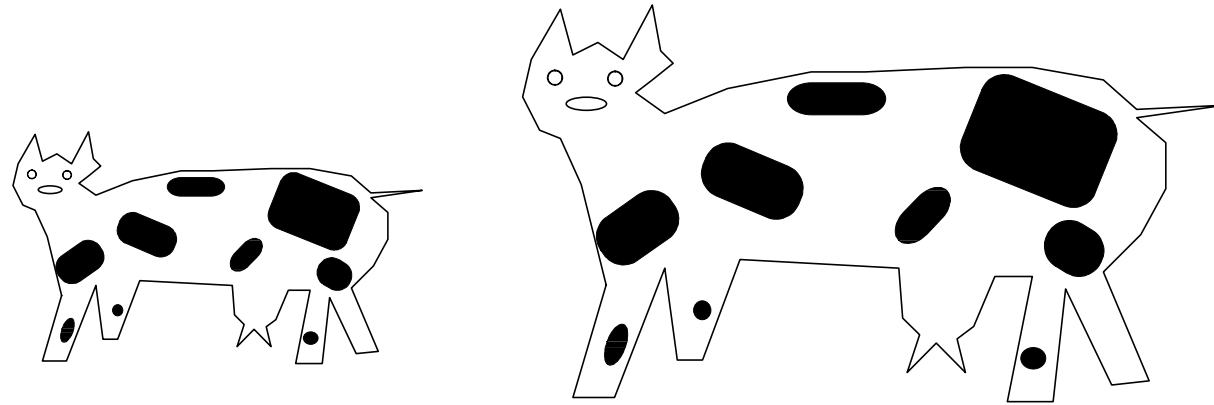


**Bootstrap confidence intervals —  
ein **statistisches** Werkzeug für  
landwirtschaftliche und biologische Versuche**

Manfred Mudelsee

Institute of Meteorology, University of Leipzig, FRG



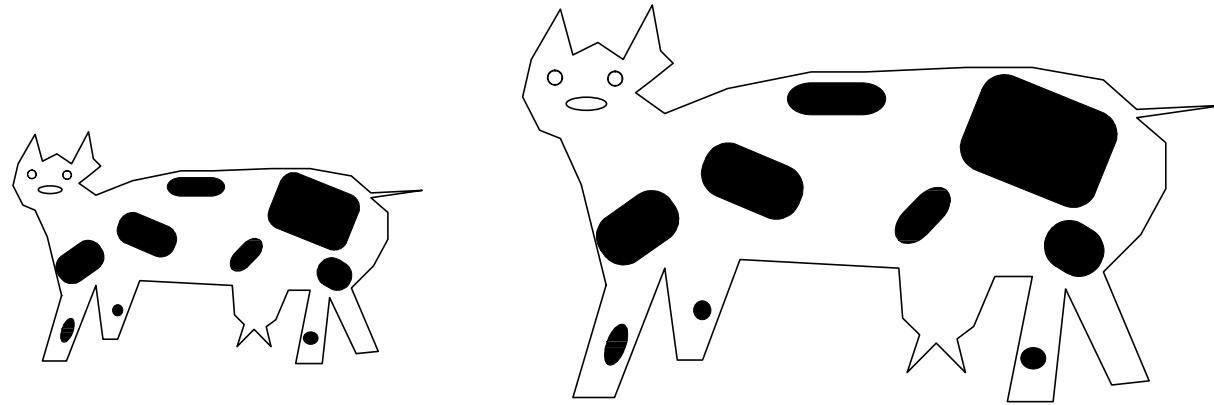
## Exact case

Hypothesis: “Cow on the right (with SUPERFEED™) is heavier.”

vs.

Estimate: “Cow on the right (with SUPERFEED™) is heavier  
by 1.5 tons.”

⇒ surplus information



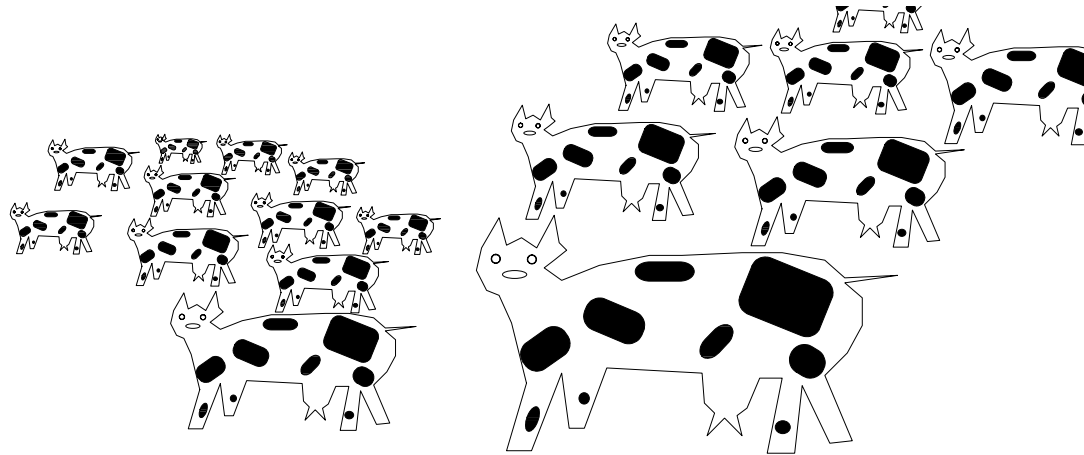
## Uncertain case: **statistics**

Hypothesis: “Cow on the right (with SUPERFEED™) is heavier  
(at significance level 0.9).”

vs.

Estimate: “Cow on the right (with SUPERFEED™) is heavier  
by 1.5 tons (confidence interval: 1.4–1.7 tons).”

⇒ **surplus information (test: zero included ?)**

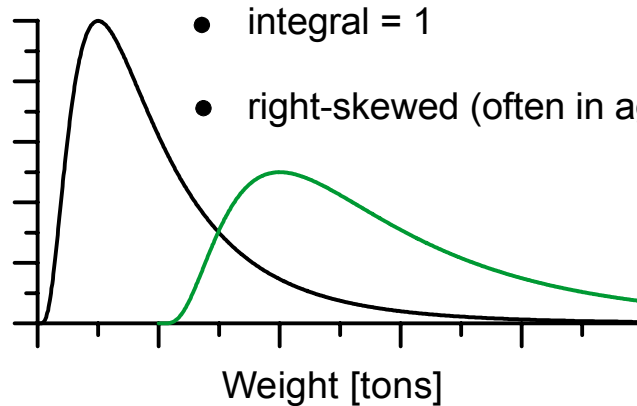


## Statistics

“Cows with SUPERFEED™ have more weight than ordinary cows.”

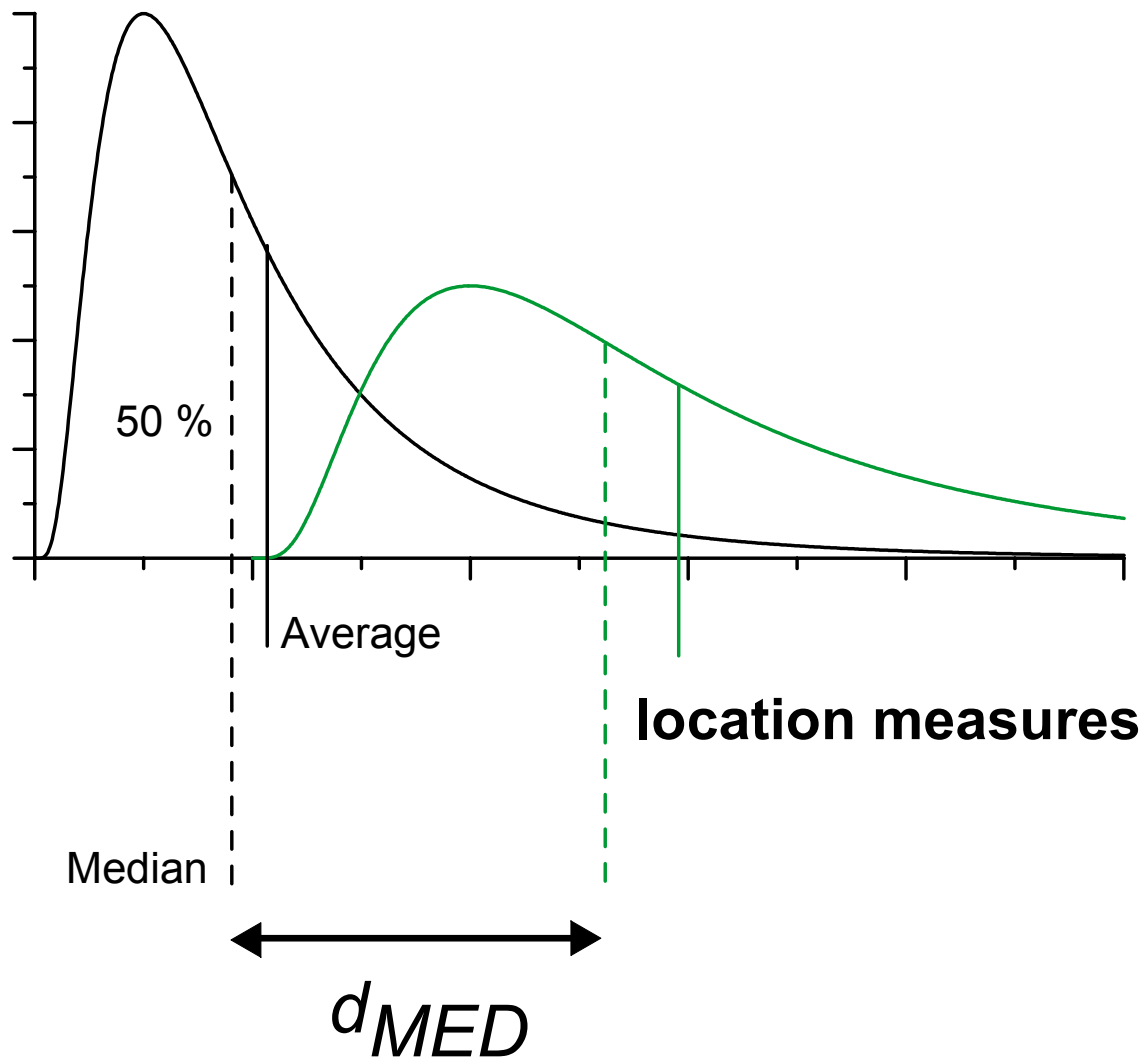
Probability density function  
(population)

with SUPERFEED™  
without SUPERFEED™



Probability density function  
(population)

with SUPERFEED™  
without SUPERFEED™



## Estimation of population parameters with samples

$n_1$  cows with SUPERFEED<sup>TM</sup>, weights  $x_1(i), i = 1, \dots, n_1$   
 $n_2$  cows without SUPERFEED<sup>TM</sup>, weights  $x_2(i), i = 1, \dots, n_2$

$$\begin{aligned} \widehat{d}_{MED} &= \text{sample median}\{x_1(1), \dots, x_1(n_1)\} - \\ &\quad \text{sample median}\{x_2(1), \dots, x_2(n_2)\} \\ &= 1.5 \text{ tons} \end{aligned}$$

“hat” notation

robust estimation:

only slightly affected by departures from assumptions

robust estimation:

only slightly affected by departures from assumptions

example normal assumption:

median = robust, mean = non-robust (large estimation error)

normal distribution:

Mean income in Giebichenstein = 700 EUR

median income = 700 EUR

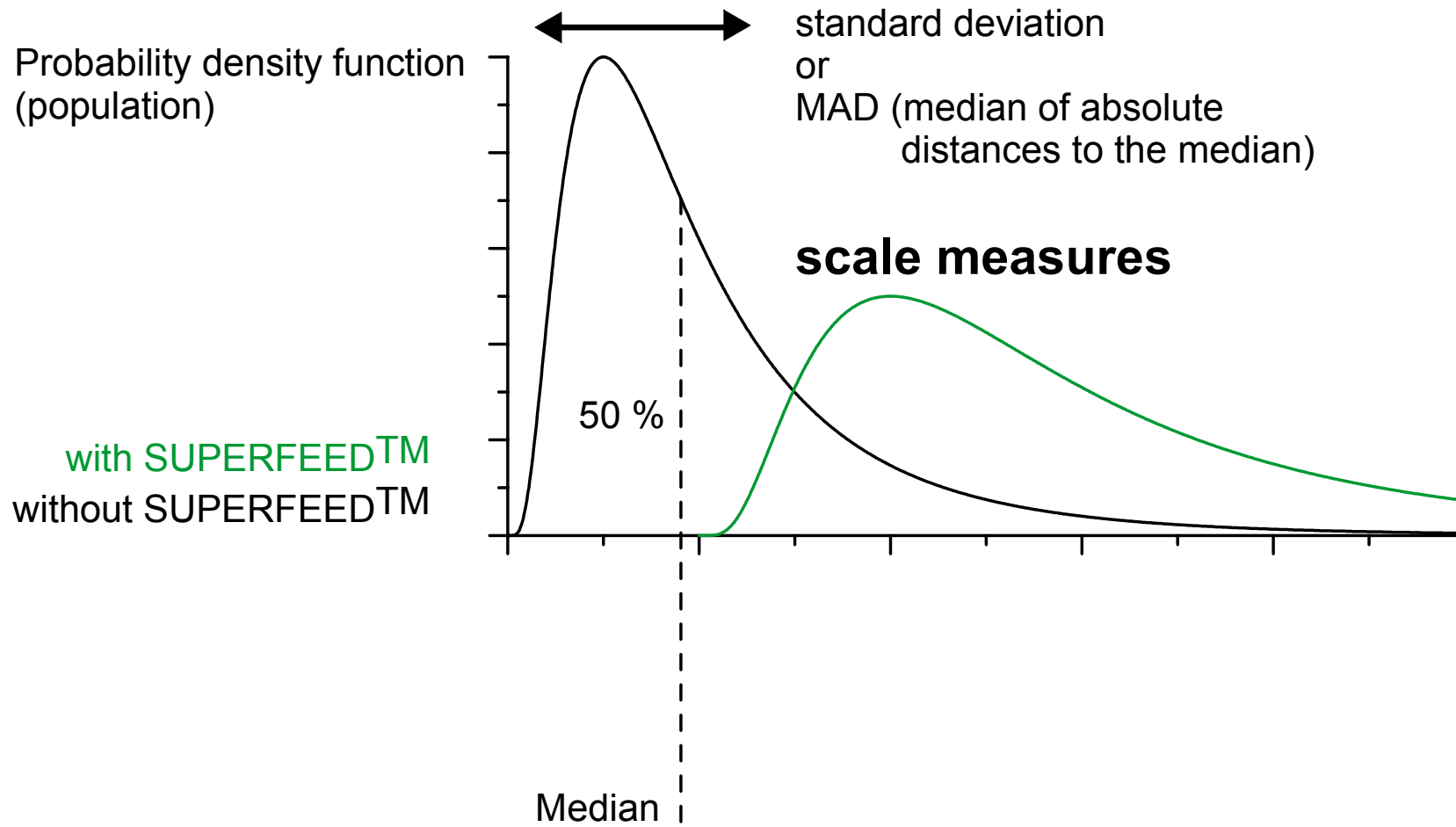
Michael Schumacher moves to Giebichenstein

⇒ right-skewed distribution

Mean income in Giebichenstein = 20000 EUR

median income = 700 EUR

Sample estimates—of mean but not of median—depend critically on whether or not Michael Schumacher is included.





	Treatment	Control	$\nu$
Data	$x1$	$x2$	
<b>Unpaired</b>			
Size	$n1$	$n2$	
$\widehat{d}_{AVE}$	$= AVE(x1)$	$- AVE(x2)$	$n1 + n2 - 2$
$\widehat{d}_{MED}$	$= MED(x1)$	$- MED(x2)$	$n1 + n2 - 2$
$\widehat{d}_{STD}$	$= STD(x1)$	$- STD(x2)$	$n1 + n2 - 4$
$\widehat{d}_{MAD}$	$= MAD'(x1)$	$- MAD'(x2)$	$n1 + n2 - 4$

$AVE(x1) = \sum_{i=1}^{n1} x1(i)/n1 =$  sample mean,

$STD(x1) = \sqrt{\sum_{i=1}^{n1} \{x1(i) - AVE(x1)\}^2 / (n1 - 1)}$  = sample std,

$MED(x1)$  = sample median,

$MAD'(x1) = 1.4826MAD(x1)$ ,

$MAD(x1) = \text{median}\{|x1(i) - MED(x1)|\}$  = sample  $MAD$

(normalizing: normal distribution has std =  $MAD'$ );

analogously for  $x2$ ;

$\nu$  = degrees of freedom

Unpaired experiment: independent samples

**example**: natural cows

Paired experiment: dependent samples, often: pairs

**example**: “cloned” cows (2 copies)

$x_1(1)$  = weight of 1st copy of cow 1  
(with SUPERFEED™)

$x_2(1)$  = weight of 2nd copy of cow 1  
(without SUPERFEED™)

$x_3(i) = x_1(i) - x_2(i)$

$\widehat{MED}_d = \text{sample median}\{x_3(i)\}$

advantage: inter-cow variability has no effect

	Treatment	Control	$\nu$
Data	$x1$	$x2$	
<b>Paired</b>			
Size	$n1$	$=$	$n2$
$\widehat{AVE}_d$	$=$	$AVE(x1 - x2)$	$n1 - 1$
$\widehat{MED}_d$	$=$	$MED(x1 - x2)$	$n1 - 1$

$$AVE(x1 - x2) = \sum_{i=1}^{n1} \{x1(i) - x2(i)\} / n1,$$

$$MED(x1 - x2) = \text{median of } \{x3(i) = x1(i) - x2(i)\}, i = 1, \dots, n1$$

Note: measures of scale not of interest in paired experiment

**Our aim: estimate confidence intervals for difference measures**

$\widehat{d}_{AVE}, \widehat{d}_{MED}, \widehat{d}_{STD}, \widehat{d}_{MAD}$  (**unpaired**)

$\widehat{AVE}_d, \widehat{MED}_d$  (**paired**)

Problems: (partly) complicated measures

few data (5, 10)

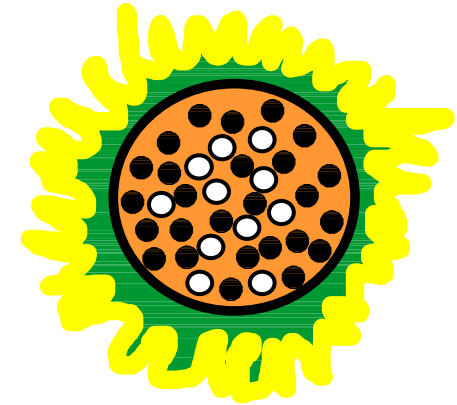
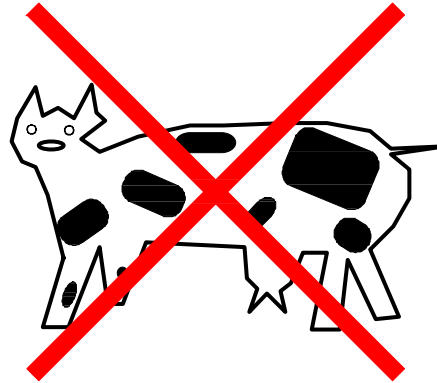
non-normal distributions

⇒ not much theory

Theory

Monte Carlo

Application



Mudelsee M, Alkio M (submitted) Bootstrap Confidence Intervals for Measures of Difference Between Two Samples: An Application to Sunflower Data. Journal of Agricultural, Biological, and Environmental Statistics.

<http://www.uni-leipzig.de/~meteo/MUDELSEE/talk/boottalk.pdf>

## Bootstrap resampling

### example

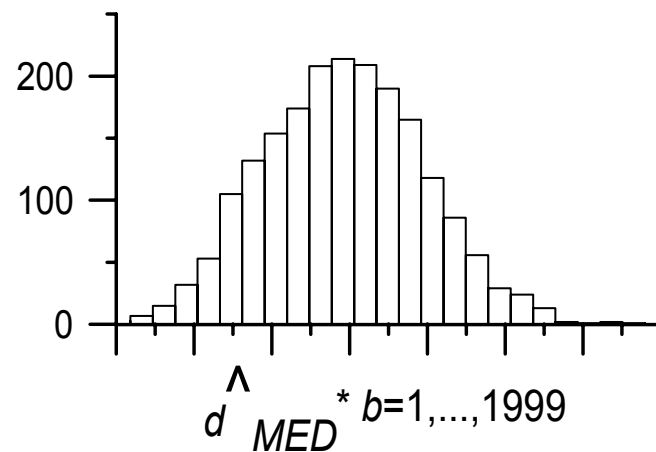
$$\begin{array}{ll}
 n_1 = 4 & \{x_1(1), x_1(2), x_1(3), x_1(4)\} \\
 n_2 = 3 & \{x_2(1), x_2(2), x_2(3)\}
 \end{array}
 \quad d_{MED}^{\widehat{\phantom{x}}}$$

resampling ( $b = 1$ ) with replacement

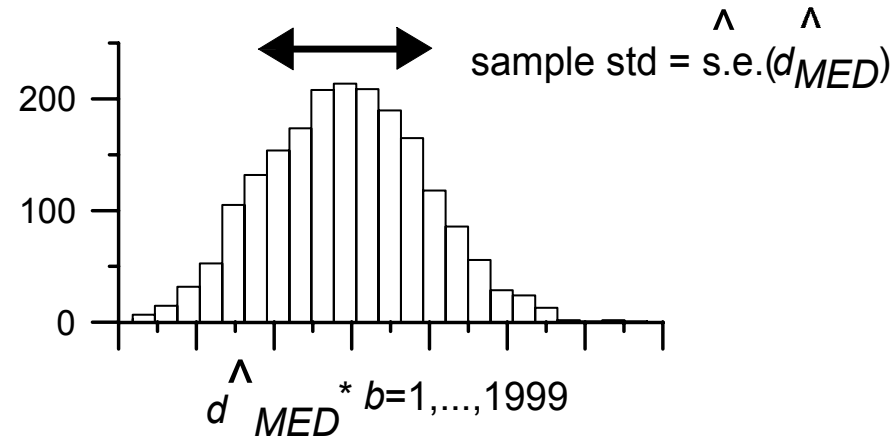
$$\begin{array}{ll}
 n_1^* = 4 & \{x_1^*(1), x_1^*(2), x_1^*(3), x_1^*(4)\} \\
 & \{x_1(3), x_1(3), x_1(4), x_1(1)\} \\
 n_2^* = 3 & \{x_2^*(1), x_2^*(2), x_2^*(3)\} \\
 & \{x_2(2), x_2(3), x_2(3)\}
 \end{array}
 \quad d_{MED}^{\widehat{\phantom{x}}*} \quad b=1$$

Repeat  $B$  times in total (typically:  $B \approx 2000$ )

$\widehat{d}_{MED}^* \quad b=1, \dots, B$  bootstrap replications



Bootstrap idea: distribution of  $\widehat{d}_{MED}^* \approx$  distribution of  $\widehat{d}_{MED}$   
(refinements account for differences)



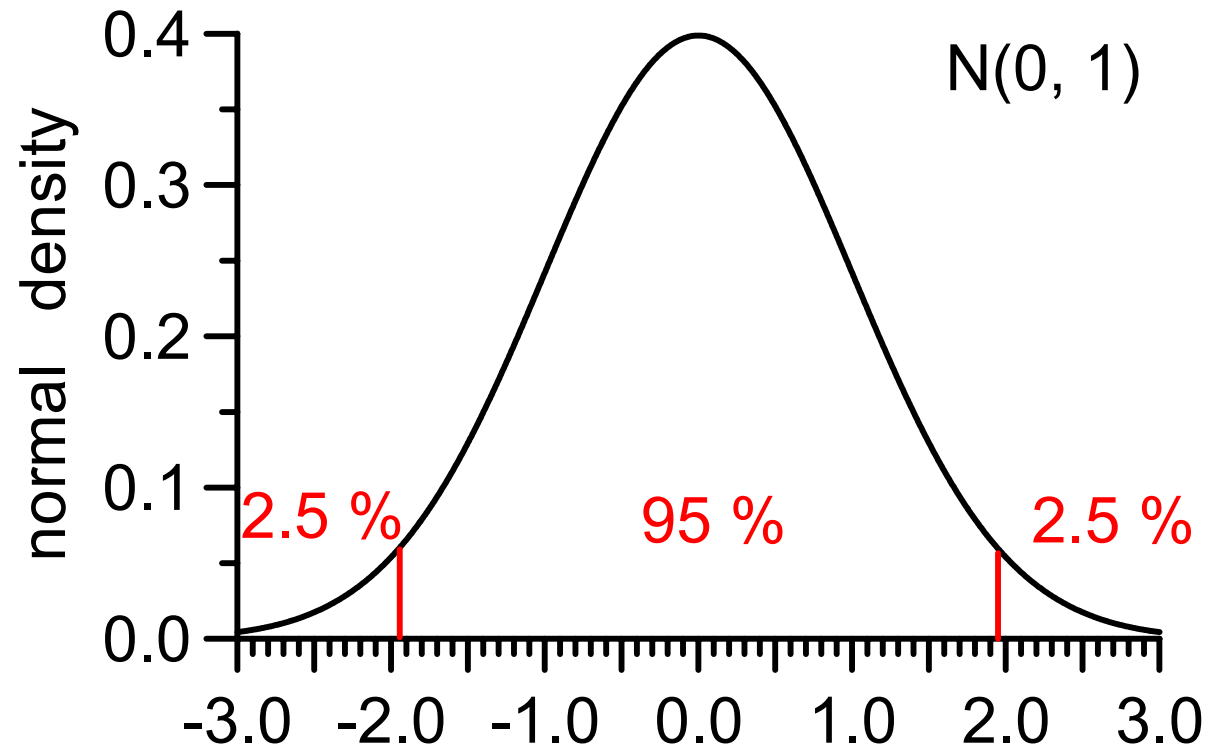
\*\*

## I Normal confidence interval

$$\left[ \hat{d}_{MED} - z^{(1-\alpha)} \cdot \widehat{se} \left( \hat{d}_{MED} \right), \hat{d}_{MED} + z^{(1-\alpha)} \cdot \widehat{se} \left( \hat{d}_{MED} \right) \right]$$

$z^{(1-\alpha)}$ : standard normal percentile (e. g.,  $z^{(0.975)} = 1.96$ ) \*

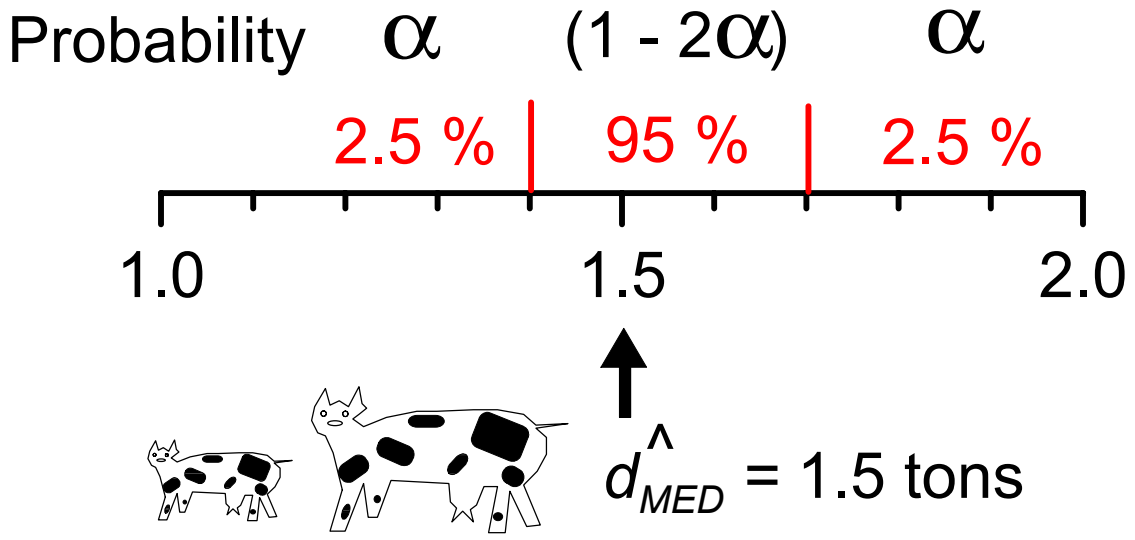




\*\*

Confidence interval:

equi-tailed  $(1 - 2\alpha)$  interval



## II Student's $t$ confidence interval

$$\left[ \widehat{d}_{MED} - t_{\nu}^{(1-\alpha)} \cdot \widehat{se}(\widehat{d}_{MED}), \widehat{d}_{MED} + t_{\nu}^{(1-\alpha)} \cdot \widehat{se}(\widehat{d}_{MED}) \right]$$

$t_{\nu}^{(1-\alpha)}$ : percentile Student's  $t$  distribution ( $\nu$  degrees of freedom)

$\widehat{se}(\widehat{d}_{MED})$ : unknown mean

$\Rightarrow$  wider confidence interval than normal

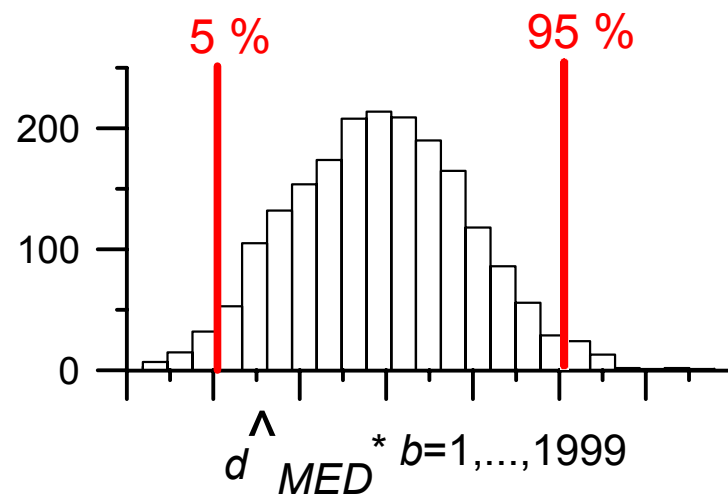
data sizes  $> 30$ : differences negligible

**Coverage error,  $C$** 

Probability  $\left[ d_{MED} \leq \text{lower confidence bound for } \widehat{d}_{MED} \right] = \alpha + C$

terms of  $C \propto 1/\sqrt{n}$  and “faster”: first -order accurate

terms of  $C \propto 1/n$  and “faster”: second-order accurate

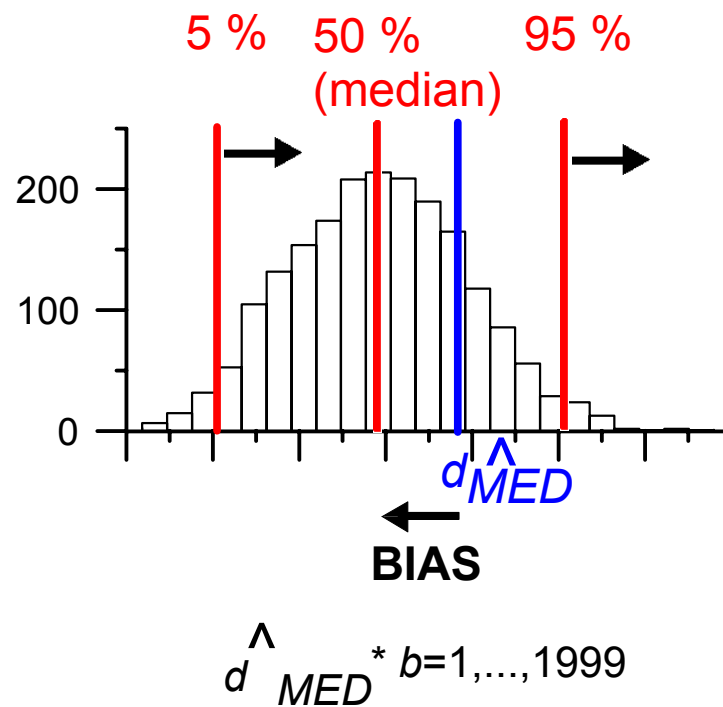


### III Percentile confidence interval

$$\left[ \widehat{d_{MED}}^{* b=1, \dots, B=\infty}(\alpha), \widehat{d_{MED}}^{* b=1, \dots, B=\infty}(1-\alpha) \right]$$

practice ( $B < \infty$ ): approximate interval

**example**:  $B = 1999$ , 95th percentile =  $0.95 \cdot (1999 + 1)$ th  
 = 1900th largest  $\widehat{d_{MED}}^*$  value



## IV BCa confidence interval

bias correction  $\hat{z}_0$

“acceleration”  $\hat{a}$ :  $(\widehat{s.e.}(d_{MED}^{\wedge}))$  needs not to be constant)  
rate of change of  $\widehat{s.e.}(d_{MED}^{\wedge})$  with  $d_{MED}$

BCa interval:  $\left[ \widehat{d}_{MED}^{*b=1, \dots, B=\infty} (\alpha 1), \widehat{d}_{MED}^{*b=1, \dots, B=\infty} (\alpha 2) \right]$

$$\alpha 1 = \Phi \left[ \widehat{z}_0 + \frac{\widehat{z}_0 + z(\alpha)}{1 - \widehat{a} (\widehat{z}_0 + z(\alpha))} \right],$$

$$\alpha 2 = \Phi \left[ \widehat{z}_0 + \frac{\widehat{z}_0 + z(1-\alpha)}{1 - \widehat{a} (\widehat{z}_0 + z(1-\alpha))} \right],$$

$$\widehat{z}_0 = \Phi^{-1} \left( \frac{\text{number of replications where } \widehat{d}_{MED}^{*b} < \widehat{d}_{MED}}{B} \right), \quad [\text{Note: } \Phi^{-1}(1/2) = 0]$$

$$\widehat{a} = \frac{\sum_{i=1}^{n1} \sum_{j=1}^{n2} \left( \langle \widehat{d}_{MED\text{jack}(i, j)} \rangle - \widehat{d}_{MED\text{jack}(i, j)} \right)^3}{6 \left[ \sum_{i=1}^{n1} \sum_{j=1}^{n2} \left( \langle \widehat{d}_{MED\text{jack}(i, j)} \rangle - \widehat{d}_{MED\text{jack}(i, j)} \right)^2 \right]^{3/2}}.$$

# Theory

standard-error based

accuracy (order)

**I Normal confidence interval**

1st [estimator  $\sim$  normal: 2nd]

**II Student's  $t$  confidence interval**

1st [estimator  $\sim$  normal: 2nd]

percentile-based

**III Percentile confidence interval**

1st

**IV BCa confidence interval**

2nd



## Why all that stuff ???

Treatment and control populations: in general non-normal  
⇒ no “normal” theory

small data sizes  
⇒ no asymptotic theory

$d_{MED}$ ,  $d_{MAD}$ ,  $MED_d$ : order statistics  
⇒ analytical derivations difficult/impossible

## Why not use transformations of data to normality ???

- Idea: use of something like  $\sqrt{STD(x1)^2/n1 + STD(x2)^2/n2}$  (normal theory) for calculating confidence intervals
- But: Can be laborious if many data sets are to be analyzed.
- But: This method needs not provide the same information. \*

\*

	$x1$	$x2$	data
	$x1' = g_1(x1)$	$x2' = g_2(x2)$	transformation
	$x1' = \log(x1)$	$x2' = \log(x2)$	<b>example</b>
$d'_{MED}$	$= MED(x1') -$	$MED(x2')$	
	$= MED[\log(x1)] -$	$MED[\log(x2)]$	
	$= \log[MED(x1)] -$	$\log[MED(x2)]$	log = monotonic
	$= \log \left[ \frac{MED(x1)}{MED(x2)} \right]$		

$d'_{MED}$  offers information about **ratio** of original medians (not about difference)

$g_1, g_2$  different and/or more complex:  $d'_{MED}$  difficult to interpret

## Notes

- Further types of bootstrap confidence intervals, “double bootstrap”, bootstrap calibration etc.  $\Rightarrow$  computationally more expensive
- bootstrap: becoming standard method in Statistics, current research: (e. g.) confidence intervals
- bootstrap: seen from time to time in Physics, Geosciences, has been seen in Agriculture

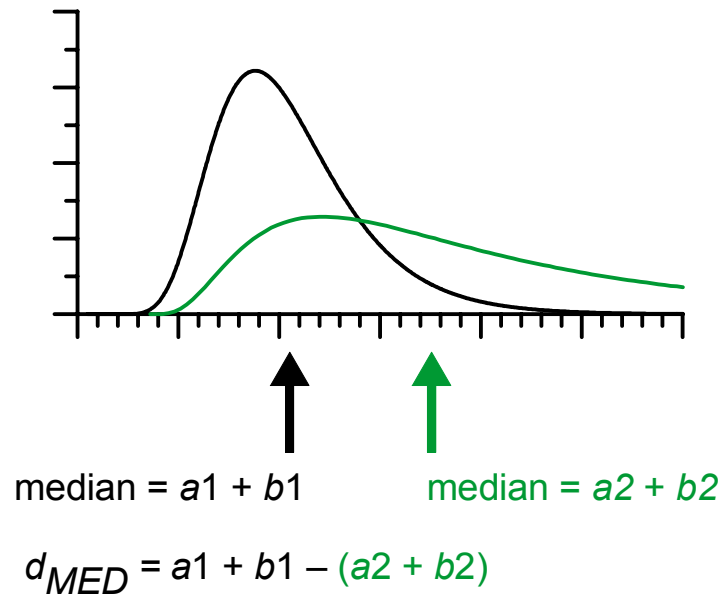
### Reading:

- Efron B (1979) Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7:1–26. [introduces bootstrap]
- Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall, London, pp 436. [accessible textbook]
- Hall P (1988) Theoretical comparison of bootstrap confidence intervals (with discussion). The Annals of Statistics 16:927–985. [calculates coverage errors]

## Lognormal treatment and control

$$f1(x) = \frac{1}{s1(x-a1)} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2s1^2} \left[ \ln \left( \frac{x-a1}{b1} \right) \right]^2 \right] = \text{LN}(a1, b1, s1)$$

$$f2(x) = \text{LN}(a2, b2, s2)$$



Draw random sample  $x_1 \sim \text{LN}(a_1, b_1, s_1)$ , size  $n_1$

Draw random sample  $x_2 \sim \text{LN}(a_2, b_2, s_2)$ , size  $n_2$

Correlation:  $\rho_{\text{LN}}$

Calculate Bootstrap confidence interval  $(1 - 2\alpha)$  for  $\widehat{d}_{\text{MED}}$

Count:  $d_{\text{MED}} <$  lower bound of confidence interval (nominal:  $P = \alpha$ )

$d_{\text{MED}} \in$  confidence interval (nominal:  $P = 1 - 2\alpha$ )

$d_{\text{MED}} >$  upper bound of confidence interval (nominal:  $P = \alpha$ )

Repeat procedure: 10000 Monte Carlo simulations

Compare empirical with nominal coverages

## Monte Carlo

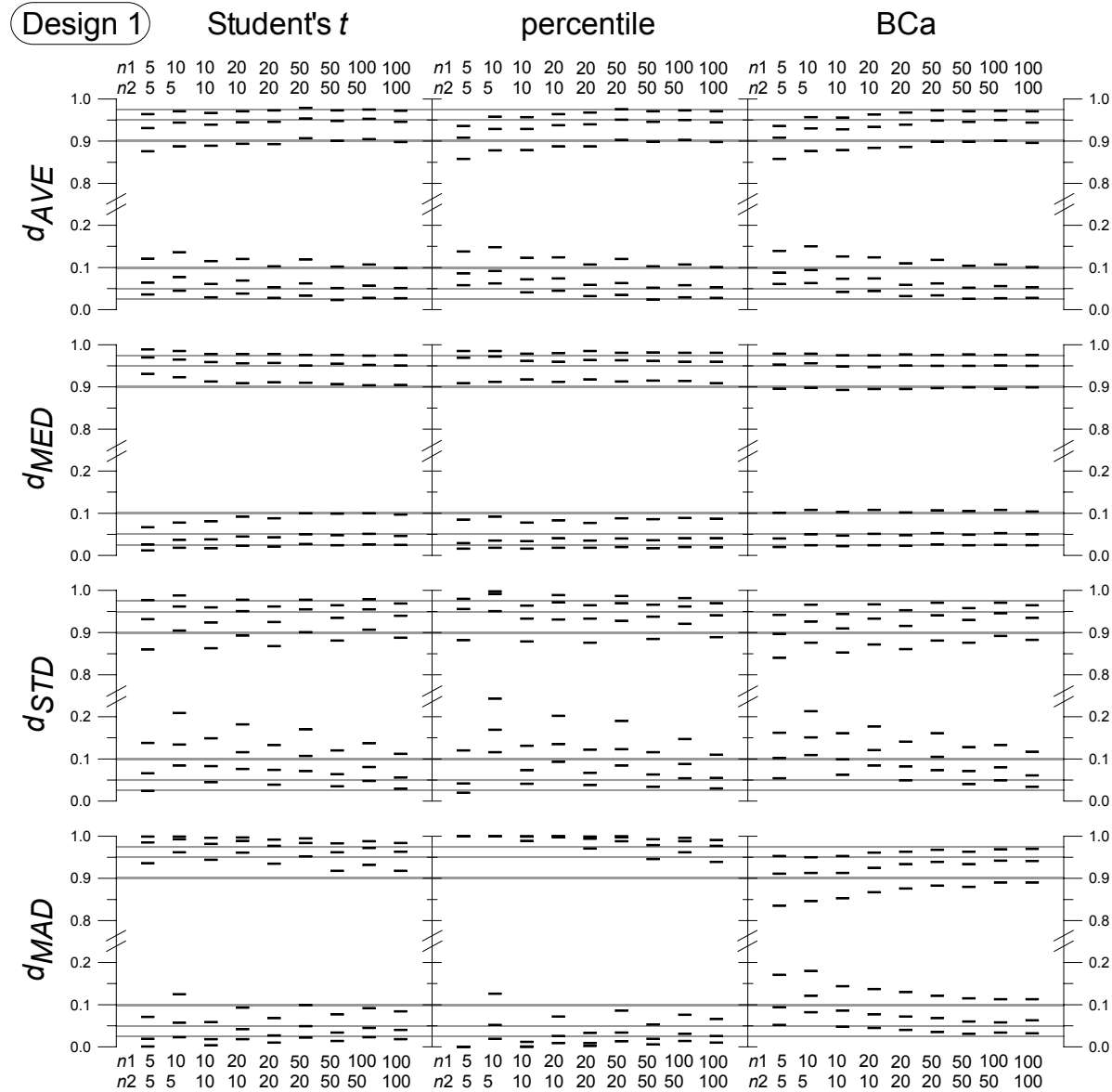
Monte Carlo Designs; Lognormal Parameters and Theoretical Difference Measures

Design	$a1$	$b1$	$s1$	$a2$	$b2$	$s2$	$d_{AVE}$	$d_{MED}$	$d_{STD}$	$d_{MAD}$	$MED_d$
1	0	4	.2	0	4	.2	0	0	0	0	0
2	0	25.3	.2	0	4	.2	21.73	21.30	4.39	4.24	21.24
3	0	4	.8	0	4	.2	1.43	0	4.39	2.17	0
4	0	25.3	.8	0	25.3	.8	0	0	0	0	0

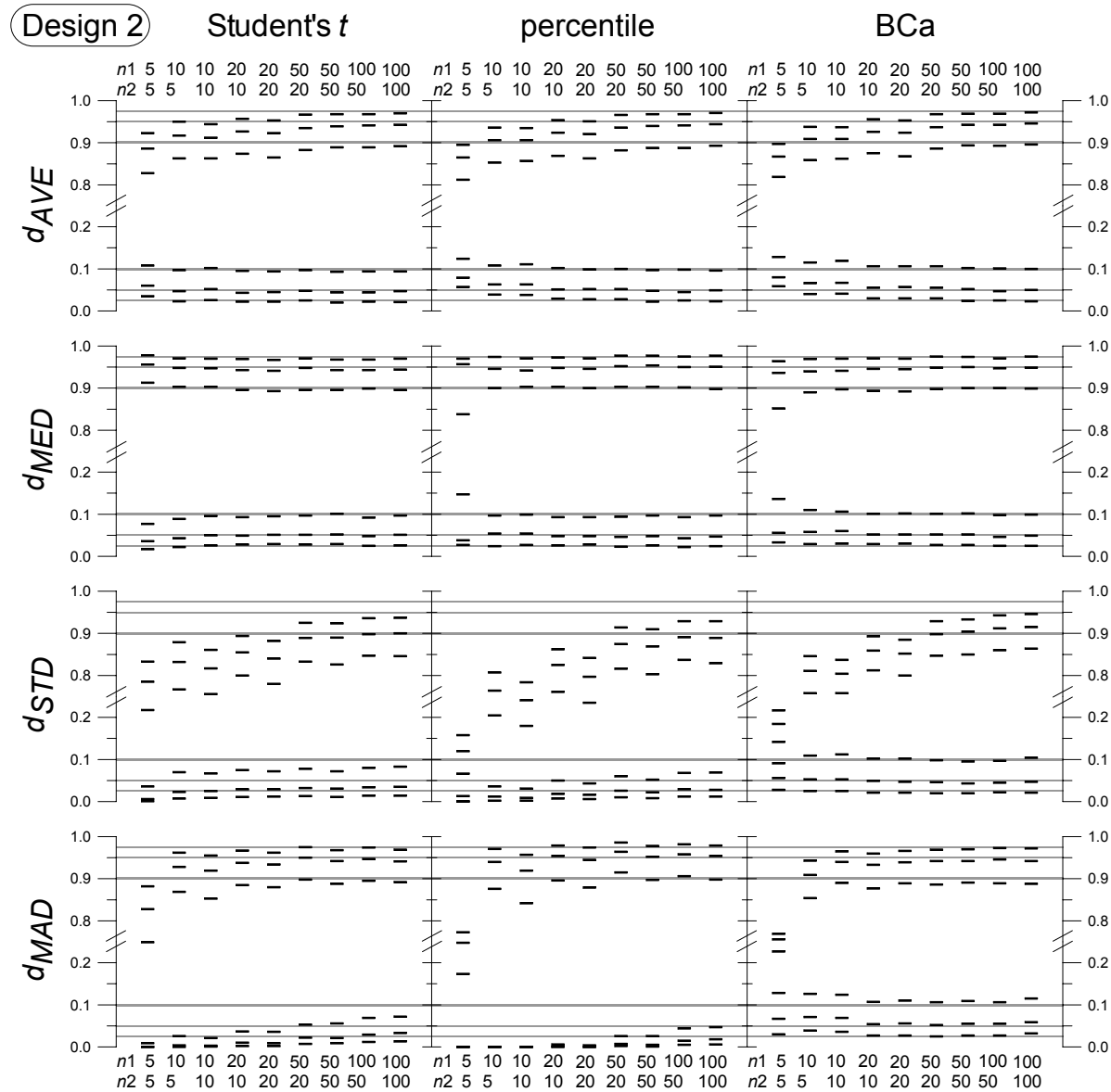
NOTES:  $AVE_d = d_{AVE}$ ,  $MED_d$  determined by preliminary simulation



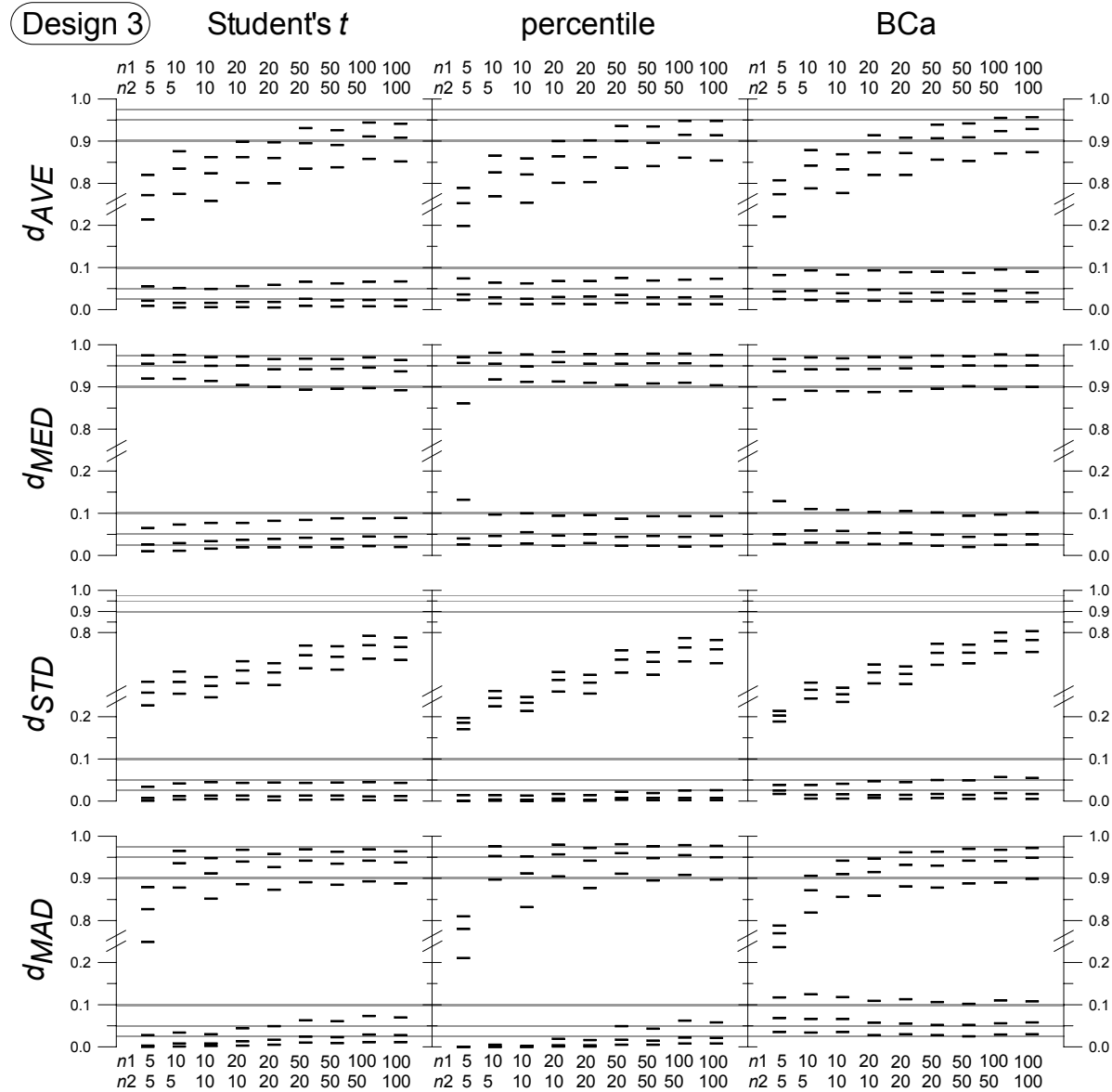
$\rho_{LN} = 0$  (correctly specified unpaired)



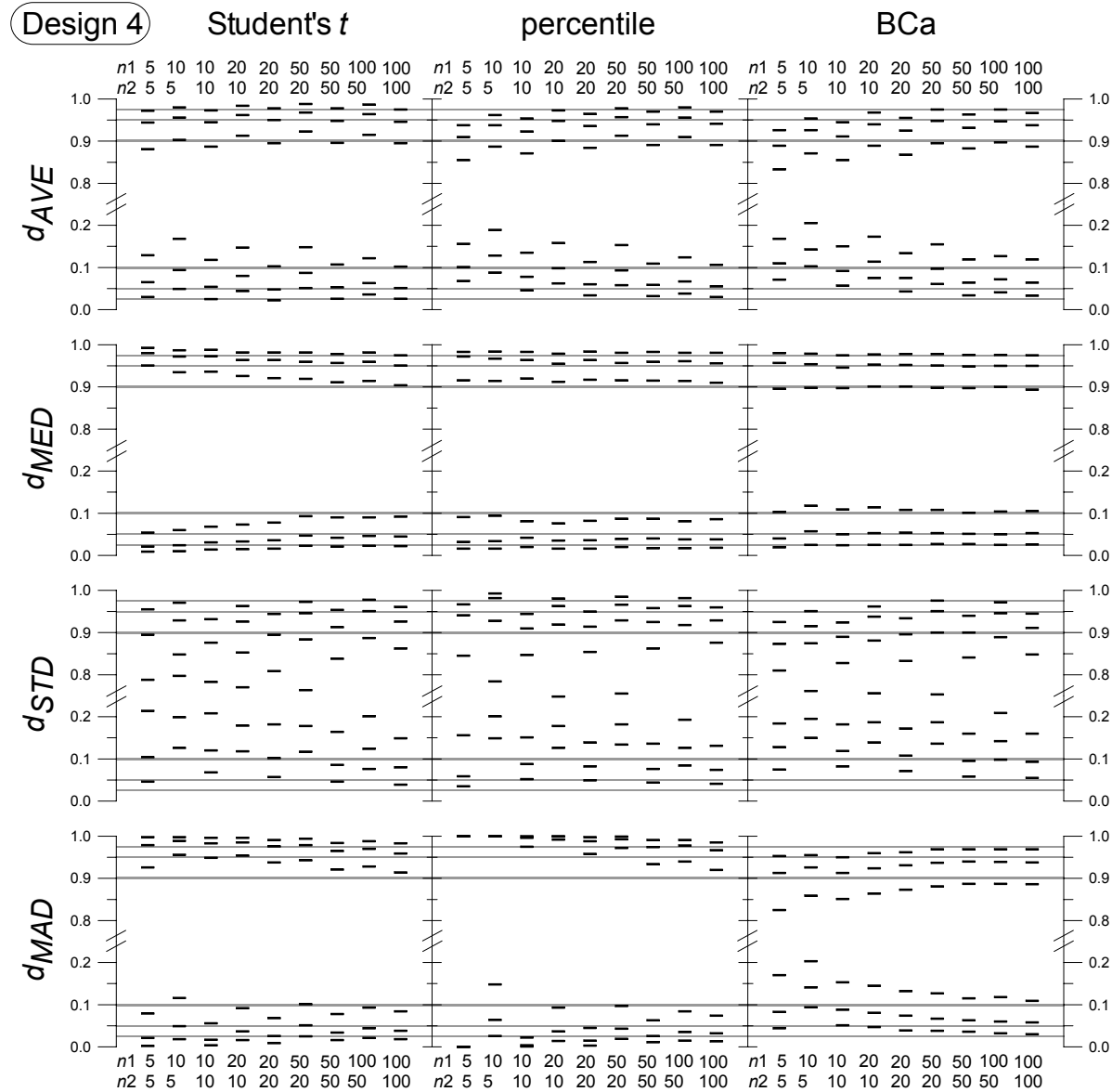
$\rho_{LN} = 0$  (correctly specified unpaired)



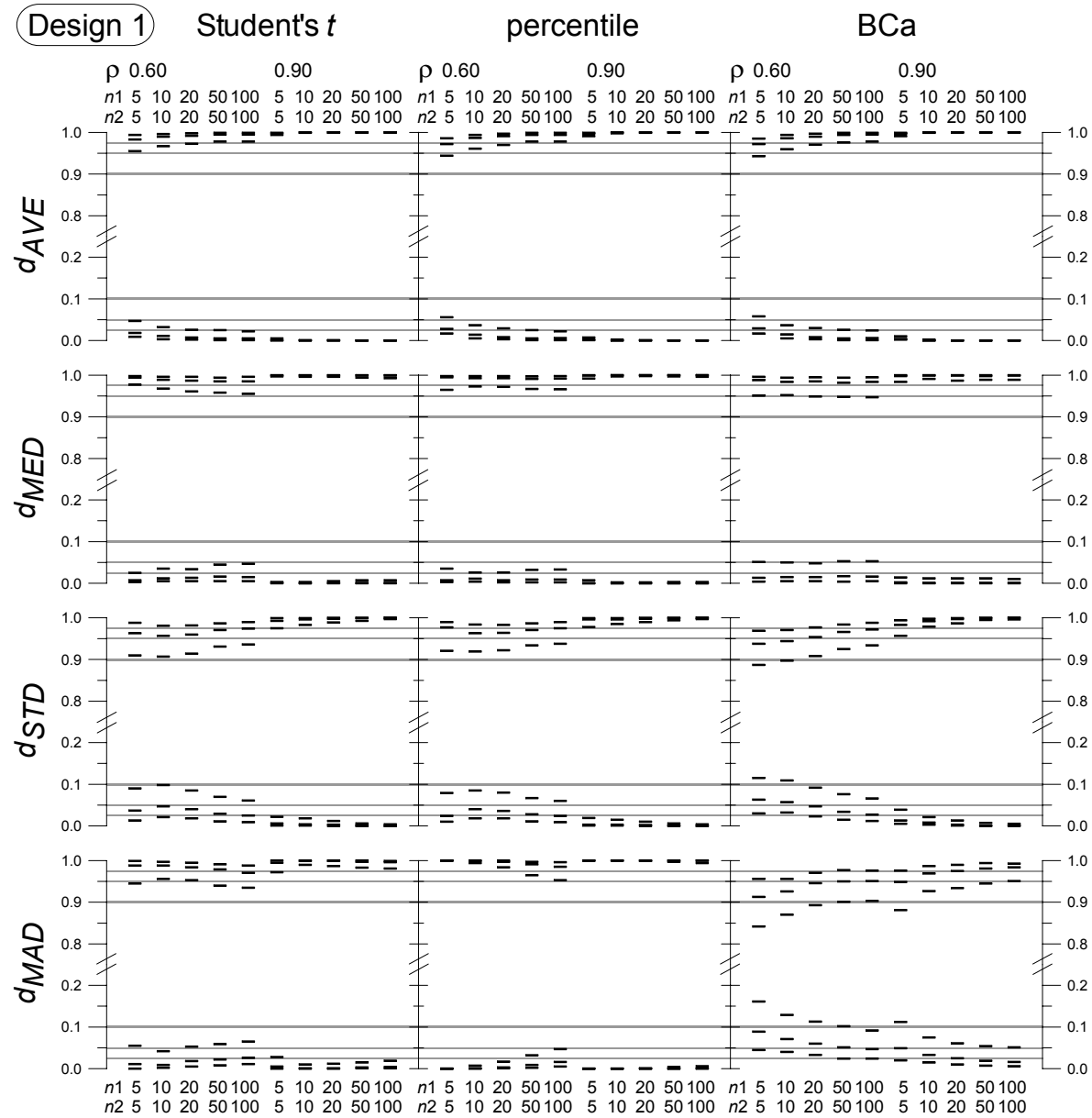
$\rho_{LN} = 0$  (correctly specified unpaired)



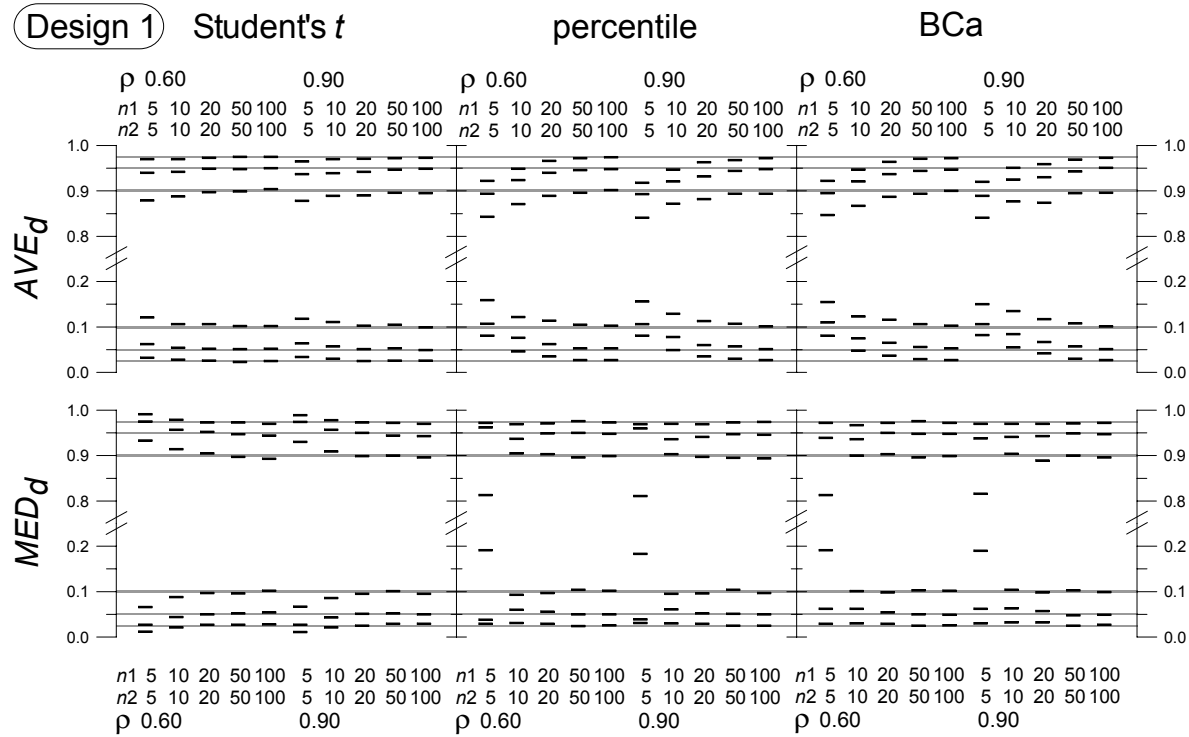
$\rho_{LN} = 0$  (correctly specified unpaired)



$\rho_{LN} \neq 0$  (mis-specified unpaired)



$\rho_{LN} \neq 0$  (correctly specified paired)



## Result

- **BCa** and **Student's  $t$**  confidence intervals of  $\widehat{MED}_d$  as measure of location of difference offer good coverage performance in **paired** experiments for  $n_1 \gtrsim 10$  and  $n_2 \gtrsim 10$ .
- **BCa** confidence intervals of  $\widehat{d}_{MED}$  as measure of difference in location offer good coverage performance in **unpaired** experiments for  $n_1 \gtrsim 20$  and  $n_2 \gtrsim 20$ .
- **BCa** confidence intervals of  $\widehat{d}_{MAD}$  as measure of difference in scale offer acceptably coverage performance in **unpaired** experiments for  $n_1 \gtrsim 50$  and  $n_2 \gtrsim 50$ .

# Application

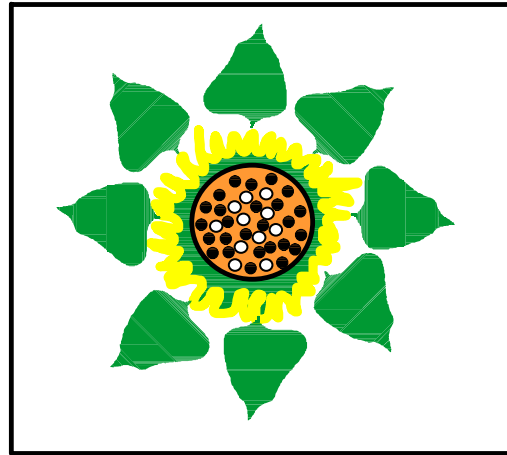
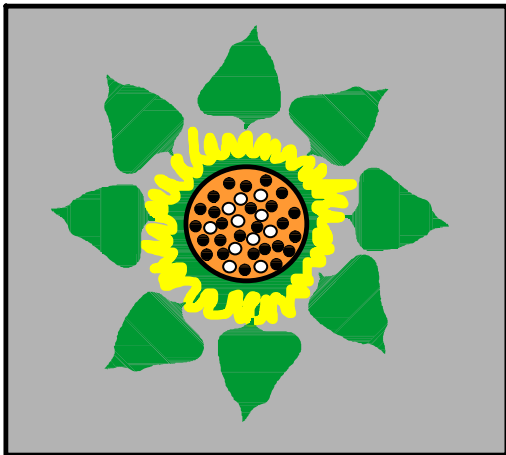
*Helianthus annuus* L. — source–sink ratio manipulations

$x$  = percentage of unfilled achenes

shading

ratio  $\uparrow$

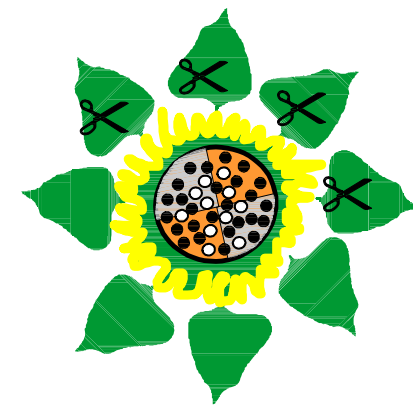
unpaired experiment



defoliation

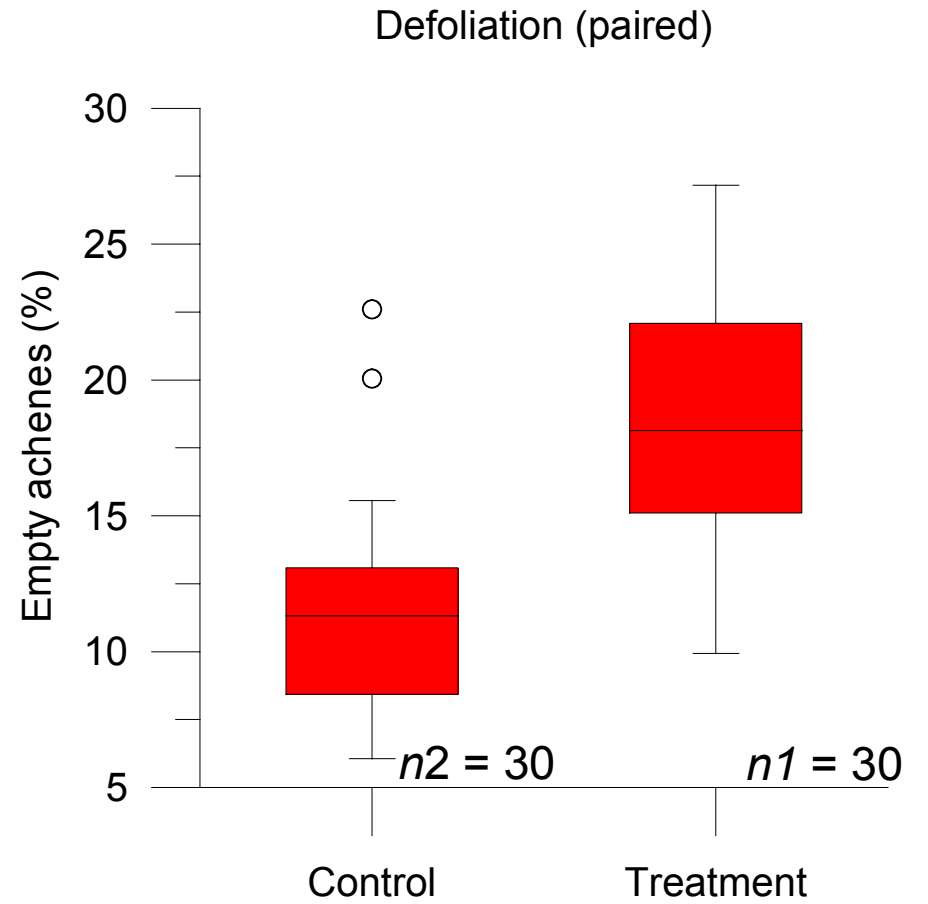
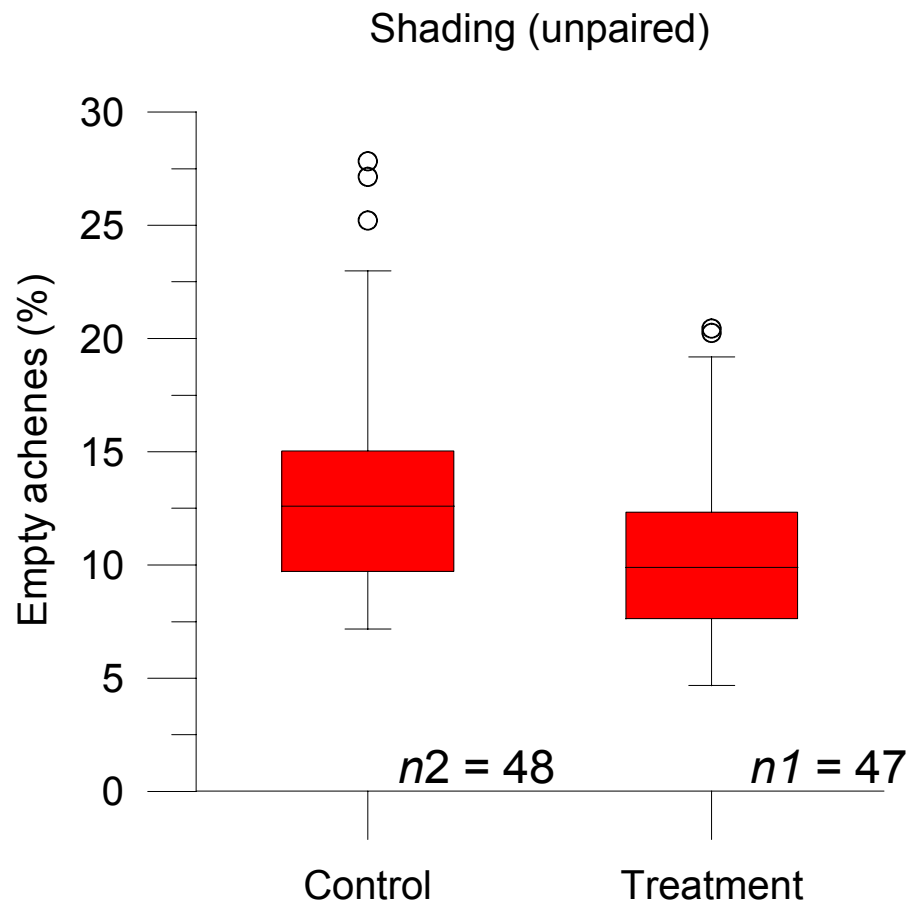
ratio  $\downarrow$

paired experiment

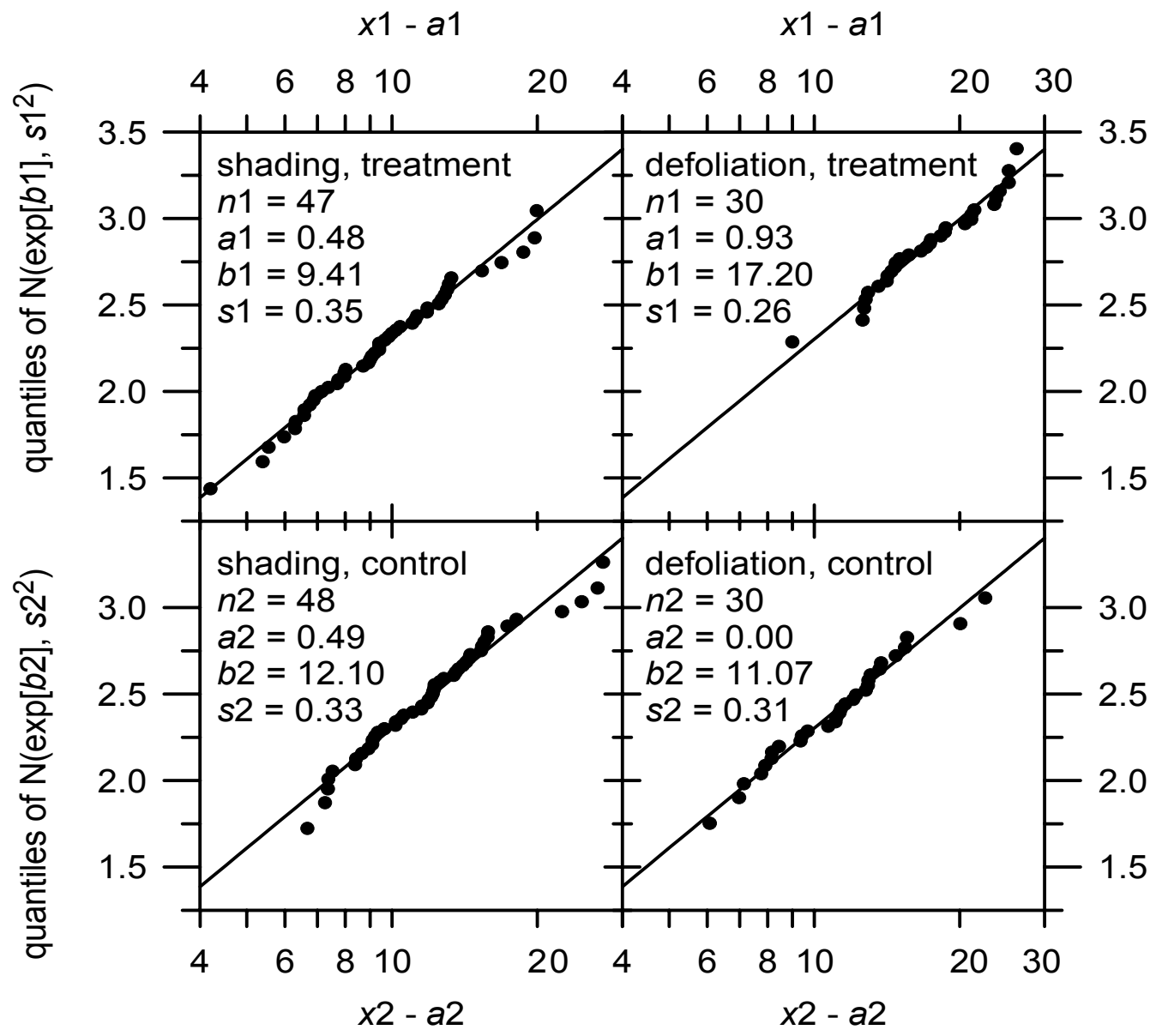




# Application



# Application



# 2SAMPLES demonstration

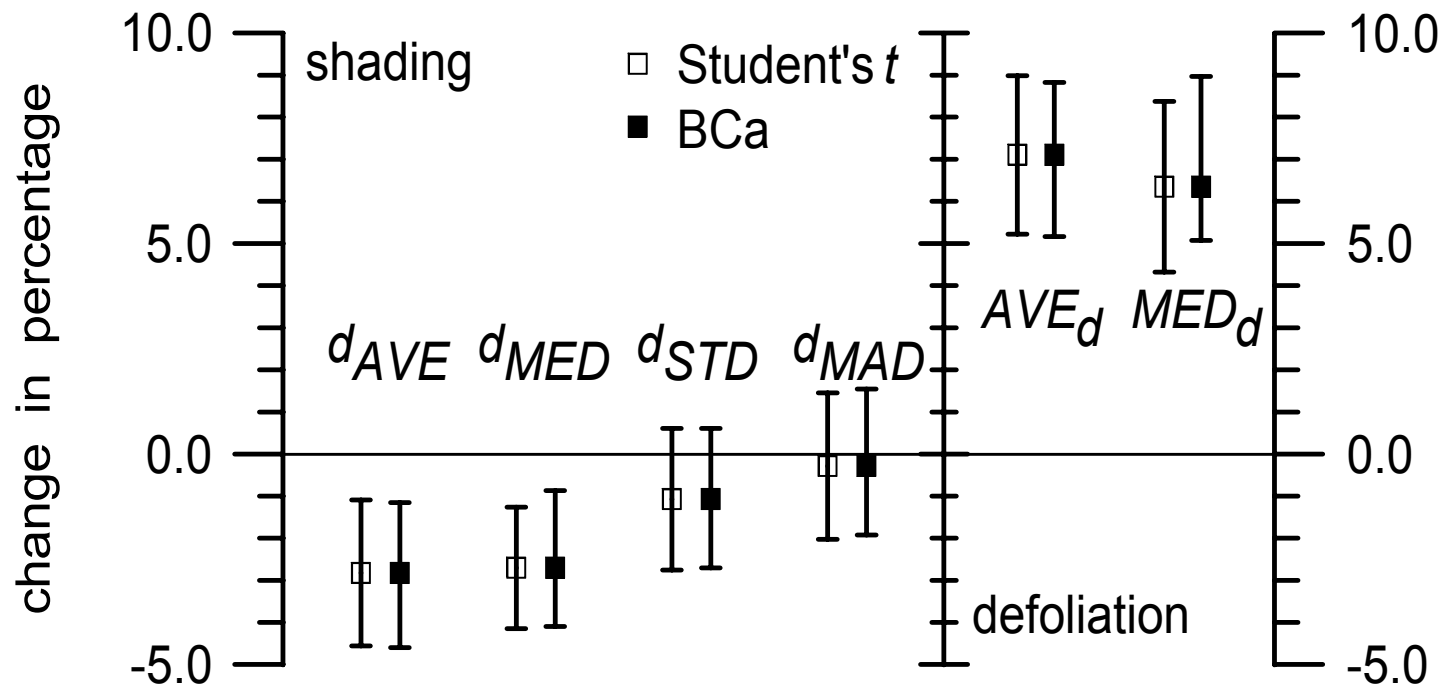
f:\2samples\data\u\_p\_\_123.t

f:\2samples\data\u\_p\_\_123.k

f:\2samples\data\p\_p\_\_1\_6.t

f:\2samples\data\p\_p\_\_1\_6.k

## Final result



## Bootstrap confidence intervals

- works for relatively small data sizes
- works for non-normal distributions
- works also for complicated measures
- avoids transformations
- **surplus information (test: zero included ?)**