Moderate Deviant	16.7%	16.7%	21.1%	26.3 %	26.3 %	15.8 %	21.1 %	15.8 %
Extreme Deviant	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Max. Deviant	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

$$\alpha = \beta = 0.15$$

An example from the mountains of Montana

Mountain “impaired” Streams

Total Score	I	II	III	IV	V	VI	VII	VIII
Non-Deviant	30.3%	27.3%	30.3%	**18.2%**	**18.2%**	24.2%	33.3%	27.3%
Moderate Deviant	36.4%	39.4%	39.4%	**57.6%**	**57.6%**	51.5%	36.4%	48.5%
Extreme Deviant	30.3%	30.3%	27.3%	**21.2%**	**21.2%**	21.2%	27.3%	21.2%
Max. Deviant	3.0%	3.0%	3.0%	**3.0%**	**3.0%**	3.0%	3.0%	3.0%



An example from the mountains of Montana

When we examined the differential classification*,
One metric suite stood out among the others…

Total Score	I	II	III	IV	V	VI	VII	VIII
Non-Deviant	53.0%	56.1%	48.6%	55.5%	55.5%	60.0%	45.6%	56.9%
Moderate Deviant	-19.7%	-22.7%	-18.3%	-31.3%	-31.3%	-35.7%	-15.3%	-32.7%
Extreme Deviant	-30.3%	-30.3%	-27.3%	-21.2%	-21.2%	-21.2%	-27.3%	-21.2%
Max. Deviant	-3.0%	-3.0%	-3.0%	-3.0%	-3.0%	-3.0%	-3.0%	-3.0%

*reference correct – impaired correct for each category

An example from the mountains of Montana

When we examined the differential classification, One metric suite stood out among the others…

Total Score	I	II	III	IV	V	VI	VII	VIII
Non-Deviant	53.0%	56.1%	48.6%	55.5%	55.5%	**60.0%**	45.6%	56.9%
Moderate Deviant	-19.7%	-22.7%	-18.3%	-31.3%	-31.3%	**-35.7%**	-15.3%	-32.7%
Extreme Deviant	-30.3%	-30.3%	-27.3%	-21.2%	-21.2%	**-21.2%**	-27.3%	-21.2%
Max. Deviant	-3.0%	-3.0%	-3.0%	-3.0%	-3.0%	**-3.0%**	-3.0%	-3.0%

An example from the mountains of Montana

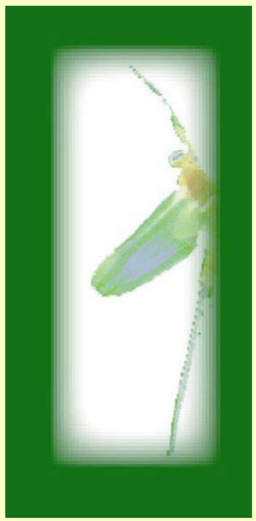
METRIC	δ_1	δ_2
Taxa Richness	< 25	< 22
Simpson's Index	> 0.172	> 0.206
EPT-Richness	< 17	< 15
B:E	> 0.213	> 0.283
% Chironomidae	> 0.128	> 0.18
% Diptera	> 0.168	> 0.232
% Sprawlers & Burrowers	> 0.213	> 0.217
% Collectors	>0.422	> 0.555
% Gatherers	> 0.340	> 0.451

Summary:

❖ Statistical Power can be used to score multimetric indices.

❖ We believe this method strengthens the defensibility of Montana's Rapid bioassessment protocols, because power is implicitly imbedded in the development of decision thresholds.

❖ Brett is looking for more data to test these ideas further.



Conclusion

However you develop your biocriteria, consider statistical power.

It is important to understand how likely your assessment protocol is to detect real impairments.

Many investigators use the incorrect statistical design for power…

If you would like help, please contact Brett.

Bmarshall@ecoanalysts.com


THANKS!