

Resampling Methods

Advanced Biostatistics

Dean Adams

Lecture 2

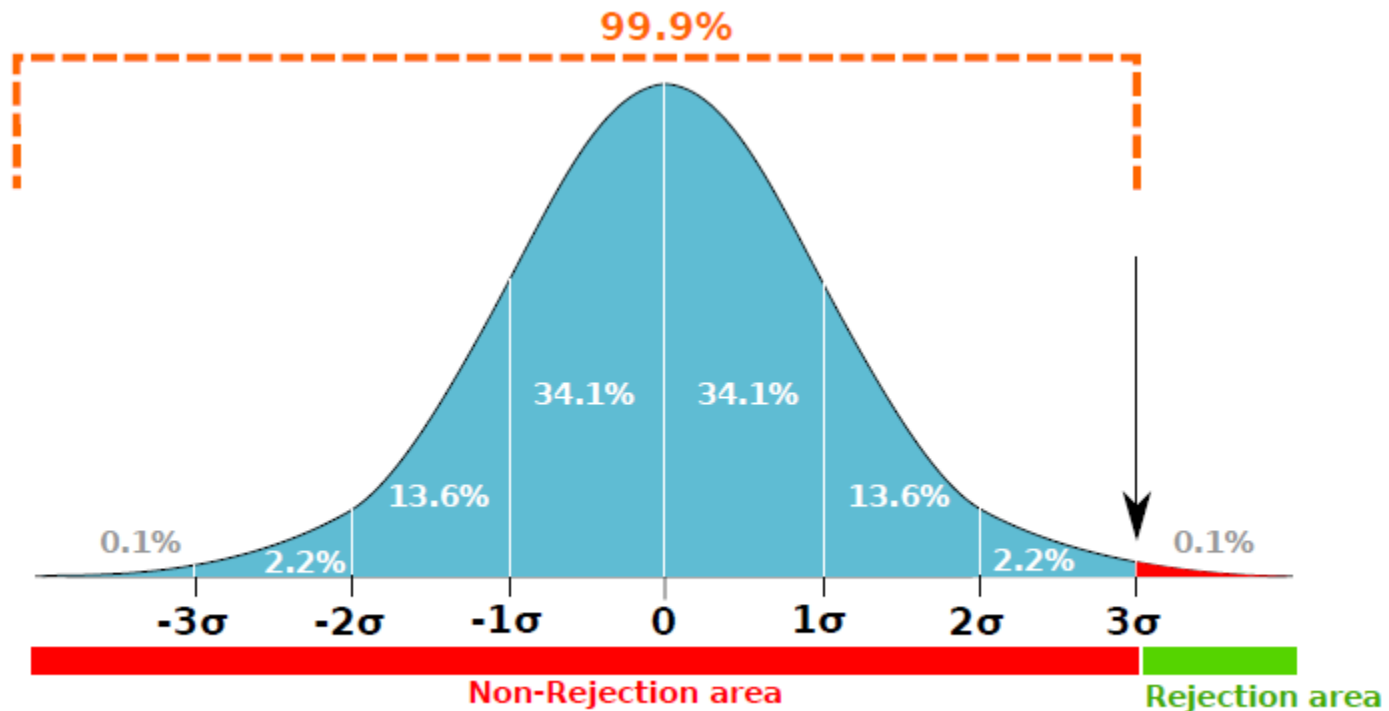
EEOB 590C

Inferential statistics: the process of making generalizations about populations, based on samples drawn from these populations.

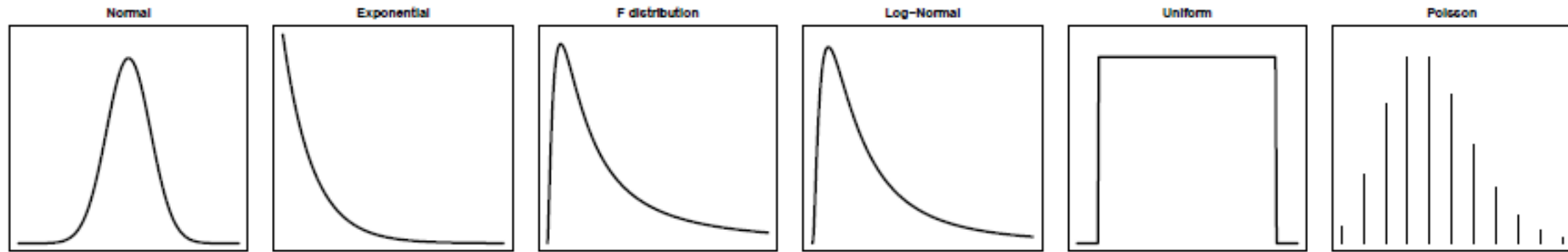
Formulate a null hypothesis (H_0)

H_0 is described by a statistic measure, which has a null distribution

Compare the observed statistic with the value expected under H_0

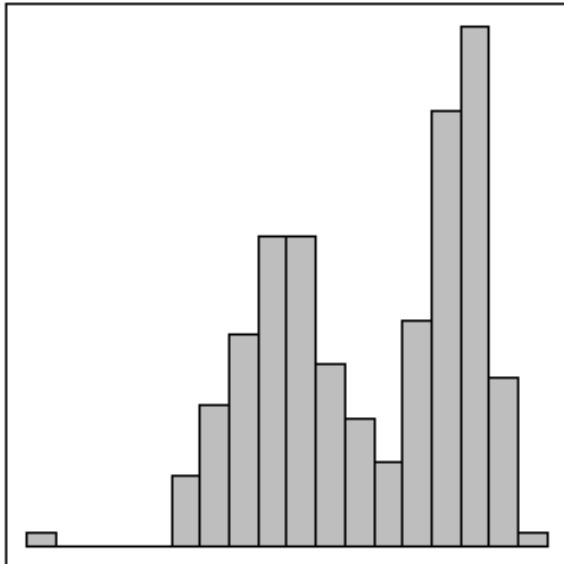


Frequentist statistics: Theoretical distributions obtained from theory



Resampling statistics: we generate expected distributions from the data

↖ Data driven distribution



Take many samples from original dataset in some manner

Evaluate significance of original data based on these samples

Methods are:

- Nonparametric (no theoretical distribution)

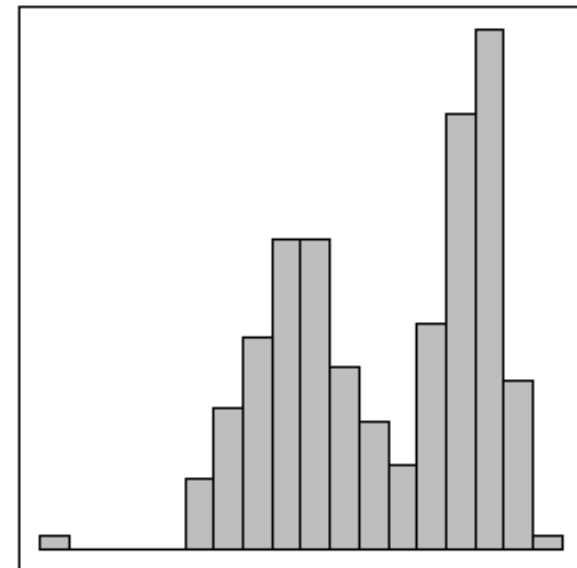
- Flexible (easy to assess complex and non-standard designs)

Useful for testing:

- Standard designs

- Non-standard designs

- High-dimensional data ($p > N$)



Names are interchangeable

Randomization/Permutation

Bootstrap

Jackknife

Monte Carlo simulations

Random sampling of original data

Resampling without replacement ($n' = n$)

Resampling with replacement ($n' = n$)

Leave-one-out resampling ($n' = n-1$)

Data simulated based on a model and parameters from original data

Full permutation of the data (without replacement) to create null distribution

First implementation: Fisher's exact test (1935)

Lady tasting tea experiment

"A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup."

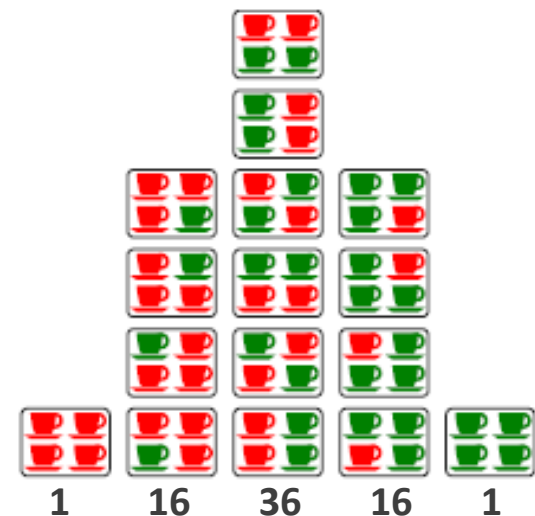
"Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order."

"Her task is to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received."

H_0 : the lady has no such ability

Total combinations: $\frac{8!}{4! \times 4!} = 70$

Probability of success: $\frac{1}{70} \times 100\% = 1.4\%$



Are males larger in body size than females?

```
> m <- c(-0.01, -0.41, -0.44, 0.58, 0.80, 0.1, -1.09, 0.13, 0.14,
0.34, -0.30, 0.22, 0.36, 1.04, 0.77)
> f <- c(0.87, 0.57, 1.76, 0.44, 0.87, 0.51, 1.12, 0.36, 1.91, 0.20,
0.82, -0.18, 1.23, -1.22, 0.01)
> mean(m)
```

0.62

```
> mean(f)
```

0.15

```
> mean(m) - mean(f)
```

Unstandardized test
measure! Dependent upon
units. Valid only in this test.

0.47

Total combinations: $\binom{30}{15} = 155\,117\,520$

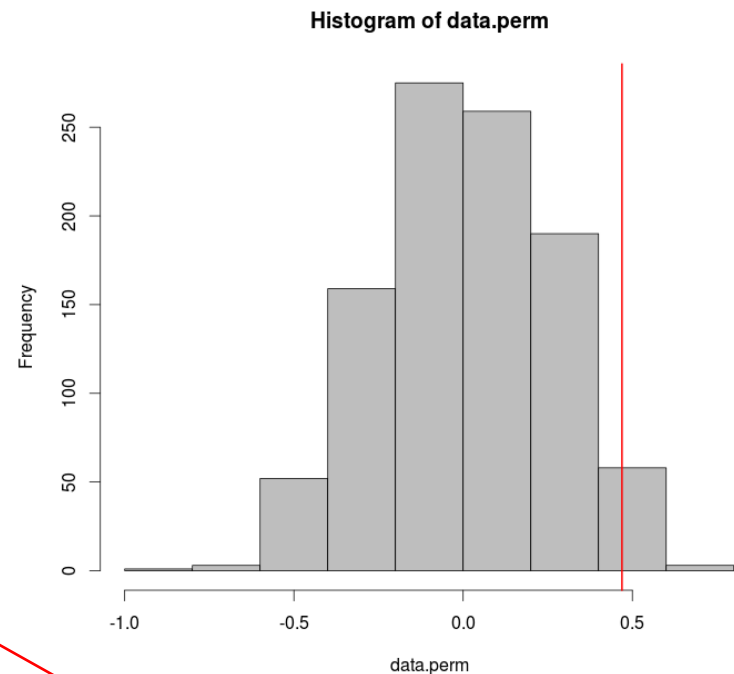
Take random sample i times

perm = i

```
> pval <- length(which(data.perm >= SDobs)) / (perm + 1)
```

```
> pval
```

0.027



Calculate the observed value of the chosen statistic

Shuffle the data and calculate the statistic based on the randomized data

Repeat i times to generate distribution of statistic

Calculate significance (typically as the proportion of $E_{rand} \geq E_{obs}$)

Can be used to evaluate both “standard” and custom statistics

e.g. Are the niches of two species different?

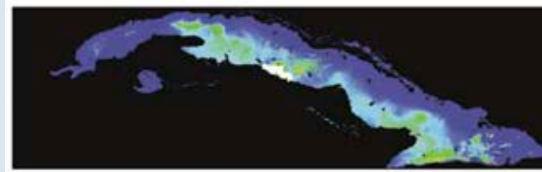
H_0 : the niches are equivalent

Metrics of niche overlap:

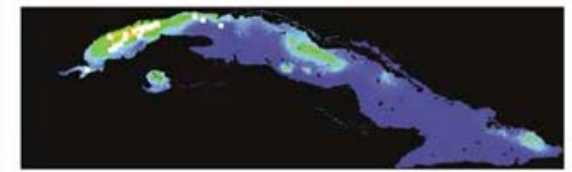
$$D_{px,py} = 1 - \frac{1}{2} \sum |p_{x,i} - p_{y,i}|$$

$$I_{px,py} = 1 - \frac{1}{2} \sqrt{\sum (\sqrt{p_{x,i}} - \sqrt{p_{y,i}})^2}$$

(Combine locals for species, randomly assign to species, re-assess ENM, and test statistics)

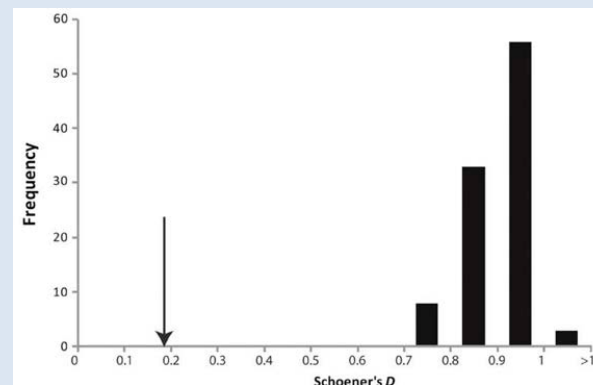


Anolis ahli



Anolis allogus (west)

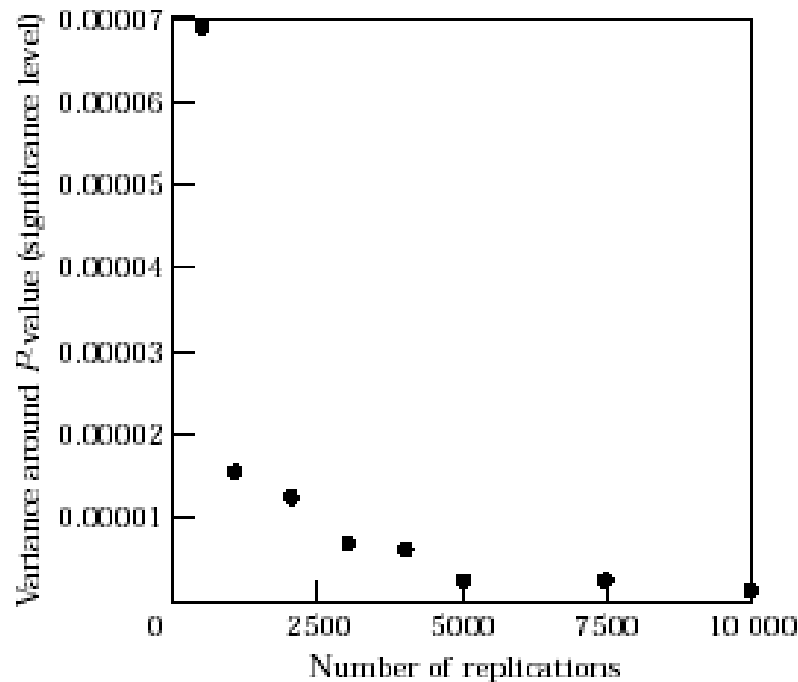
Niche Similarity: $I = 0.47$, $D = 0.19$



All possible permutations not feasible for most cases

Use large number of iterations instead (e.g., 4,999, 9,999 etc.)

↑# iterations improves precision of estimated significance



Extremely flexible for both standard and custom designs

Works for:

- Standard designs (ANOVA, regression, factorial ANOVA)

- Non-standard designs

- High-dimensional data ($p > N$)

When parametric assumptions are met, empirical distributions converge back to theoretical ones (parametric statistics are a special case of permutation-based approaches!!!)

Critical issue: **What** and **How** to resample

Designing a proper resampling test requires:

- 1 - Identifying the null hypothesis (H_0)
- 2 - Having a known expected value under H_0
- 3 - Identifying which values may be shuffled and how to estimate expected distribution under H_0

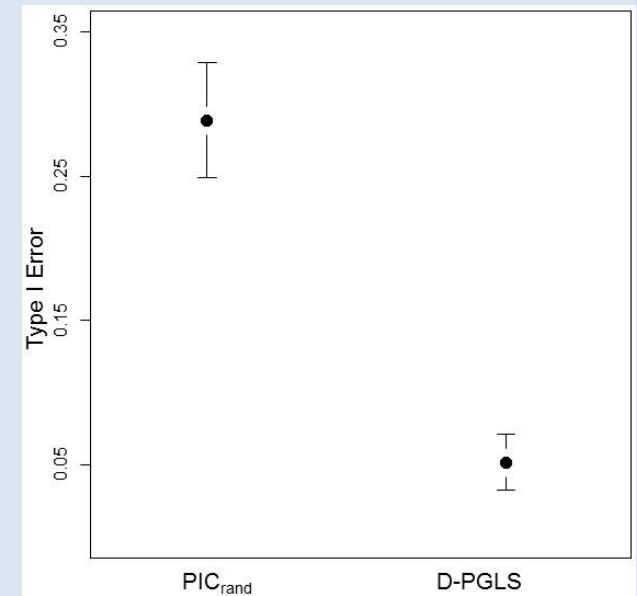
See Collyer, Sekora and Adams (2015). *Heredity*.

e.g. Phylogenetic Comparative tests with high-dimensional data

- Shuffle values at the tips and re-calculate everything
- Shuffle Phylogenetic Independent Contrasts

<u>D-PGLS</u>	df	SS	MS	F	R ²	P _{Yrand}
SVL	1	0.0006586	0.0006586	3.0288	0.07039	0.221 NS
Residual	40	0.0086976	0.0086976			
Total	41	0.0093562	0.00021744			

<u>PIC</u>	df	SS	MS	F	R ²	P _{PICrand}	P _{Yrand}
SVL	1	0.0006586	0.0006586	3.0288	0.07039	0.026	0.221 NS
Residual	40	0.0086976	0.0086976				
Total	41	0.0093562	0.00021744				



PICs are not adequate exchangeable units for this design!

Adams and Collyer (2015). *Evolution*.

Known as "Full Randomization"

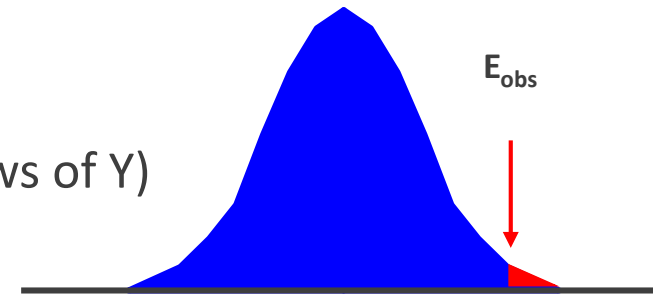
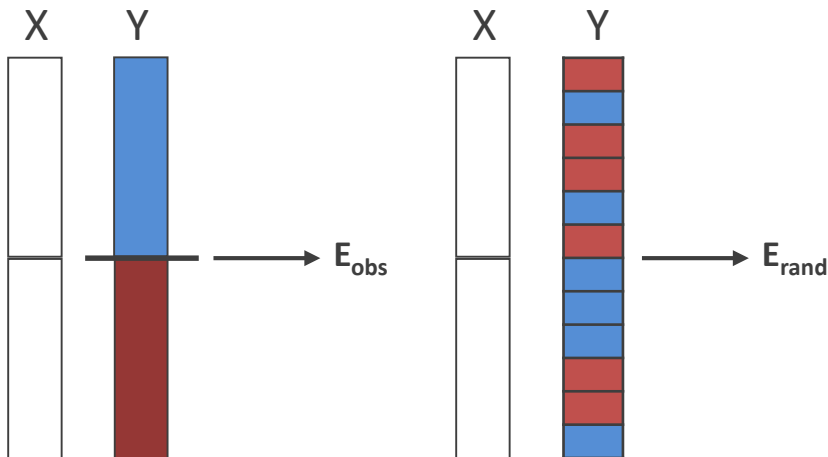
Objective: assess association between X and Y

Choose and calculate observed statistic E_{obs} (e.g. T for T-test, SS or F for ANOVA, F, r or R^2 for regression/correlation)

H_0 : lack of association

Shuffle Y in relation to X (if Y is multivariate, shuffle rows of Y)

Calculate distribution of E_{rand} and compare to E_{obs}



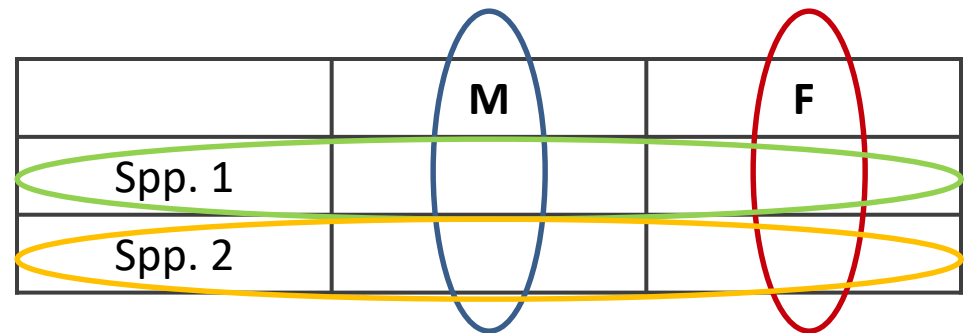
Y	X	Y	X
-0.4	1	-0.4	2.5
0.2	1	0.2	3.7
1.6	1	1.6	5.25
-2.5	1	-2.5	0.3
.	.	.	.
.	.	.	.
.	.	.	.
0.6	2	0.6	2.5
-0.1	2	-0.1	1.25
2.5	2	2.5	5.6
-1.2	2	-1.2	0.4
.	.	.	.
.	.	.	.
.	.	.	.

Restrict the resampling to subsets of data

Useful (and potentially more adequate) in cases where some combinations do not make sense, or when specific combinations are of interest

e.g. two species, with two sexes*

Y	sp	sex
-0.4	1	M
0.2	1	M
1.6	1	F
-2.5	1	F
0.6	2	M
-0.1	2	M
2.5	2	F
-1.2	2	F



Compare species while preserving sex effects

Compare the sexes within each species

*This is just an example, *NOT* a claim that species/sex must be evaluated with restricted randomization!

$$Y \sim A + B + A:B$$

Interaction

Unrestricted randomization (permute Y vs. $A+B+A:B$)

Can test all terms (MS_A , MS_B , & $MS_{A:B}$) as well as MS_{Model}

Often the **wrong** H_0 ! Conflates MS across terms (can yield uninterpretable results)

Restricted randomization (permute Y within A , then within B)

Can test MS_A & MS_B , but not $MS_{A:B}$ (can't evaluate interaction)

DO THIS



See Edgington 1995
Manly 1998

Residual randomization (permute Y_{resid} from sequential H_0 models)

Can test all terms (MS_A , MS_B , & $MS_{A:B}$)

Appropriate H_0 for each term in model

See Collyer, Sekora and Adams (2015). *Heredity*.

$$Y \sim A + B + A:B$$

A set of sequential hypotheses tests

Model	Test	H_0
$Y \sim A$	Does A explain more variation than the grand mean?	$Y \sim 1$
$Y \sim A + B$	Does B A explain more variation than A alone?	$Y \sim A$
$Y \sim A + B + A:B$	Does the interaction help in explaining more variation than A and B separately?	$Y \sim A + B$

Does adding B improve the fit of the model?

Does adding the interaction improve the fit of the model?

Requires appropriate resampling procedure with exchangeable units for each H_0

Residual randomization is the most appropriate for factorial models

Important paper

Anderson and ter Braak 2003. *J. Stat. Comput. Simul.*

Collyer and Adams 2007. *Ecology*.

Collyer et al. 2015. *Heredity*.

Sequentially permute Y_{resid} from reduced model ($H_{0.r}$) with fewer terms

Tests SS for H_1 , while holding constant SS in corresponding H_0

Procedure

- 1: Estimate parameters and observed test statistic (E_{obs}) from full model
- 2: Remove one term from the model; calculate predicted values (\hat{Y}) and residuals (ε) for the reduced model
- 3: Permute residuals (ε) and add them to the predicted values to obtain randomized values
- 4: Calculate test statistic for random data (E_{rand})
- 5: Repeat i times to obtain empirical distribution of statistic

Higher statistical power for factorial designs (Anderson and terBraak, 2003)

Tests appropriate H_0 for each term in model (Collyer, Sekora, and Adams 2015)

Uses correct exchangeable units for hypothesis testing (Collyer, Sekora, and Adams 2015)

Permutation especially useful when non theoretical distribution exists for H_0

Very Common in biology. MUST IDENTIFY CORRECT EXCHANGEABLE UNITS!

e.g. Is there evidence of multivariate character displacement?

H_0 : No difference in shape distance

Test measure: $T = (D_{symp} - D_{allo})$

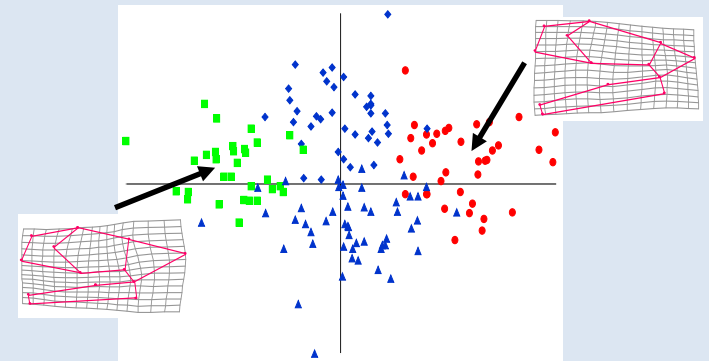
(shuffle individuals among groups)

$$D_{symp} = 0.0753$$

$$D_{allo} = 0.0444$$

$$T = 0.0308$$

$$P_{rand} = 0.0001$$



sympatric *P. cinereus* (green) and sympatric *P. hoffmani* (red)

Conclusion: Strong evidence of character displacement

Adams and Rohlf, 2000. *PNAS*.

High-dimensional data are increasingly common (e.g. morphometrics, genomics, proteomics etc. –omics)

When $p > N$ matrix algebra underlying parametric statistics may fail

e.g. MANOVA design with $p > N$

$|\text{SSCP}_{\text{full}}| = 0$ (matrix is singular)

← Sums of squares and cross products

$\text{SSCP}_{\text{full}}^{-1}$ (inverse) cannot be calculated (divide by zero)

Thus, MANOVA calculations cannot be completed

Check for divide
by ALMOST zero!

Solution: Use resampling-based tools

- 1: Assess significance using other model parameters
- 2: Use distance-based approaches

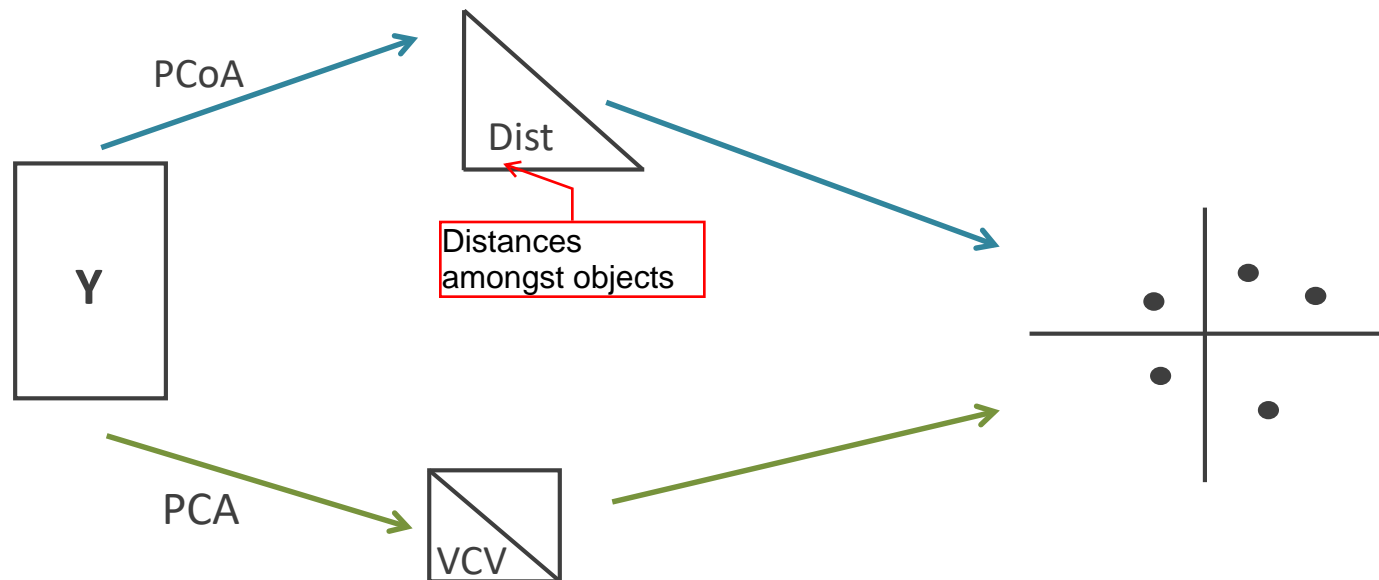
Test significance using parameters estimable when $p > N$

- 1: Obtain original test-statistics (T_{obs}): $\text{tr}(\text{SSPC}_{\text{model}})$, $D_{\text{gp1, gp2}}$, etc.
- 2: Shuffle data & calculate T_{rand}
- 3: Compare T_{obs} vs. T_{rand}
- 4: Repeat

Doesn't require inverting covariance matrix, so general solution

Test significance based on distances between objects

Relies on equivalence between covariance matrix and distance matrix



MANOVA: covariance-based

Permutational-MANOVA: distance-based version

Sums of Squares found from distances

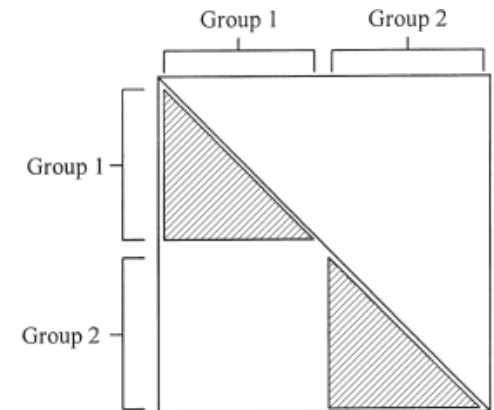
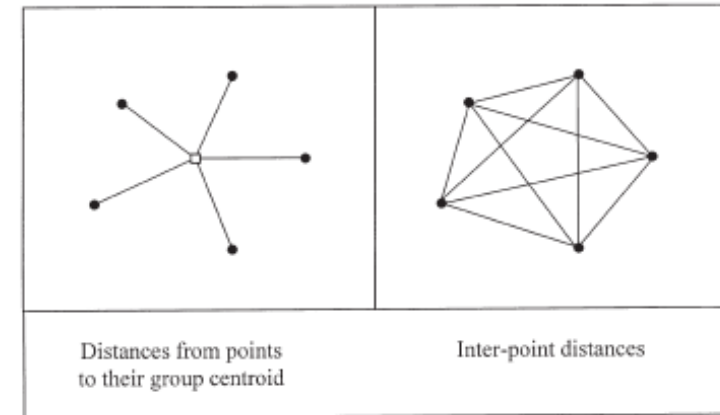
$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2$$

$$SS_W = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 e_{ij}$$

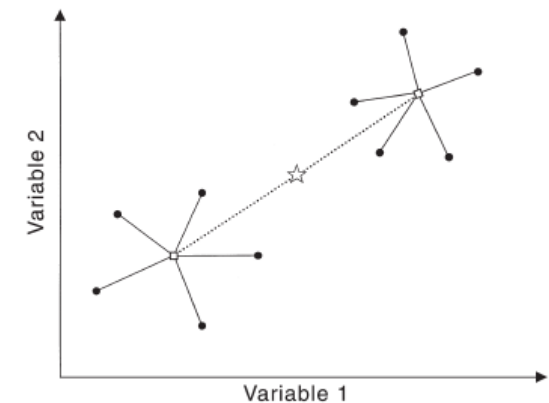
Estimate F_{obs}

$$F_{\text{obs}} = \frac{(SS_T - SS_W)/(a - 1)}{SS_W/(N - a)}$$

This does not
require more
specimens than
variables



Shuffle data, estimate F_{rand} , compare, repeat



*Method identical to Procrustes ANOVA and AMOVA

Randomization/Permutation

Resampling without replacement ($n' = n$)


Bootstrap

Resampling with replacement ($n' = n$)

Jackknife

Leave-one-out resampling ($n' = n-1$)

Monte Carlo simulations



Take objects out of bag, observe,
and put object back in bag.

Useful for estimating confidence intervals (among other uses)

Several approaches exist

Proposed to alleviate biases in estimating σ

Procedure:

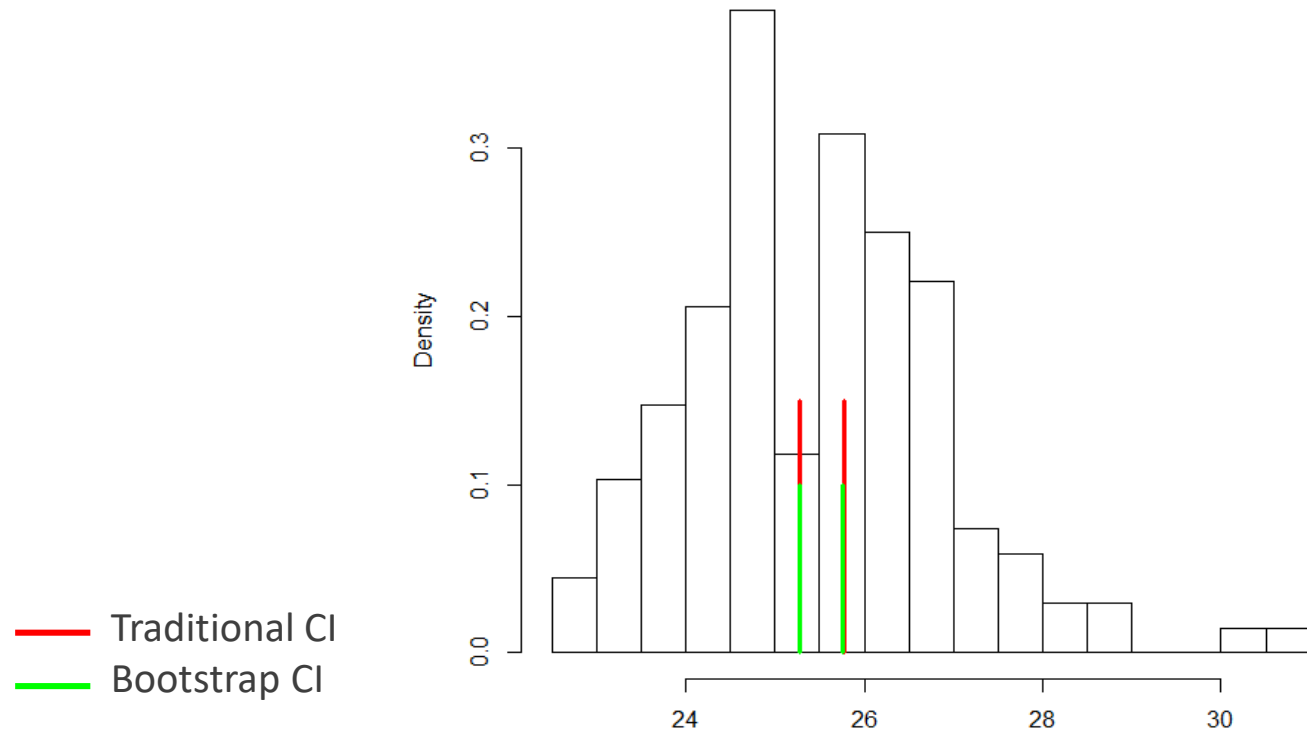
Generate i bootstrap datasets

Estimate test statistic (e.g. the mean) for each

Find σ^* from bootstrap test statistics

Calculate $CI = statistic \pm Z_{\alpha/2}\sigma^*$

For bootstrapping: substitute standard deviation with bootstrapped standard deviation



Proposed to alleviate deviances from the normal distribution

Procedure:

Generate i bootstrap datasets

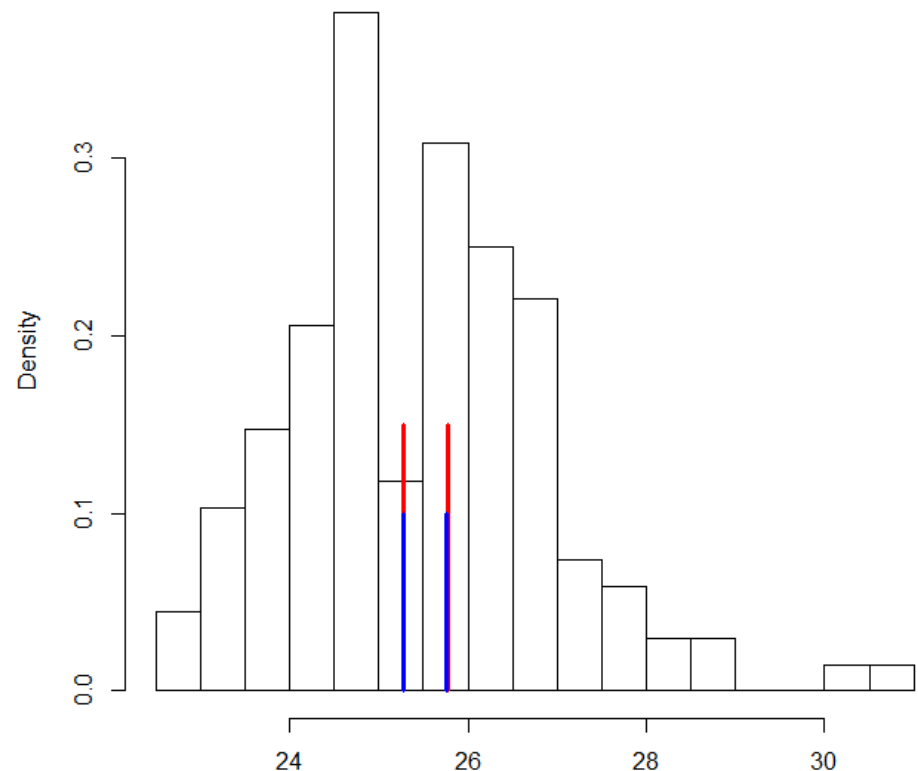
Estimate test statistic (e.g. the mean) for each

Bootstrap CI: upper and lower $\alpha/2$ % (usually 0.025 and 0.975, for $\alpha=0.05$)

Similar to doing standard deviations, but take upper and lower percentages and that's the CI

This approach assumes that the distribution of the bootstrap test statistic is centered on the observed test statistic

— Traditional CI
— Bootstrap CI



Accounts for when >50% of bootstrap test statistics are above or below the observed value (i.e. skews in the distribution)

“Slides” the percentiles a bit

Procedure:

- Generate i bootstrap datasets

- Estimate test statistic (e.g. the mean) for each

- Find fraction (Fr) of bootstrap values above /below observed test statistic

$$CI = \Phi[2\Phi^{-1}(Fr) \pm Z_{\alpha/2}]$$

Φ is the cumulative normal distribution

Felsenstein (1985) proposed using the bootstrap to assess confidence in the topology of phylogenetic trees

Procedure:

- Calculate phylogenetic tree from data (e.g. using ML, Bayesian, or other approach)

- Bootstrap dataset and recalculate tree

- Proportion of nodes in bootstrapped trees is used to measure support for that node in the observed tree

The measured characters are assumed to be representative of true character set

The bootstraps generate alternative character matrices

Caution with interpretations:

- Bootstrap estimates on nodes are NOT independent

- Bootstrap values are often large at base and tips, smaller in middle
(result of combinatorics branching theory)

Randomization/Permutation

Resampling without replacement ($n' = n$)

Bootstrap

Resampling with replacement ($n' = n$)

Jackknife

SUCKS

Leave-one-out resampling ($n' = n-1$)

Monte Carlo simulations

... like a Boy Scout's jackknife, it is a "rough and ready" tool that can solve a variety of problems even though specific problems may be more efficiently solved with a purpose-designed tool.

Asks how precise is the observed estimate (i.e. how sensitive it is to specific values in the dataset)

Typically used to estimate bias, standard errors, CIs of test statistics


Calculate observed test statistic E_{obs}

Remove one observation and calculate jackknife estimate of statistic E_{jack}

Repeat many times, removing a different object in each iteration

Calculate mean of jackknife estimates \bar{E}_{jack}

$$\text{Bias} = E_{\text{obs}} - \bar{E}_{\text{jack}}$$



Provides a measure of bias; seeing how influential each piece is on the total.

Randomization/Permutation

Bootstrap

Jackknife

Random sampling of original data

Resample original data set

Monte Carlo simulations

Data simulated based on a model & parameters from original data

Simulate more data sets

Use model parameters to simulate data, from which E_{rand} is generated

Values in simulated dataset are NOT from the original dataset

However, parameters for the model are estimated from the original data

aka “parametric bootstrap”

Assumes that the observed data is a **representative sample**, such that this is used to generate other representative samples, and evaluate patterns in the original dataset as compared to other, simulated datasets

1: Are plants distributed randomly in forest?

Calculate point-pattern statistic of actual plants

Simulate random plant locations (using RandUnif, or other model) and compare patterns

2: Are species 'evenly' distributed among communities?

Calculate evenness measure (E) for actual communities

Simulate random communities from a community-assembly model and compare E_{rand} to E_{obs}

In E&E, one often hears of 'parametric bootstrap' for hypothesis testing and generation of confidence intervals. This is a Monte Carlo procedure

Very useful and flexible

Much more powerful than rank-based non parametric approaches

Can be used to assess significance when data don't meet some assumptions

Useful when no theoretical distribution exists (e.g. 2B-PLS)

Can be implemented for any kind of custom design

Can be easily implemented in any programming language

Careful when designing custom permutational procedures: require correct H_0 and exchangeable units

What you shuffle matters!!!