Synthesizing information from multivariate data: Inference methods for global and local questions

Arne Bathke
Arne.Bathke@sbg.ac.at

Universität Salzburg, Austria

Wien, 28 May 2021

Thanks to: Solomon W. Harrar, Yvonne Höller, Robert Knapp, Frank Konietschke, Markus Pauly, Sarah Friedrich

PARIS LODRON UNIVERSITÄT SALZBURG

Global View

- Interesting data are multivariate
- Here: multivariate = multiple endpoints

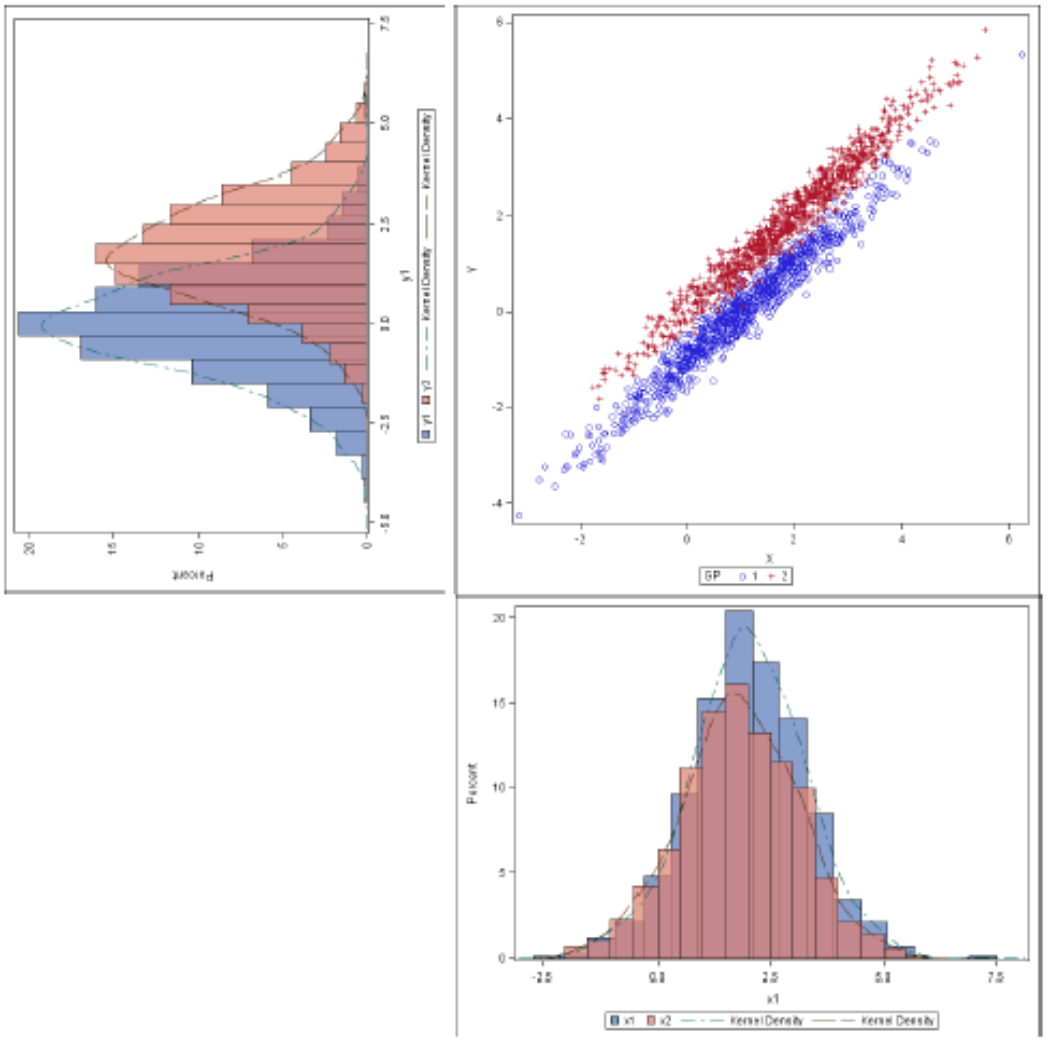
Strategies

Strategies for inference

- One response variable at a time
  - parametric / semi- / nonparametric
  - adjusting for multiple testing
  - combining p-values
- classical MANOVA
- semiparametric MANOVA (using bootstrap)
- semiparametric MANCOVA (using bootstrap)
- nonparametric MANOVA (using ranks)
- supplementing the above by an appropriate multiple testing tree

PARIS LODRON UNIVERSITÄT SALZBURG

Joint and marginal distributions



Major differences may be hidden from the univariate eye.

Strategies

Strategies for inference

- ~~One response variable at a time~~
  - parametric / semi- / nonparametric
  - adjusting for multiple testing
  - combining p-values
- classical MANOVA
- semiparametric MANOVA (using bootstrap)
- semiparametric MANCOVA (using bootstrap)
- nonparametric MANOVA (using ranks)
- supplementing the above by an appropriate multiple testing tree

PARIS LODRON UNIVERSITÄT SALZBURG

Research questions

- Situation: data with several endpoints (responses) and several factor levels (experimental conditions)
- Global questions
  - Is there a difference between experimental conditions when using the information from all endpoints?
- Local questions ("finding the needle in the haystack")
  - If yes, on which endpoints?
  - And between which conditions?
  - Is there a condition that works better than others?
  - *Control familywise α when answering local questions*
    [trying to support research reproducibility!]

PARIS LODRON UNIVERSITÄT SALZBURG

Data Example I: Insect-Flower Interaction

- insects from different taxa may visit different flowers
- flower traits = response variables (height, nectar-tube depth, display size, etc.)
- taxa = subpopulations = factor levels (bees, beetles, bumble bees, ants, etc.)
- individual insects = experimental units
- as illustration, consider $a = 4$ taxa and $p = 3$ traits

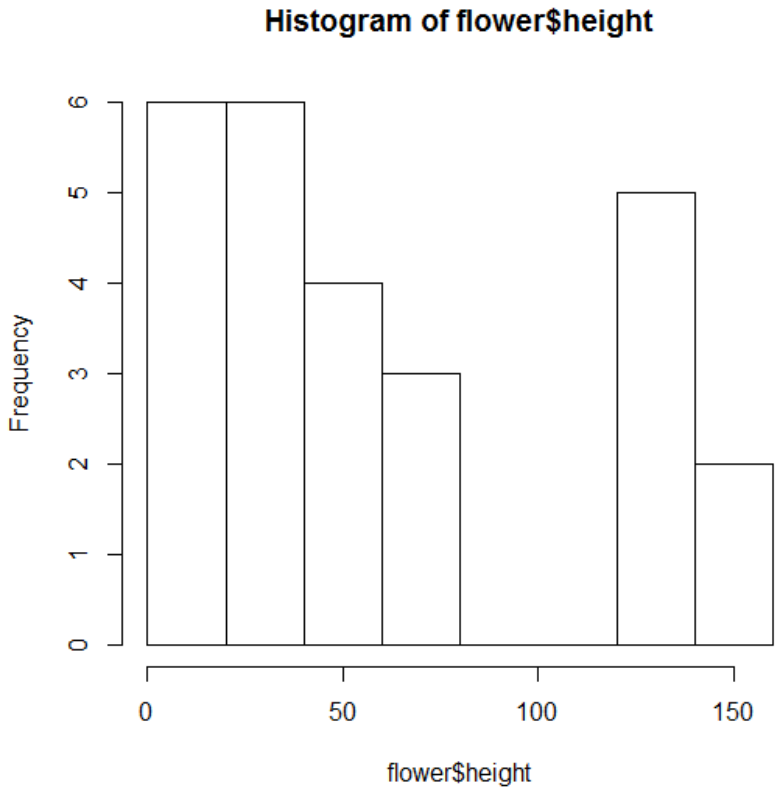
PARIS LODRON UNIVERSITÄT SALZBURG

Response variables = Traits

- height
- nectar tube depth
- display size (how many flowers/inflorescences per species)

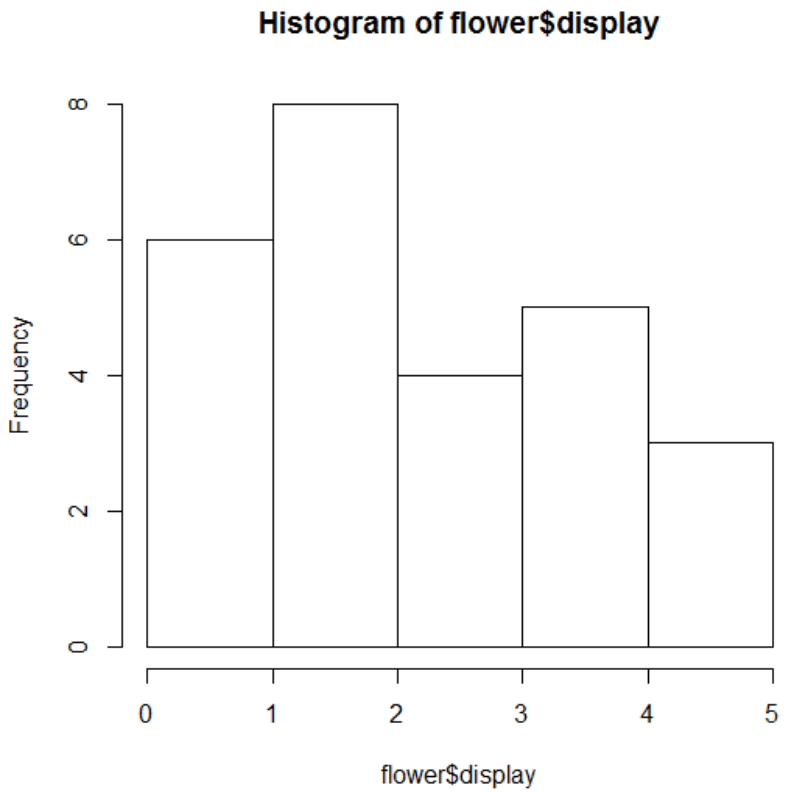
MANOVA? Or looking at one variable at a time... (1)

- height



MANOVA?? Or looking at one variable at a time... (2)

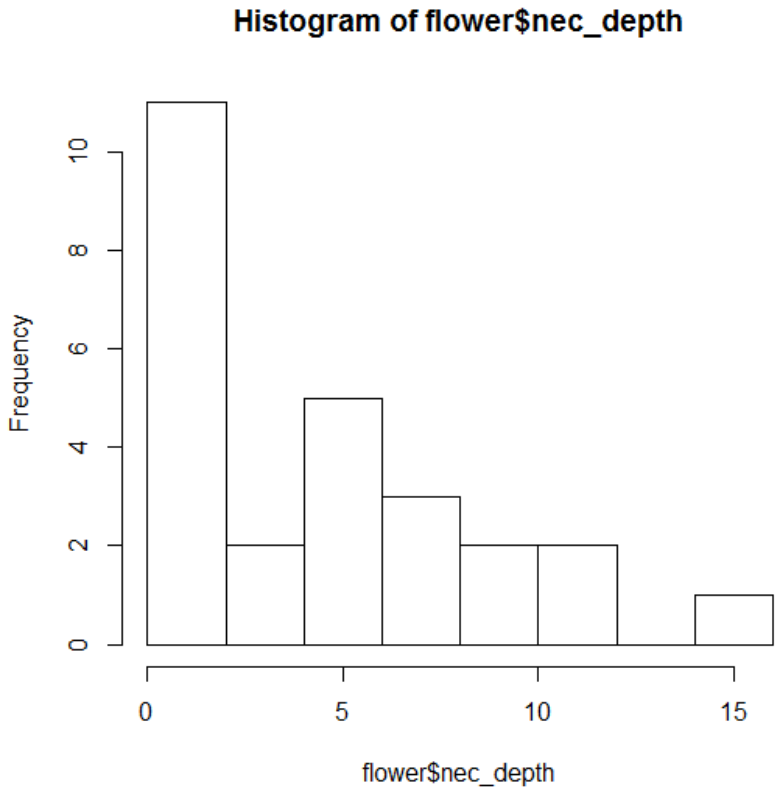
- display size



PARIS LODRON UNIVERSITÄT SALZBURG

MANOVA??? Or looking at one variable at a time... (3)

- nectar tube depth



Let's try parametric MANOVA

```
Y=cbind(flower$height,flower$nec_depth,flower$display)
trt=as.factor(flower$animal)
flower.manova=manova(Y ~ trt,data=flower)
summary(flower.manova,test="Pillai")
summary(flower.manova,test="Wilks")
```

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
trt	3	0.62199	1.9181	9	66	0.06438

	Df	Wilks	approx F	num Df	den Df	Pr(>F)
trt	3	0.42897	2.2563	9	48.825	0.03358

PARIS LODRON UNIVERSITÄT SALZBURG

Let's try MANOVA (2)

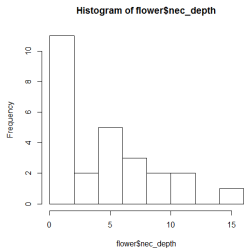
	Df	Pillai	approx F	num Df	den Df	Pr(>F)
trt	3	0.62199	1.9181	9	66	0.06438

	Df	Wilks	approx F	num Df	den Df	Pr(>F)
trt	3	0.42897	2.2563	9	48.825	0.03358

"Nectar tube depth" was rather skewed, take logarithm instead ...

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
trt	3	0.67318	2.1216	9	66	0.03972

	Df	Wilks	approx F	num Df	den Df	Pr(>F)
trt	3	0.39646	2.5091	9	48.825	0.01902



PARIS LODRON UNIVERSITÄT SALZBURG

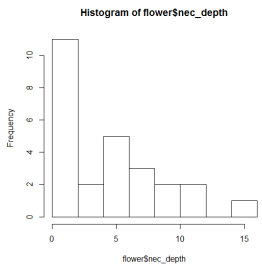
Let's try MANOVA (3)

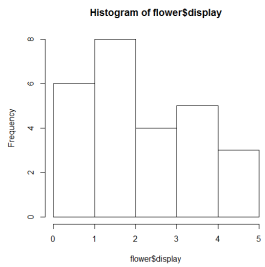
raw data		Df	statistic	approx F	num Df	den Df	Pr(>F)
Pillai	trt	3	0.62199	1.9181	9	66	0.06438
Wilks	trt	3	0.42897	2.2563	9	48.825	0.03358

log:nec		Df	statistic	approx F	num Df	den Df	Pr(>F)
Pillai	trt	3	0.67318	2.1216	9	66	0.03972
Wilks	trt	3	0.39646	2.5091	9	48.825	0.01902

count variable "display size" also skewed, take its logarithm as well ...

log:nec,display		Df	statistic	approx F	num Df	den Df	Pr(>F)
Pillai	trt	3	0.75693	2.4746	9	66	0.01689
Wilks	trt	3	0.35551	2.8726	9	48.825	0.008395





PARIS LODRON UNIVERSITÄT SALZBURG

Preliminary Confusion

1. parametric MANOVA may lead to confusing results
2. Reminder: parametric MANOVA assumes multivariate normality of the data
3. univariate normality is sometimes hard to justify
4. multivariate normality is usually quasi impossible to justify

Strategies

Strategies for inference

- ~~One response variable at a time~~
- ~~classical MANOVA~~
- semiparametric MANOVA (using bootstrap)
- semiparametric MANCOVA (using bootstrap)
- nonparametric MANOVA (using ranks)
- supplementing the above by an appropriate multiple testing tree
  - parametric / semi- / nonparametric
  - adjusting for multiple testing
  - combining p-values

PARIS LODRON UNIVERSITÄT SALZBURG

Data Example II: Dementia, Alzheimer's Disease (AD)

- Demographic development in western countries comes with growing incidence of dementia (about 150,000 affected in Austria)
- Accurate and early diagnosis desirable
- Facilitates early treatment and perhaps prevention of dementing course

PARIS LODRON UNIVERSITÄT SALZBURG

Alzheimer / EEG / SPECT

- 160 patients with either Alzheimer's disease (AD), mild cognitive impairment (MCI), or subjective cognitive complaints without clinically significant deficits (SCC)
- neuropsychological diagnostics for evaluation of cognitive impairment included test batteries for dementia, memory, intelligence, education and emotional status
- do the groups differ w.r.t. single photon emission computed tomography (SPECT) or electroencephalography (EEG)?
- EEG values of activity, complexity, mobility, and brain rate at five regions, 46 different SPECT variables
- high-dimensional response setting
- potential addtional factors: age, sex
- interactions between these and the neuropsychological diagnosis?

PARIS LODRON UNIVERSITÄT SALZBURG

SPECT

- single-photon emission computed tomography (SPECT)
- well examined tool
- used to differentiate AD from other forms of dementia (frontotemporal dementia, dementia with Lewy bodies)
- considered cheap

EEG

- considered much cheaper
- electroencephalogram (SPECT)
- highly available
- free of radiation hazards, non-invasive
- useful as diagnostic tool in early-onset dementia

We'll get back to these data soon, but first some theory...

Notation for (non/semi/parametric) multivariate CR1F

- p different response variables (endpoints) $\quad k = 1, \ldots, p$
- a different experimental conditions (treatments, sub-populations) $i = 1, \ldots, a$
- n_i subjects (experimental units) per condition $\quad j = 1, \ldots, n_i$

Sample 1				Sample 2				...	Sample a			
$X_{11}^{(1)}$	$X_{12}^{(1)}$	...	$X_{1n_1}^{(1)}$	$X_{21}^{(1)}$	$X_{22}^{(1)}$	...	$X_{2n_2}^{(1)}$	...	$X_{a1}^{(1)}$	$X_{a2}^{(1)}$	...	$X_{a,n_a}^{(1)}$
$X_{11}^{(2)}$	$X_{12}^{(2)}$	...	$X_{1n_1}^{(2)}$	$X_{21}^{(2)}$	$X_{22}^{(2)}$	...	$X_{2n_2}^{(2)}$	...	$X_{a1}^{(2)}$	$X_{a2}^{(2)}$	...	$X_{a,n_a}^{(2)}$
	...				...			...		...		
$X_{11}^{(p)}$	$X_{12}^{(p)}$	...	$X_{1n_1}^{(p)}$	$X_{21}^{(p)}$	$X_{22}^{(p)}$	...	$X_{2n_2}^{(p)}$	...	$X_{a1}^{(p)}$	$X_{a2}^{(p)}$	...	$X_{a,n_a}^{(p)}$

- Ranks denoted by R instead of X, where each row (each variable) is ranked separately

PARIS LODRON UNIVERSITÄT SALZBURG

Classical vs. Semiparametric (Additive Location) Model

- Classical parametric MANOVA

  $$(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim N_p(\mu_i, \Sigma), \;\; i = 1, \ldots, a; \;\; j = 1, \ldots, n_i;$$

  X_{ij} independent random vectors

- Alternative Model 1: Semiparametric MANOVA using multivariate linear model

  $$X_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, a; j = 1, \ldots, n_i.$$

  For each i, the $\varepsilon_{i1}, \ldots, \varepsilon_{in_i}$ are i.i.d. p-dimensional random vectors satisfying

  $$E(\varepsilon_{i1}) = 0,$$
  $$Cov(\varepsilon_{i1}) = \Sigma_i > 0, \;\; i = 1, \ldots, a,$$
  $$E(\|\varepsilon_{i1}\|^4) < \infty, \;\; i = 1, \ldots, a.$$

- Null hypothesis: $H_0^{\mu} : \mu_1 = \cdots = \mu_a$

PARIS LODRON UNIVERSITÄT SALZBURG

Assumption of Homoscedasticity

- Parametric MANOVA assumes that covariance matrices are the same for each group
- severe restriction in practice!
- Violation of the covariance matrix homogeneity assumption may cause serious problems with MANOVA,
  even under normality
- Similar to univariate case, in particular for unbalanced designs
- Distinguish *positive* and *negative pairing* of group size n_i with variance σ_i^2
- For nominal $\alpha = 0.05$, the simulated α may be ...
  - ... around 0.01 for *positive pairing* ($n_2 = 2 \cdot n_1$, $\sigma_2^2 = 3 \cdot \sigma_1^2$)
  - ... around 0.20 for *negative pairing*

PARIS LODRON UNIVERSITÄT SALZBURG

Equal Covariance Matrices? EEG Data

Table 1: *Covariance matrices for the three impairment diagnosis groups AD, MCI and SCC, calculated for six EEG response variables. Variables 1-3 are temporal, frontal, and cerebellal values for brain rate, variables 4-6 corresponding values for complexity. For ease of presentation, the covariance matrices are displayed in tabular form.*

AD	1	2	3	4	5	6
	5.14	5.04	4.94	5.63	4.36	4.46
	5.04	6.55	5.21	5.74	5.82	4.83
	4.94	5.21	6.35	5.39	4.55	6.63
	5.63	5.74	5.39	8.88	6.92	6.64
	4.36	5.82	4.55	6.92	7.88	7.15
	4.46	4.83	6.63	6.64	7.15	13.84
MCI	1	2	3	4	5	6
	2.10	1.95	1.76	1.45	1.25	0.69
	1.95	2.18	1.82	1.59	1.61	0.86
	1.76	1.82	2.11	1.41	1.21	1.08
	1.45	1.59	1.41	2.23	2.35	1.19
	1.25	1.61	1.21	2.35	2.95	1.23
	0.69	0.86	1.08	1.19	1.23	1.03
SCC	1	2	3	4	5	6
	1.62	1.17	1.17	0.76	0.49	0.32
	1.17	1.41	1.10	0.63	0.75	0.37
	1.17	1.10	1.26	0.54	0.39	0.41
	0.76	0.63	0.54	0.64	0.53	0.30
	0.49	0.75	0.39	0.53	0.94	0.28
	0.32	0.37	0.41	0.30	0.28	0.28

PARIS LODRON UNIVERSITÄT SALZBURG

Classical vs. Semiparametric (Additive Location) Model

- Classical parametric MANOVA

  $$(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim N_p(\mu_i, \Sigma), \;\; i = 1, \ldots, a; \;\; j = 1, \ldots, n_i;$$

  X_{ij} independent random vectors

- Alternative Model 1: Semiparametric MANOVA using multivariate linear model

  $$X_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, a; j = 1, \ldots, n_i.$$

  For each i, the $\varepsilon_{i1}, \ldots, \varepsilon_{in_i}$ are i.i.d. p-dimensional random vectors satisfying

  $$E(\varepsilon_{i1}) = 0,$$
  $$Cov(\varepsilon_{i1}) = \Sigma_i > 0, \;\; i = 1, \ldots, a,$$
  $$E(\|\varepsilon_{i1}\|^4) < \infty, \;\; i = 1, \ldots, a.$$

- Null hypothesis: $H_0^\mu : \mu_1 = \cdots = \mu_a$

PARIS LODRON UNIVERSITÄT SALZBURG

Classical vs. Nonparametric Model

- Classical parametric MANOVA assumes the model

$$(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim N_p(\mu_i, \Sigma), \; i = 1, \ldots, a; \; j = 1, \ldots, n_i;$$

X_{ij} independent random vectors

(note multivariate normality and equal covariance matrices)

- Alternative Model 2: Nonparametric (rank-based) MANOVA:

$$(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim F_i, \; i = 1, \ldots, a; \; j = 1, \ldots, n_i;$$

X_{ij} independent random vectors

- Null hypotheses: $H_0^\mu : \mu_1 = \cdots = \mu_a$ or $H_0^F : F_1 = \cdots = F_a$

PARIS LODRON UNIVERSITÄT SALZBURG

Review: Deriving (M)ANOVA test statistics

- Goal: Multivariate (M) Analysis of Variance (ANOVA)
- Recall: ANOVA
- a groups with respective sample sizes n_i; $N = \sum_{i=1}^{a} n_i$

$$F = H/E \qquad \text{where}$$

$$H = \frac{1}{a-1} \sum_{i=1}^{a} n_i (\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{..})^2 \qquad \text{and}$$

$$E = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{i.})^2.$$

- Under normality, equal variances, and null hypothesis: $F \sim F(a-1, N-a)$.

Review: Deriving (M)ANOVA test statistics

- Multivariate (M) Analysis of Variance (ANOVA)
- a groups with respective sample sizes n_i; $N = \sum_{i=1}^{a} n_i$
- p variables

$$H(\mathbf{X}) = \frac{1}{a-1} \sum_{i=1}^{a} n_i (\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{..})(\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{..})' \qquad \text{and}$$

$$E(\mathbf{X}) = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{i.})'.$$

- How to combine these into one test statistic?

PARIS LODRON UNIVERSITÄT SALZBURG

Review: Classical MANOVA test statistics

Lawley-Hotelling's trace: $T_{LH} = \mathrm{tr}\left(HE^{-}\right) = \sum \lambda_l$

Bartlett-Nanda-Pillai: $T_{BNP} = \mathrm{tr}\left(H(H+E)^{-}\right) = \sum \frac{\lambda_l}{1+\lambda_l}$

Wilks' Lambda: $T_{WL} = -\log \frac{\det(E)}{\det(E+H)} = \prod \frac{1}{1+\lambda_l}$

where A^{-} is the Moore-Penrose generalized inverse of A,
λ_l are the eigenvalues of HE^{-1}

- Classical MANOVA assumes multivariate normality.
- Still, null distributions rather complicated.

PARIS LODRON UNIVERSITÄT SALZBURG

Proposed Test Statistics

- Nonparametric
  - Rank-based variation on Wilks' Lambda
  - Sampling distribution: approximated by F with estimated d.f. (moment approximation)
- Semiparametric
  - Wald-Type Statistic $N \cdot \bar{\mathbf{X}}_{\cdot}' \mathbf{T} (\mathbf{T} \hat{\mathbf{V}}_N \mathbf{T})^{+} \mathbf{T} \bar{\mathbf{X}}_{\cdot}$
  - Sampling distribution: Parametric bootstrap; generating normal random vectors using group-specific empirical covariance matrices (*iid* only within the groups)

PARIS LODRON UNIVERSITÄT SALZBURG

Not so Many Assumptions: Two Models

- Alternative Model 1:
- Additive location model
  - endpoints should be metric variables
  - additivity should be justifiable
  - hypotheses formulated using mean vectors
  - advantages when performing a closed testing procedure in order to choose relevant variables
  - test statistics in terms of observed values
  - sampling distribution: we propose asymptotic model-based bootstrap
  - very flexible theory, works for pretty much any factorial design with multiple endpoints (even for repeated measures)
  - R package MANOVA.RM (0.5.1)

PARIS LODRON UNIVERSITÄT SALZBURG

Not so Many Assumptions: Two Models

- Alternative Model 2:
- Fully nonparametric model
  - endpoints may be metric, ordinal, binary (or mix thereof)
  - hypotheses formulated using multivariate distributions
  - test statistics expressed in terms of endpoint-wise ranks, and based on nonparametric relative effect
  - sampling distribution: we propose F with moment approximation
  - asymptotic distributions (large a, large n), Cornish Fisher expansions, permutations also available
  - highly robust; invariant under endpoint-wise strictly monotone (isotone or antitone!) transformations
  - higher-way layouts somewhat tedious, but possible
  - R package npmv (2.4.0) for multivariate one-way layout

PARIS LODRON UNIVERSITÄT SALZBURG

Literature

- Nonparametric MANOVA:
  AR Ellis & WW Burchett & SW Harrar & AB 2017, Nonparametric inference for multivariate data: the R package npmv. J Stat Soft 76, 4.
- Semiparametric MANOVA:
  F Konietschke & AB & SW Harrar & M Pauly 2015, Parametric and nonparametric bootstrap methods for general MANOVA. J Multivariate An 140, 291–301.
- Semiparametric MANOVA and Repeated Measures:
  AB & S Friedrich & M Pauly & F Konietschke & W Staffen & N Strobl & Y Höller 2018, Testing mean differences among groups: multivariate and repeated measures analysis with minimal assumptions. Mult Beh Res
- Nonparametric Combination:
  R Arboretti & AB & E Carrozzo & F Pesarin & L Salmaso 2019, Multivariate permutation tests for two sample testing . . . Stat Meth Med Res
- Semiparametric MANCOVA:
  G Zimmermann & M Pauly & AB 2020 J Multivariate An

PARIS LODRON UNIVERSITÄT SALZBURG

Nonparametric MANOVA

- Nonparametric MANOVA:

$$(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim F_i, \; i = 1, \ldots, a; \; j = 1, \ldots, n_i;$$

X_{ij} independent random vectors

F_i are p-variate distributions

- Null hypothesis: $H_0^F : F_1 = \cdots = F_a$
- Based on this model, we have developed
  1. asymptotic theory
  2. small sample approximations (expansions, moment estimators, permutations)
  3. and the R package `npmv`

for nonparametric inference of multivariate data

PARIS LODRON UNIVERSITÄT SALZBURG

Nonparametric MANOVA and npmv: flower data

```
install.packages("npmv")
library(npmv)
library(Formula)

nonpartest(height | nec_depth | display ~ animal, flower)
```

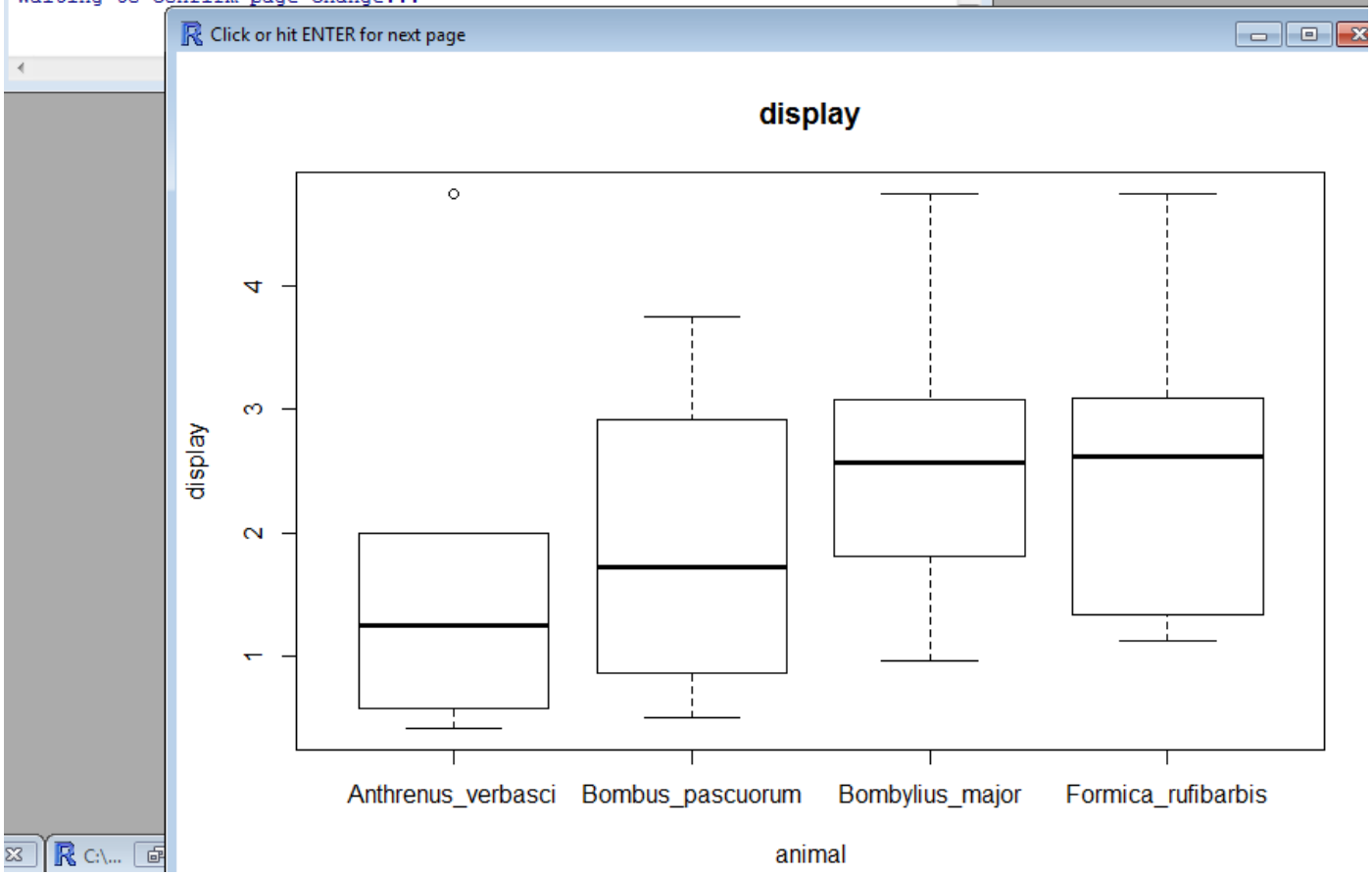
Nonparametric MANOVA using npmv

Results . . .

No! First a few pictures!

Nonparametric MANOVA using npmv – First: Figures

```
> nonpartest(height | nec_depth| display ~ animal, flower, permreps=1000)
Waiting to confirm page change...
```



Nonparametric MANOVA using npmv – Second: Numbers

$results

	Test Statistic	P-value	Permutation Test P-value
ANOVA type test	2.588	0.020	0.016
McKeon approx. for the Lawley Hotelling Test	3.292	0.007	0.008
Muller approx. for the Bartlett-Nanda-Pillai Test	2.238	0.021	0.021
Wilks Lambda	2.813	0.010	0.013

$releffects	height	nec_depth	display
Anthrenus_verbasci	0.603	0.301	0.353
Bombus_pascuorum	0.365	0.769	0.423
Bombylius_major	0.442	0.611	0.596
Formica_rufibarbis	0.609	0.282	0.596

PARIS LODRON UNIVERSITÄT SALZBURG

Sidenote: Relative Effects

- the natural quantity upon which these (and several other) nonparametric tests are based
- basically an extension of $P(X < Y)$
- "Assume that you randomly choose a B.p. and you randomly choose any insect from your trial. Then, the estimated probability that B.p. seeks out the flower with longer nectar-tube depth is 0.769"
- larger relative effects for one group indicate a tendency to larger observations
- 1/2 indicates "no tendency"
- (basically) a univariate measure

$releffects	height	nec_depth	display
Anthrenus_verbasci	0.603	0.301	0.353
Bombus_pascuorum	0.365	0.769	0.423
Bombylius_major	0.442	0.611	0.596
Formica_rufibarbis	0.609	0.282	0.596

PARIS LODRON UNIVERSITÄT SALZBURG

Preliminary Conclusions

1. It is possible to do valid inference for multivariate data that does not follow normal distributions
2. The assumed model underlying the nonparametric method is quite general
3. Inference invariant under monotone transformations of individual variables
4. Inference should never stand alone – practitioners shall always be confronted with graphs of their data, and with numerical effect measures that (hopefully) can be interpreted

PARIS LODRON UNIVERSITÄT SALZBURG

Some facts about the R package npmv

- Altogether, eight tests are being calculated
- Generally, Wilks' Lambda is recommended
- For high-dimensional data, the only test that can always be used is the ANOVA-type test: use whenever Wilks' Lambda is not available
- $N < 10$: Permutation test;
  $10 \leq N < 30$: Randomization test with 10,000 permutations;
  $30 \leq N$: F approximation

PARIS LODRON UNIVERSITÄT SALZBURG

Now what?

- Global (nonparametric) MANOVA has shown that the (a) taxa differ in their preferred niches of the (p) traits
- So what?
- Maybe not the main question that is of interest to the biologist?
- Real questions:
  - which traits matter
  - which taxa differ from each other
- Conjectures from the figures and relative effects:
  - nectar depth seems to be a discriminating trait
  - Anthrenus verbasci (and Formica rufibarbis) seem to act differently from the other insects

PARIS LODRON UNIVERSITÄT SALZBURG

Finding the interesting conditions / treatments

- Multiple testing approach with familywise error control
- Possible hypotheses using three treatments A, B, C and the multivariate distribution functions

  A B C $\left[\widehat{=} \quad F_A^{(multiv.)} = F_B^{(multiv.)} = F_C^{(multiv.)}\right]$

  A B A C B C
- This family of hypotheses is closed under intersection
  $\Longrightarrow$ Closed testing procedure (Marcus/Peritz/Gabriel 1976) can be used
- closure test is coherent and controls familywise error rate ... without having to adjust the local α for the individual hypothesis tests

PARIS LODRON UNIVERSITÄT SALZBURG

Finding the interesting conditions II

- Possible hypotheses using four treatments A, B, C, D and the multivariate distribution functions

  A B C D
  A B C A B D A C D B C D
  A B A C A D B C B D C D

- This family of hypotheses is. . .
  . . . *not* closed under intersection!
- need to take the partial hypotheses into account, even though they may not be of interest themselves:
  (A B) (C D) (A C) (B D) (A D) (B C)

PARIS LODRON UNIVERSITÄT SALZBURG

Finding the interesting conditions III

- Possible hypotheses using four treatments A, B, C, D and the multivariate distribution functions

  A B C D

  A B C A B D A C D B C D

  (A B) (C D) (A C) (B D) (A D) (B C)

  A B A C A D B C B D C D

- This family of hypotheses is closed under intersection!
  $\Longrightarrow$ Closed testing procedure (Marcus/Peritz/Gabriel 1976) could technically be used

PARIS LODRON UNIVERSITÄT SALZBURG

When finding interesting conditions, take care of partial hypotheses

- However: the number of partial hypotheses grows rather fast

factor levels	4	5	10	13	20
all hypotheses in family closed under intersection	14	51	115,974	27,644,436	51,725,158, ,235,371

- Solution: testing partial hypotheses implicitly, with "partial bonferronization"

Partial Bonferronization

- Possible hypotheses using four treatments A, B, C, D and the multivariate distribution functions

  Total: 4 — A B C D

  Subset of 3: A B C, A B D, A C D, B C D

  Pesky partials: (A B) (C D), (A C) (B D), (A D) (B C)

  Subset of 2: A B, A C, A D, B C, B D, C D
- Test each of the subsets of size 2 at $\alpha \cdot 2/4$
- Consider (A B)(C D) "rejected" if A B or C D is rejected, etc.
- Resulting procedure is coherent, and controls familywise error rate at α
- And fast! Only $2^a - a - 1$ hypotheses need to be tested

PARIS LODRON UNIVERSITÄT SALZBURG

Partial Bonferronization saves time!

factor levels	4	5	10	13	20
all hypotheses in family closed under intersection	14	51	115,974	27,644,436	51,725,158, ,235,371
$2^a - a - 1$	11	26	1,013	8,178	1,048,555

PARIS LODRON UNIVERSITÄT SALZBURG

Finding the interesting endpoints

- Possible **nonparametric** hypotheses using three response variables / endpoints (h, n, d)
  and all factor levels (say, ABCD, as before)

  hnd $\widehat{=}$
  $$F_A^{(h,n,d)} = F_B^{(h,n,d)} = F_C^{(h,n,d)} = F_D^{(h,n,d)}$$

  hn hd nd $\widehat{=}$
  $$F_A^{(h,n)} = F_B^{(h,n)} = F_C^{(h,n)} = F_D^{(h,n)}$$
  $$F_A^{(h,d)} = F_B^{(h,d)} = F_C^{(h,d)} = F_D^{(h,d)}$$
  $$F_A^{(n,d)} = F_B^{(n,d)} = F_C^{(n,d)} = F_D^{(n,d)}$$

  h n d $\widehat{=}$
  $$F_A^{(h)} = F_B^{(h)} = F_C^{(h)} = F_D^{(h)}$$
  $$F_A^{(n)} = F_B^{(n)} = F_C^{(n)} = F_D^{(n)}$$
  $$F_A^{(d)} = F_B^{(d)} = F_C^{(d)} = F_D^{(d)}$$

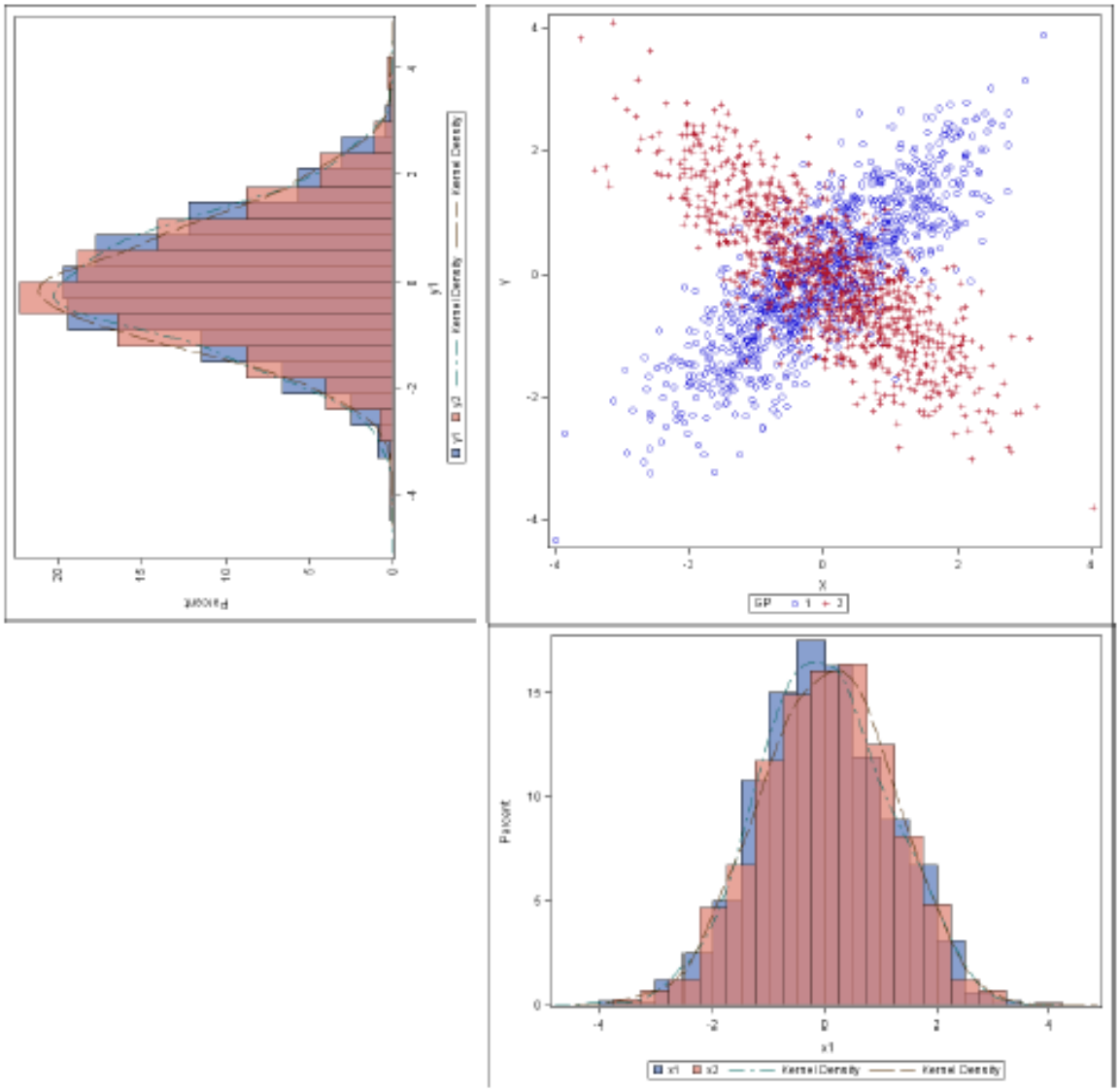
PARIS LODRON UNIVERSITÄT SALZBURG

Finding the interesting endpoints II

- This family of nonparametric hypotheses is...
  ...*not* closed under intersections!
- Consider, e.g., h, n and hn.
- Intersection of the two hypotheses $F_A^{(h)} = F_B^{(h)}$ and $F_A^{(n)} = F_B^{(n)}$ is not equal to $F_A^{(h,n)} = F_B^{(h,n)}$!
- Equality of the marginal distribution does not imply equality of the joint distribution!

PARIS LODRON UNIVERSITÄT SALZBURG

Joint and marginal distributions



Finding the interesting endpoints III

- This family of nonparametric hypotheses is not closed under intersections!
- Consider, e.g., h, n and hn.
- Intersection of the two hypotheses $F_A^{(h)} = F_B^{(h)}$ and $F_A^{(n)} = F_B^{(n)}$ is not equal to $F_A^{(h,n)} = F_B^{(h,n)}$!
- Equality of the marginal distribution does not imply equality of the joint distribution!
  $\Longrightarrow$ Closed testing procedure can not be used.
- Partial Bonferronization: E.g., divide α by the number of hypotheses being considered at each step, to obtain a local α.
- Familywise error rate is not exceeded then.
  However: Can be conservative in some situations
  (depends on the specific test statistic!)

PARIS LODRON UNIVERSITÄT SALZBURG

From nonparametric to semiparametric

- For semi/parametric hypotheses, formulated in terms of location parameters, the situation is different.
- Consider, e.g., h, n and hn.
- Intersection of the two hypotheses $\mu_A^{(h)} = \mu_B^{(h)}$ and $\mu_A^{(n)} = \mu_B^{(n)}$ equals $(\mu_A^{(h)}, \mu_A^{(n)}) = (\mu_B^{(h)}, \mu_B^{(n)})$!
- Equality of the mean vector elements implies equality of the mean vector.
- This family of semi/parametric hypotheses is closed under intersections! $\Longrightarrow$ Closed testing procedure can be used for a coherent test.
- Individual α does not need to be adjusted, familywise error still controlled.

PARIS LODRON UNIVERSITÄT SALZBURG

Subset testing for the flower-insect example

- Which traits or combinations of traits could be discriminating? (height=h, nectar tube depth=n, display size=d)
  hnd $\surd$
  hn hd nd
  h n d
- Which insects could have different trait niches? (Fr, Bp, Av, Bm)
  Fr Bp Av Bm $\surd$
  Fr Bp Av Fr Bp Bm Fr Av Bm Bp Av Bm
  Fr Bp Fr Av Fr Bm Bp Av Bp Bm Av Bm
- Answered with a closed multiple testing procedure, using the (new) nonparametric multivariate method at each step
- Coherence!

PARIS LODRON UNIVERSITÄT SALZBURG

Closed multiple testing procedure

ssnonpartest(height | nec_depth | display $\sim$ animal, flower,
alpha=.05, test=c(1,0,0,0),factors.and.variables=TRUE)

- "test": choose one of ANOVA type, LH, BNP, Wilks' Lambda
- $p \leq a$: test subsets of the variables,
  $p > a$: test subsets of factor levels
  call "factors.and.variables=TRUE" above requests both

PARIS LODRON UNIVERSITÄT SALZBURG

Closed multiple testing procedure: output from R – npmv

The Global Hypothesis is significant, subset algorithm will continue

Performing the Subset Algorithm based on Response Variables

The Hypothesis involving response variables height nec_depth display is significant
The Hypothesis involving response variables nec_depth display is significant
The Hypothesis involving response variables height nec_depth is significant
The Hypothesis involving response variables nec_depth is significant
All appropriate subsets using response variables have been checked using a closed multiple testing procedure, which controls the maximum overall type I error rate at alpha= 0.05

<u>hnd</u>

<u>hn</u> hd <u>nd</u>

h <u>n</u> d

PARIS LODRON UNIVERSITÄT SALZBURG

Closed multiple testing procedure: output from R – npmv

The Global Hypothesis is significant, subset algorithm will continue

Performing the Subset Algorithm based on Factor levels

The Hypothesis between factor levels Anthrenus_verbasci Bombus_pascuorum Bombylius_major Formica_rufibarbis is significant
The Hypothesis between factor levels Bombus_pascuorum Bombylius_major Formica_rufibarbis is significant
The Hypothesis between factor levels Anthrenus_verbasci Bombus_pascuorum Formica_rufibarbis is significant
The Hypothesis between factor levels Anthrenus_verbasci Bombus_pascuorum Bombylius_major is significant
The Hypothesis between factor levels Bombus_pascuorum Formica_rufibarbis is significant
The Hypothesis between factor levels Anthrenus_verbasci Bombus_pascuorum is significant
All appropriate subsets using factor levels have been checked using a closed multiple testing procedure, which controls the maximum overall type I error rate at alpha= 0.05

Fr Bp Av Bm

Fr Bp Av Fr Bp Bm Fr Av Bm Bp Av Bm

Fr Bp Fr Av Fr Bm Bp Av Bp Bm Av Bm

PARIS LODRON UNIVERSITÄT SALZBURG

Conclusions I

- *Inference for data with multiple endpoints*
- Theory and R packages for:
  (N) Nonparametric model, rank-based test and
  (S) Semiparametric model, parametric bootstrap-based test.
- (N) and (S) answer different questions, work for different types of data, have different invariance and robustness properties, have different, although overlapping scope.
- Multivariate more powerful than univariate: Even small effects can be found if they are emitted by the same conditions and are visible *in a few* endpoints – signals don't even have to go into the same direction in the different endpoints.
- *Supplementing multivariate by multiple inference*
- Not exceeding familywise nominal α, that is, few false positives.
- (N) somewhat more conservative than (S) when using closed testing principle

PARIS LODRON UNIVERSITÄT SALZBURG

Non-/Semi-/Parametric MANOVA

- Parametric
  - Multivariate normal (metric) data, equal covariance matrices
  - Hypotheses in mean vectors
- Semiparametric
  - Metric endpoints
  - Hypotheses in mean vectors
  - Bootstrap
- Nonparametric
  - Ordinal, binary, or metric endpoints (or mix thereof)
  - Hypotheses in multivariate distributions
  - Tests based on nonparametric relative effects
  - Invariance under endpoint-wise monotone transformations
- Still lots of work to do!

PARIS LODRON UNIVERSITÄT SALZBURG