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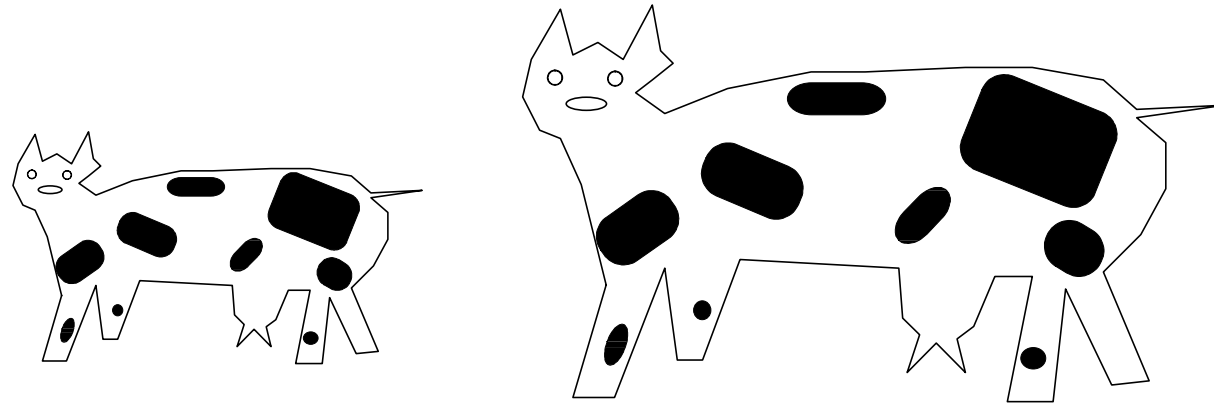 Bootstrap confidence intervals — ein statistisches Werkzeug für landwirtschaftliche und biologische Versuche

Manfred Mudelsee
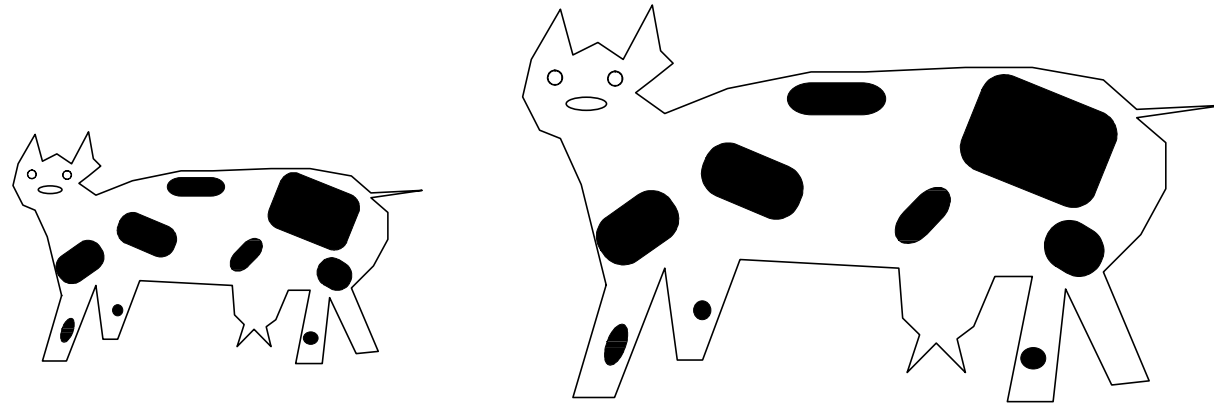
Institute of Meteorology, University of Leipzig, FRG

1

Exact case

Hypothesis: "Cow on the right (with SUPERFEED$^{\text{TM}}$) is heavier."

vs.

Estimate: "Cow on the right (with SUPERFEED$^{\text{TM}}$) is heavier by 1.5 tons."

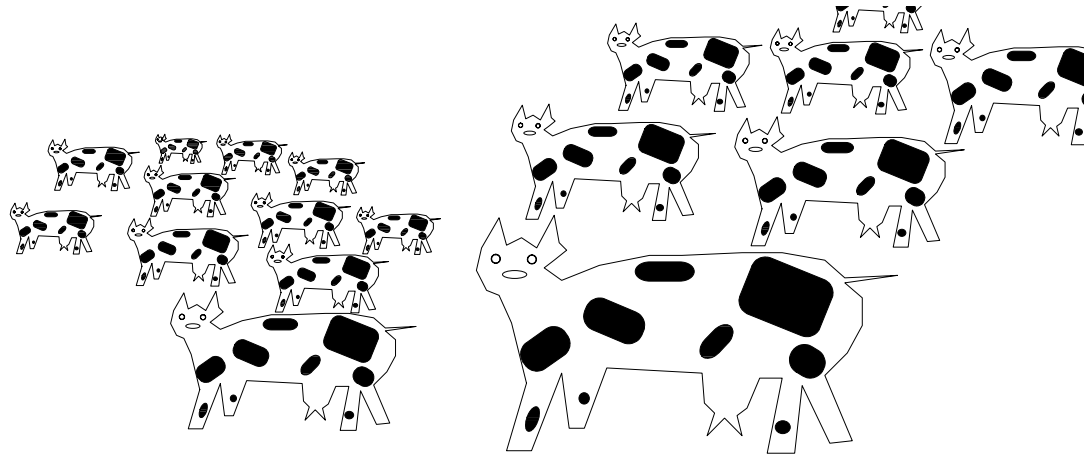$\Rightarrow$ **surplus information**

Uncertain case: statistics

Hypothesis: "Cow on the right (with SUPERFEED$^{\text{TM}}$) is heavier (at significance level 0.9)."

vs.

Estimate: "Cow on the right (with SUPERFEED$^{\text{TM}}$) is heavier by 1.5 tons (confidence interval: 1.4–1.7 tons)."
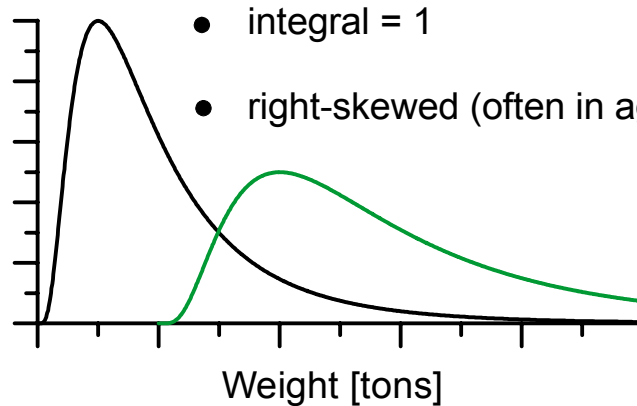
⇒ **surplus information (test: zero included ?)**

Statistics

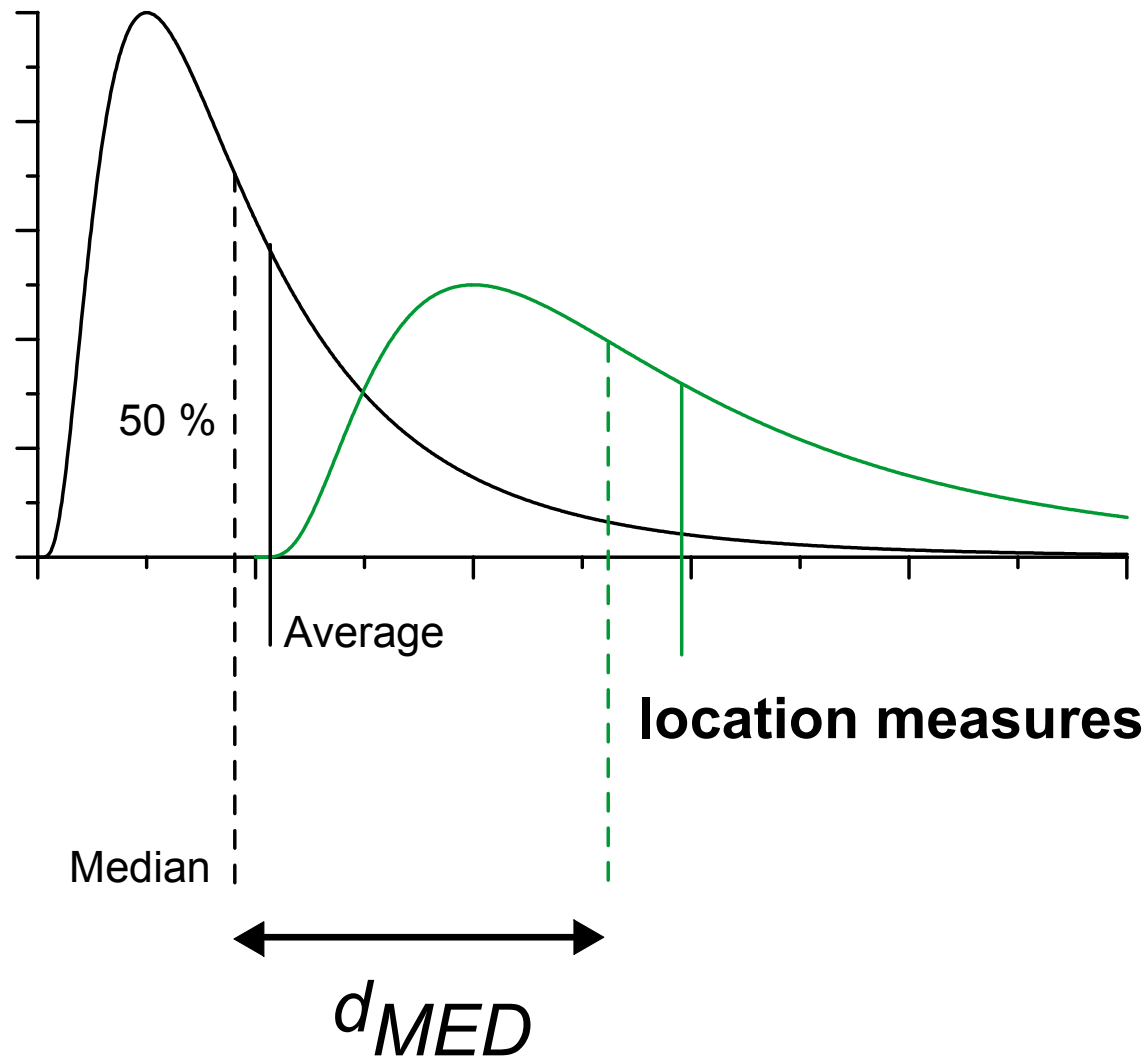"Cows with SUPERFEED^{TM} have more weight than ordinary cows."

Probability density function (population)

with SUPERFEED^{TM}
without SUPERFEED^{TM}

- integral = 1
- right-skewed (often in agronomy or biology)

Weight [tons]

Estimation of population parameters with samples

$n1$ cows with   SUPERFEED$^{\mathsf{TM}}$, weights $x1(i), i = 1, \ldots, n1$

$n2$ cows without SUPERFEED$^{\mathsf{TM}}$, weights $x2(i), i = 1, \ldots, n2$

$$
\begin{aligned}
\widehat{d_{MED}} \;=\; & \text{sample median}\{x1(1), \ldots, x1(n1)\} - \\
& \text{sample median}\{x2(1), \ldots, x2(n2)\} \\
=\; & 1.5 \text{ tons}
\end{aligned}
$$

"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
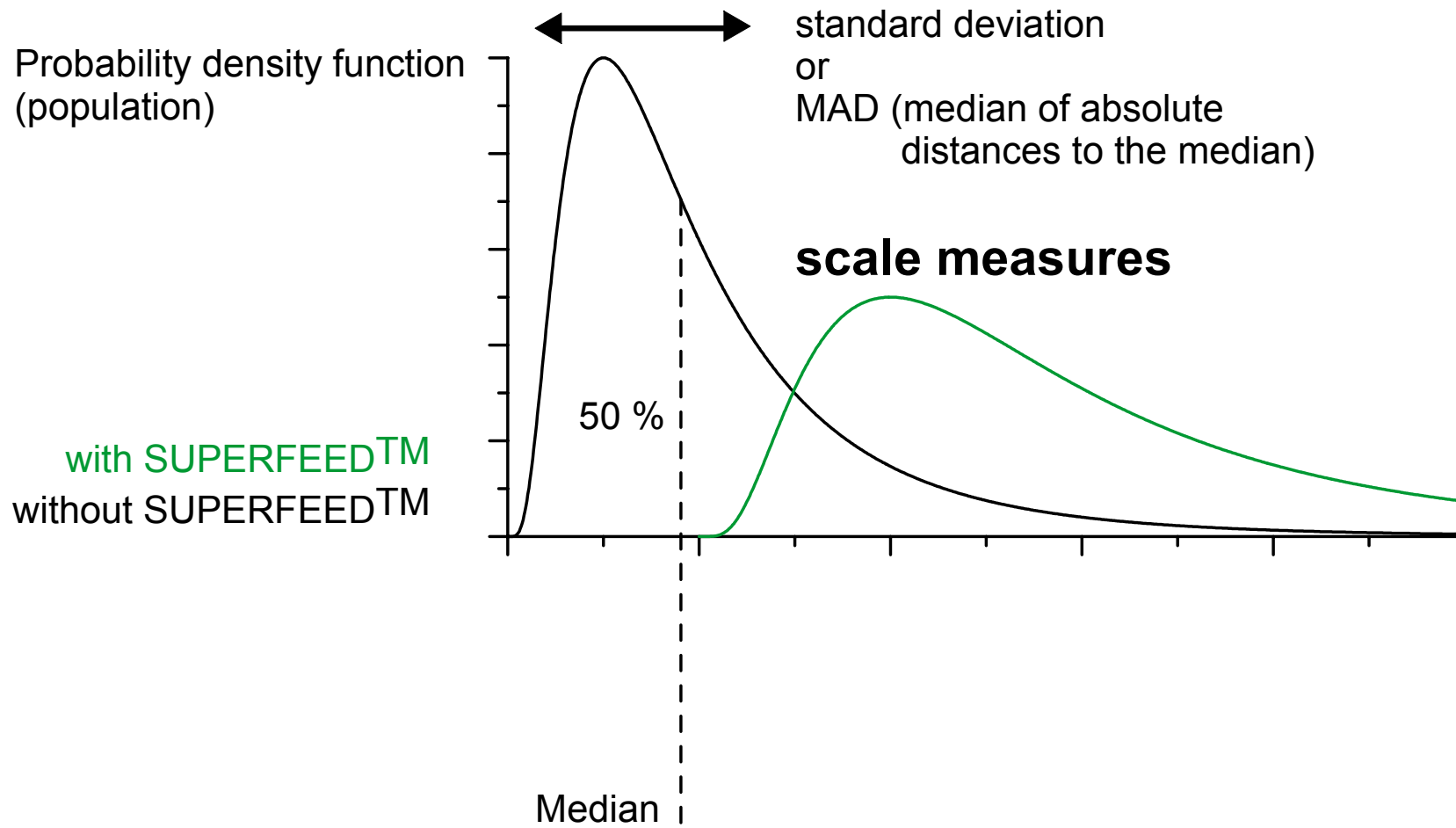
Michael Schumacher moves to Giebichenstein

⇒ 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.

Probability density function (population)

standard deviation
or
MAD (median of absolute distances to the median)

scale measures

50 %

with SUPERFEED^{TM}
without SUPERFEED^{TM}

Median

	Treatment		Control	ν
Data	$x1$		$x2$	
Unpaired				
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,

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

$MED(x1) =$ sample median,

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

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

$x1(1) =$ weight of 1st copy of cow 1
         (with SUPERFEED$^{\text{TM}}$)

$x2(1) =$ weight of 2nd copy of cow 1
         (without SUPERFEED$^{\text{TM}}$)

$x3(i) = x1(i) - x2(i)$

$\widehat{MED}_d =$ sample median$\{x3(i)\}$

advantage: inter-cow variability has no effect

	Treatment	Control	ν
Data	$x1$	$x2$	
Paired			
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) = $ median of $\{x3(i) = x1(i) - x2(i)\}, \ i = 1, \ldots, n1$

Note: measures of scale not of interest in paired experiment

Our aim: estimate confidence intervals for difference measures

$$\widehat{d_{AVE}}, \widehat{d_{MED}}, \widehat{d_{STD}}, \widehat{d_{MAD}} \text{ \textbf{(unpaired)}}$$

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

Problems: (partly) complicated measures
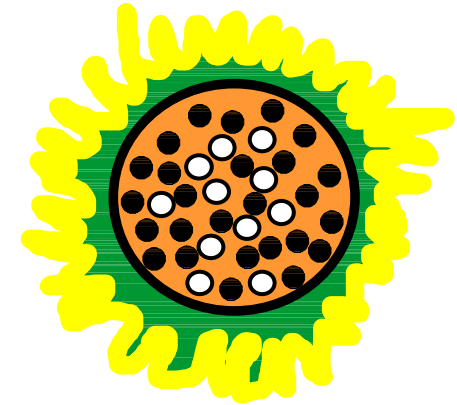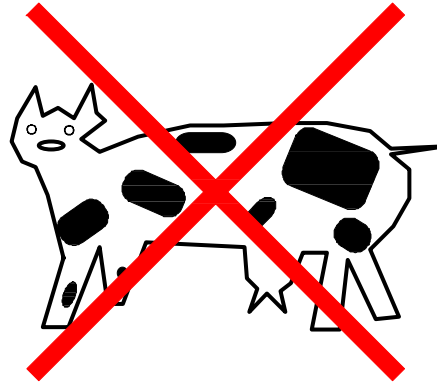
few data (5, 10)

non-normal distributions

$\Rightarrow$ 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.

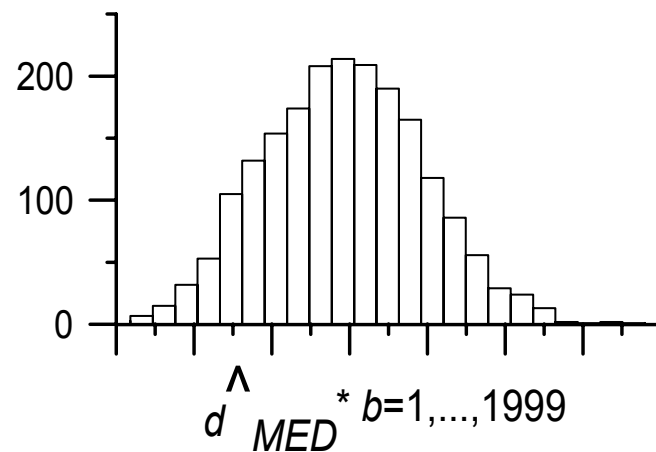http://www.uni-leipzig.de/~meteo/MUDELSEE/talk/boottalk.pdf

13

Bootstrap resampling

example

$n1 = 4$ $\{x1(1),\ \ x1(2),\ \ x1(3),\ \ x1(4)\}$ $\widehat{d_{MED}}$

$n2 = 3$ $\{x2(1),\ \ x2(2),\ \ x2(3)\}$

resampling ($b = 1$) with replacement

$n1^* = 4$ $\{x1^*(1), x1^*(2), x1^*(3), x1^*(4)\}$

          $\{x1(3),\ \ x1(3),\ \ x1(4),\ \ x1(1)\}$ $\widehat{d_{MED}}^{*\ b=1}$

$n2^* = 3$ $\{x2^*(1), x2^*(2), x2^*(3)\}$

          $\{x2(2),\ \ x2(3),\ \ x2(3)\}$

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

$\widehat{d_{MED}}^{* \; b=1,...,B}$ bootstrap replications

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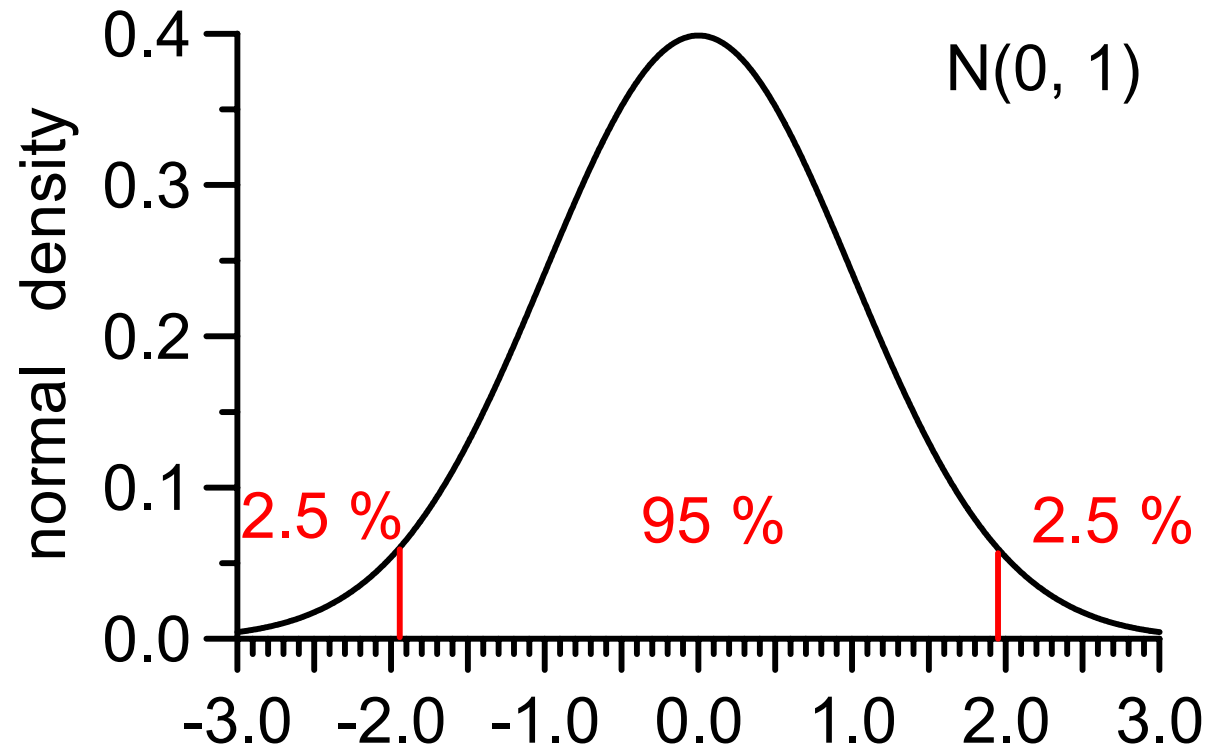(refinements account for differences)

sample std = s.e.($\hat{d}_{MED}$)

$\hat{d}_{MED}^*$ b=1,...,1999
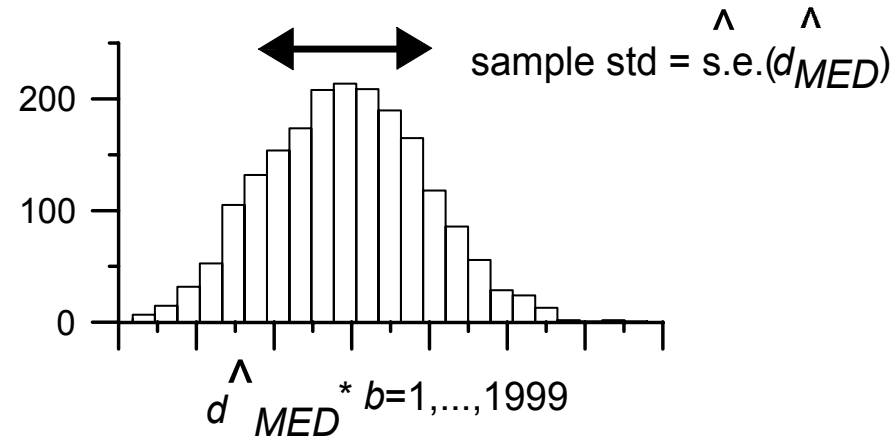
I Normal confidence interval **

$$\left[\widehat{d_{MED}} - z^{(1-\alpha)} \cdot \widehat{se}\left(\widehat{d_{MED}}\right), \ \widehat{d_{MED}} + z^{(1-\alpha)} \cdot \widehat{se}\left(\widehat{d_{MED}}\right) \right]$$
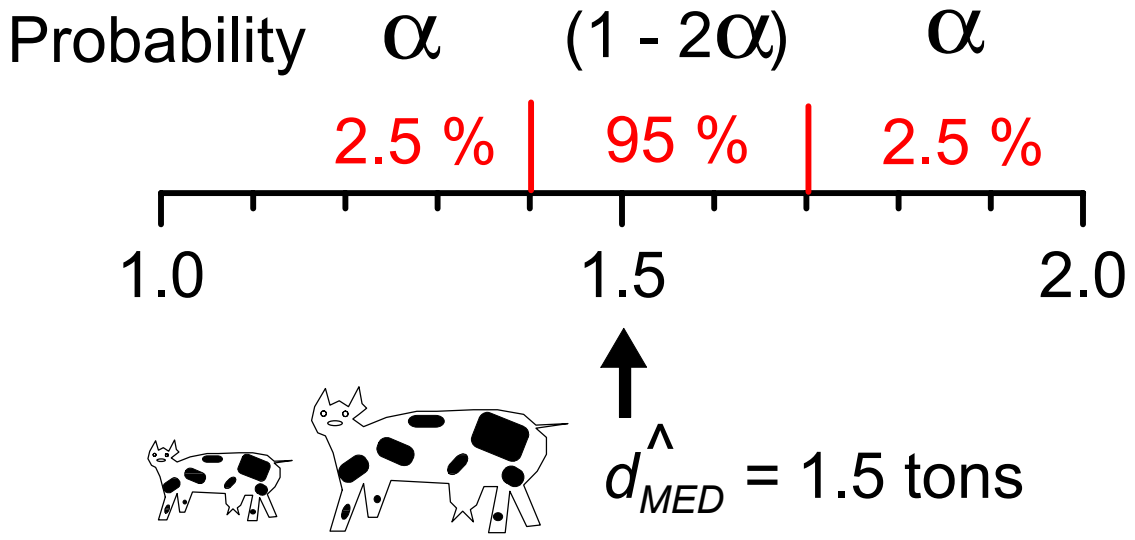
$z^{(1-\alpha)}$: standard normal percentile (e. g., $z^{(0.975)} = 1.96$) *

16

*

N(0, 1)

normal density

2.5 %   95 %   2.5 %

**

Confidence interval:

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

Probability $\quad\alpha\quad\quad (1 - 2\alpha)\quad\quad \alpha$

2.5 % | 95 % | 2.5 %

1.0   1.5   2.0

$\hat{d}_{MED}$ = 1.5 tons

18

II Student's t confidence interval

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

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

$\widehat{s.e.}(\widehat{d_{MED}})$: unkown mean
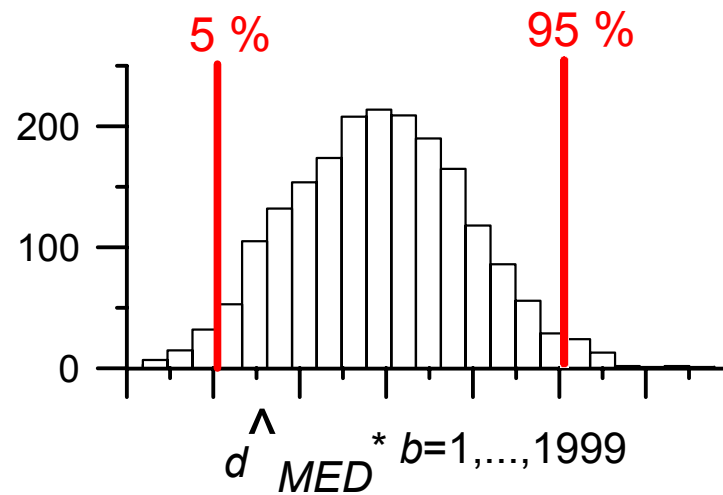
$\Rightarrow$ wider confidence interval than normal

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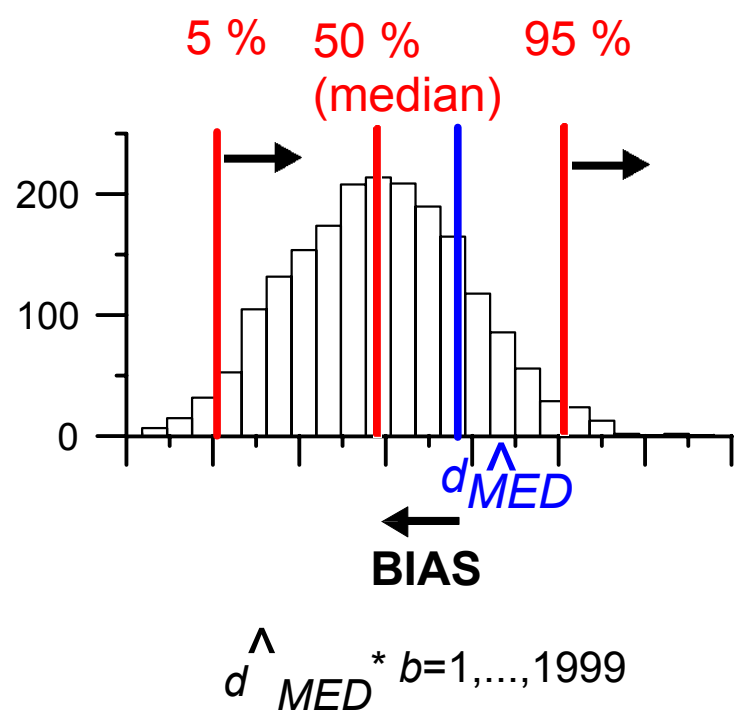terms of $C \propto 1/n$ and "faster" : second-order accurate

5 %　　　　　95 %

$\hat{d}_{MED}$　${}^{*\,b=1,...,1999}$

III Percentile confidence interval

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

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

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

5 %   50 %   95 %
       (median)

$d\widehat{}_{MED}$

BIAS

$d\widehat{}_{MED}*\,b=1,...,1999$

IV BCa confidence interval

bias correction $\widehat{z_0}$

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

BCa interval: $\left[\widehat{d_{MED}}^{*b=1,...,B=\infty}(\textcolor{red}{\alpha 1}),\ \widehat{d_{MED}}^{*b=1,...,B=\infty}(\textcolor{red}{\alpha 2})\right]$

$$\alpha 1 = \Phi\left[\widehat{z_0} + \frac{\widehat{z_0} + z^{(\alpha)}}{1 - \widehat{a}\left(\widehat{z_0} + z^{(\alpha)}\right)}\right],$$

$$\alpha 2 = \Phi\left[\widehat{z_0} + \frac{\widehat{z_0} + z^{(1-\alpha)}}{1 - \widehat{a}\left(\widehat{z_0} + z^{(1-\alpha)}\right)}\right],$$

$$\widehat{z_0} = \Phi^{-1}\left(\frac{\text{number of replications where } \widehat{d_{MED}}^{*b} < \widehat{d_{MED}}}{B}\right),\ \textcolor{red}{\left[\text{Note: } \Phi^{-1}(1/2) = 0\right]}$$

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

23

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
$\Rightarrow$ no "normal" theory

small data sizes
$\Rightarrow$ no asymptotic theory

$\widehat{d_{MED}}$, $\widehat{d_{MAD}}$, $\widehat{MED_d}$: order statistics
$\Rightarrow$ 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. *

*

$$\begin{array}{ccccc}
& & x1 & x2 & \text{data} \\
& & x1' = g_1(x1) & x2' = g_2(x2) & \text{transformation} \\
& & x1' = \log(x1) & x2' = \log(x2) & \boxed{\text{example}} \\
\widehat{d'_{MED}} & = & MED(x1')- & MED(x2') & \\
& = & MED[\log(x1)]- & MED[\log(x2)] & \\
& = & \log[MED(x1)]- & \log[MED(x2)] & \log = \text{monotonic} \\
& = & \log\left[\dfrac{MED(x1)}{MED(x2)}\right] & &
\end{array}$$

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

g_1, g_2 different and/or more complex: $\widehat{d'_{MED}}$ 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

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

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

median = *a*1 + *b*1   median = *a*2 + *b*2

d_{MED} = *a*1 + *b*1 − (*a*2 + *b*2)
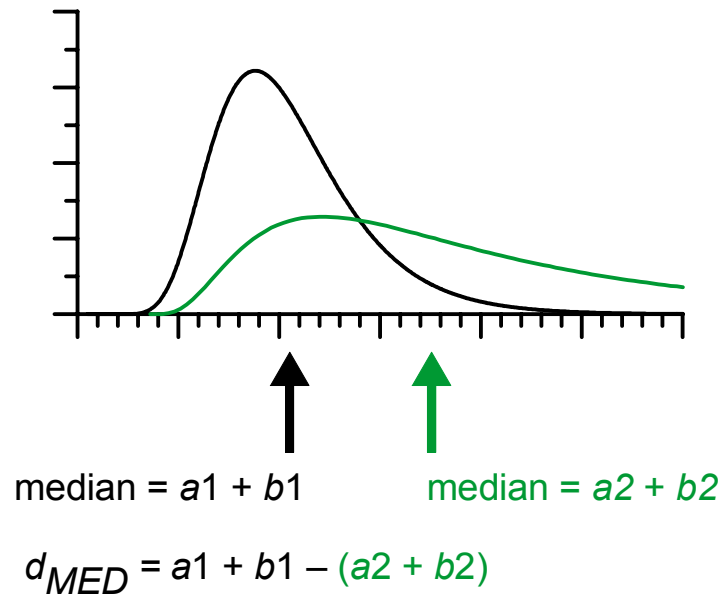
Draw random sample $x1 \sim \text{LN}(a1, b1, s1)$, size $n1$
Draw random sample $x2 \sim \text{LN}(a2, b2, s2)$, size $n2$
Correlation: ρ_{LN}

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

Count: $d_{MED} <$ lower bound of confidence interval (nominal: $P = \alpha$)
$\quad\quad d_{MED} \in$ confidence interval (nominal: $P = 1 - 2\alpha$)
$\quad\quad d_{MED} >$ upper bound of confidence interval (nominal: $P = \alpha$)

Repeat procedure: 10000 Monte Carlo simulations

Compare empirical with nominal coverages

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)
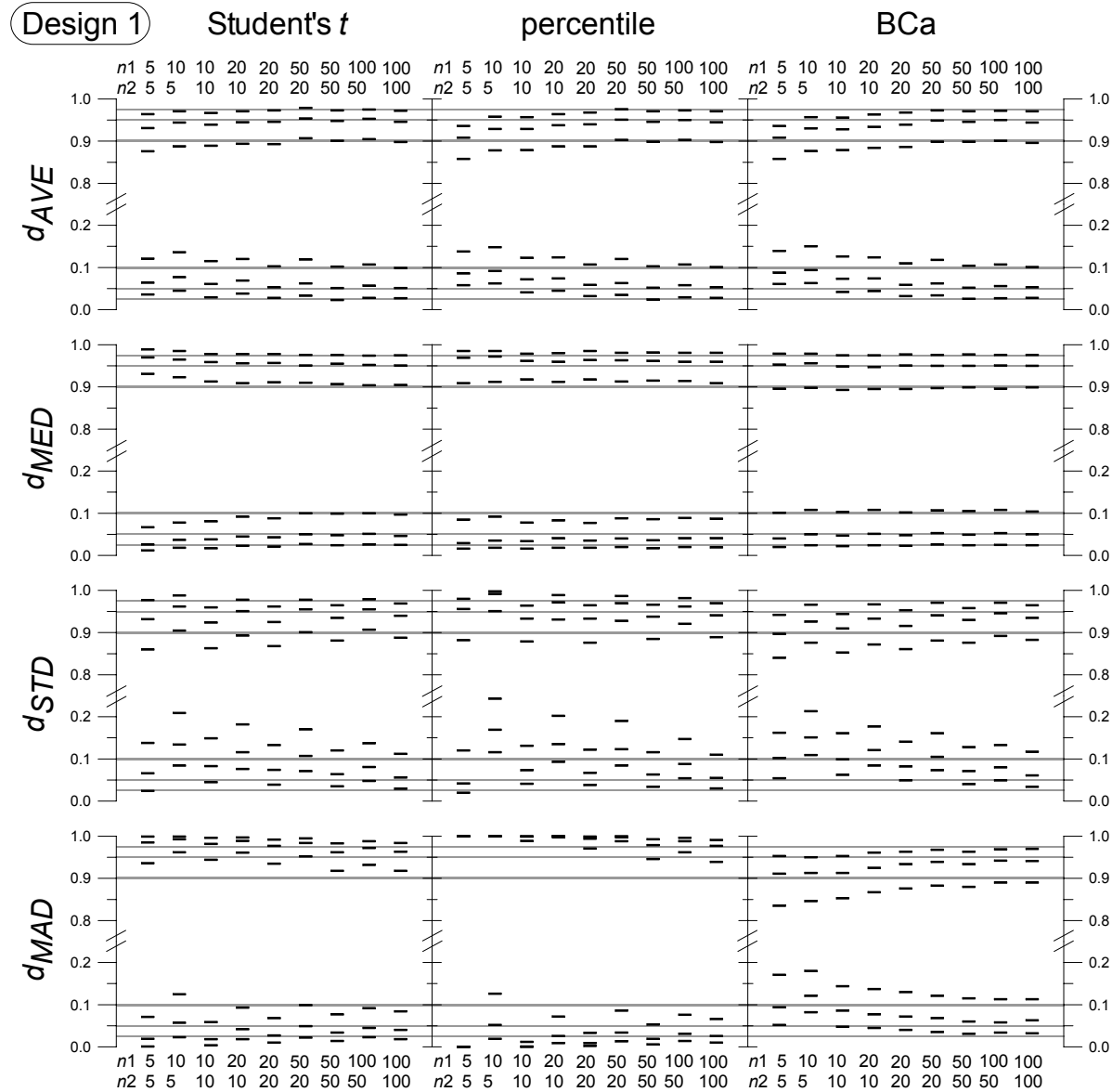
33

$\rho_{\mathsf{LN}} = 0$ (correctly specified unpaired)
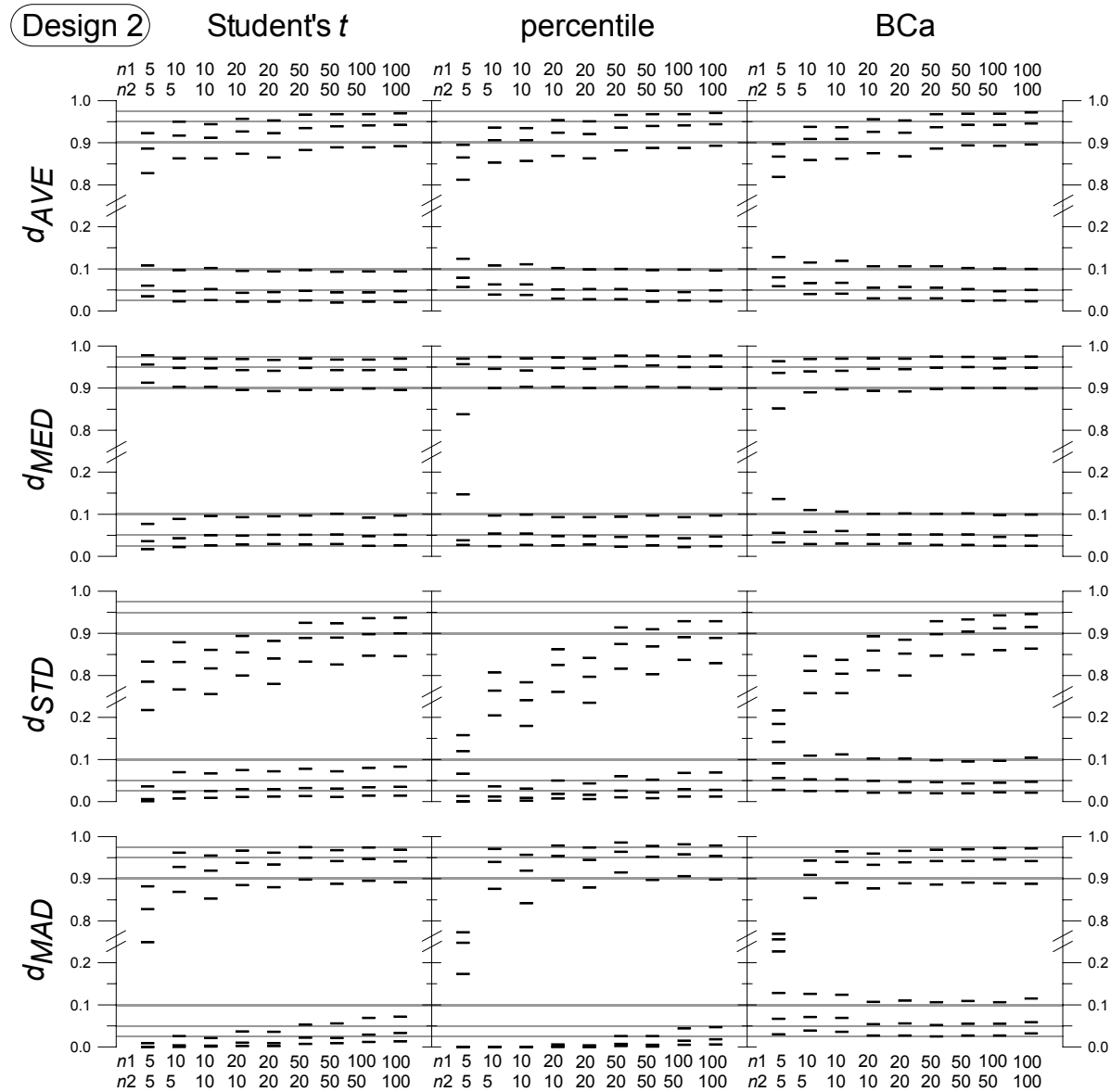
$\rho_{\mathsf{LN}} = 0$ (correctly specified unpaired)
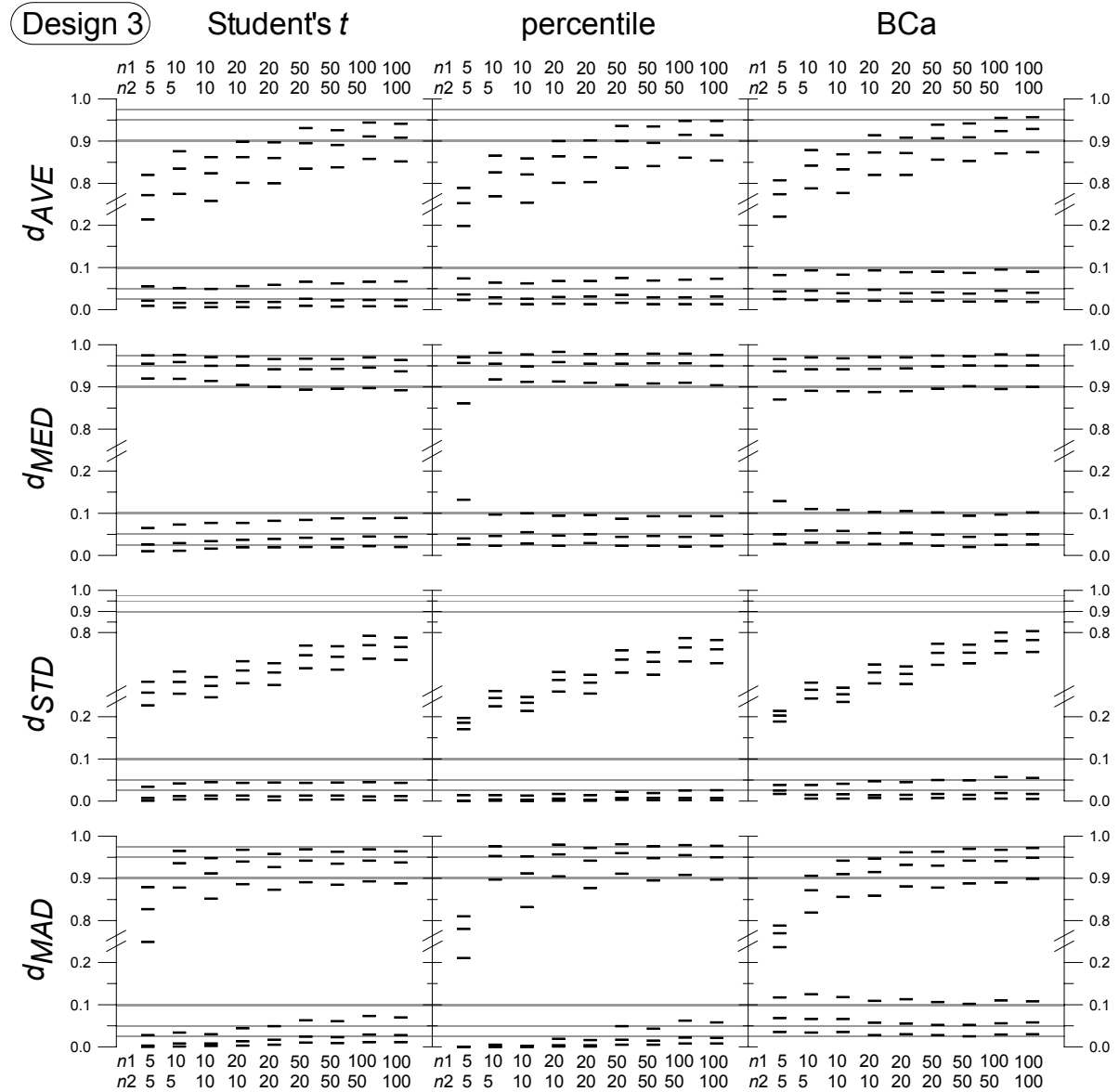
$\rho_{\text{LN}} = 0$ (correctly specified unpaired)
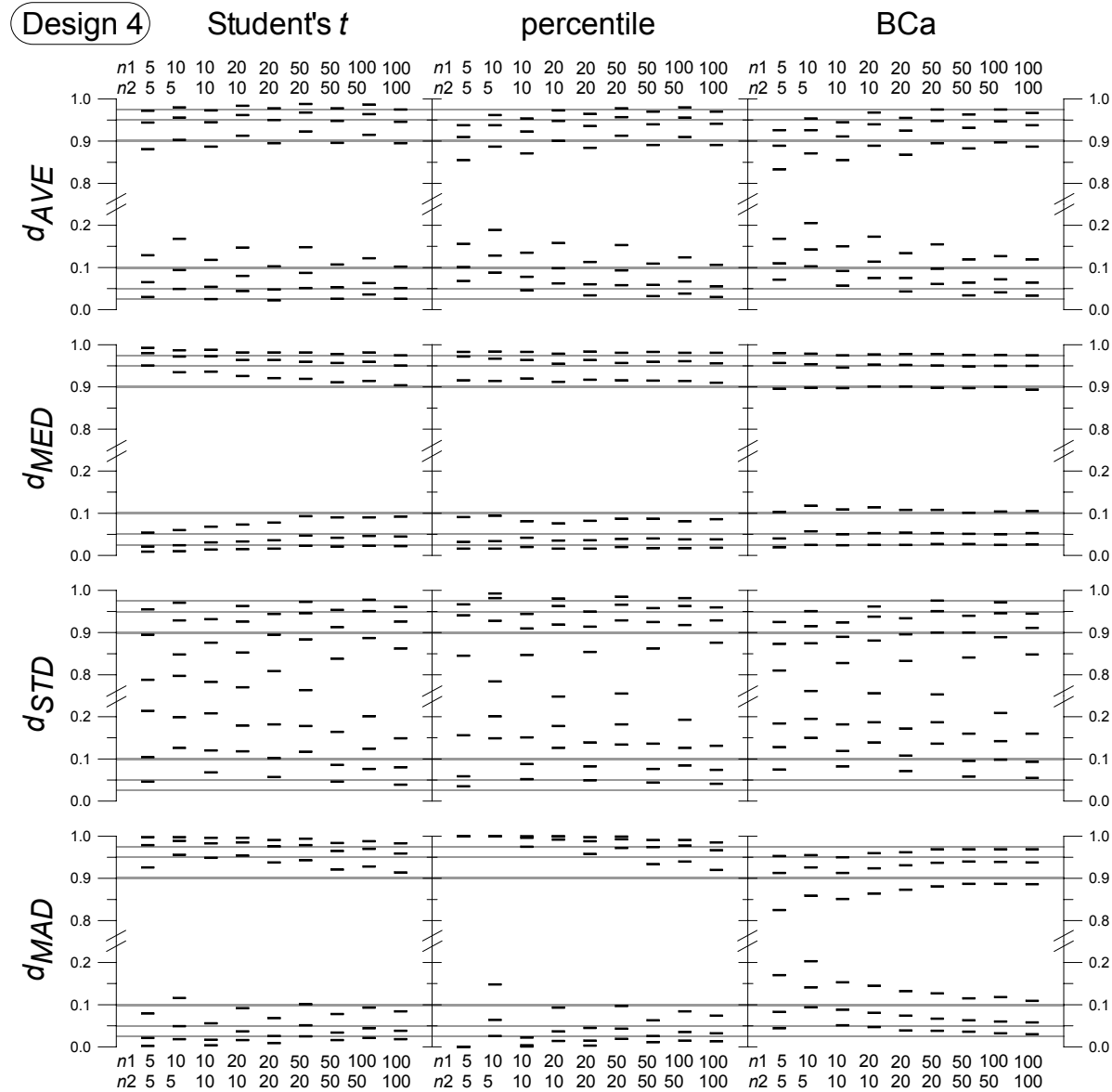
$\rho_{LN} \neq 0$ (mis-specified unpaired)
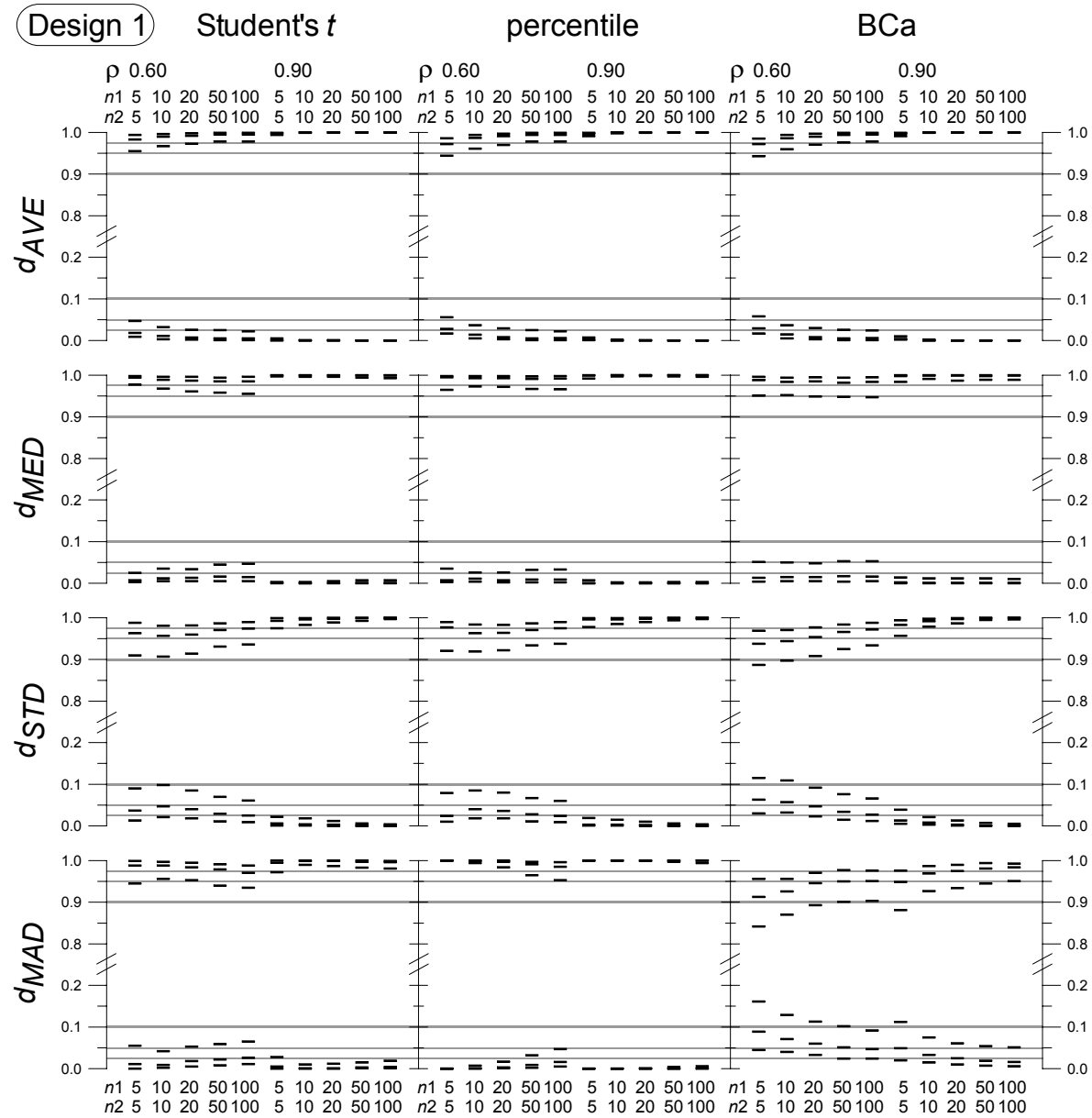
$\rho_{\mathsf{LN}} \neq 0$ (correctly specified paired)
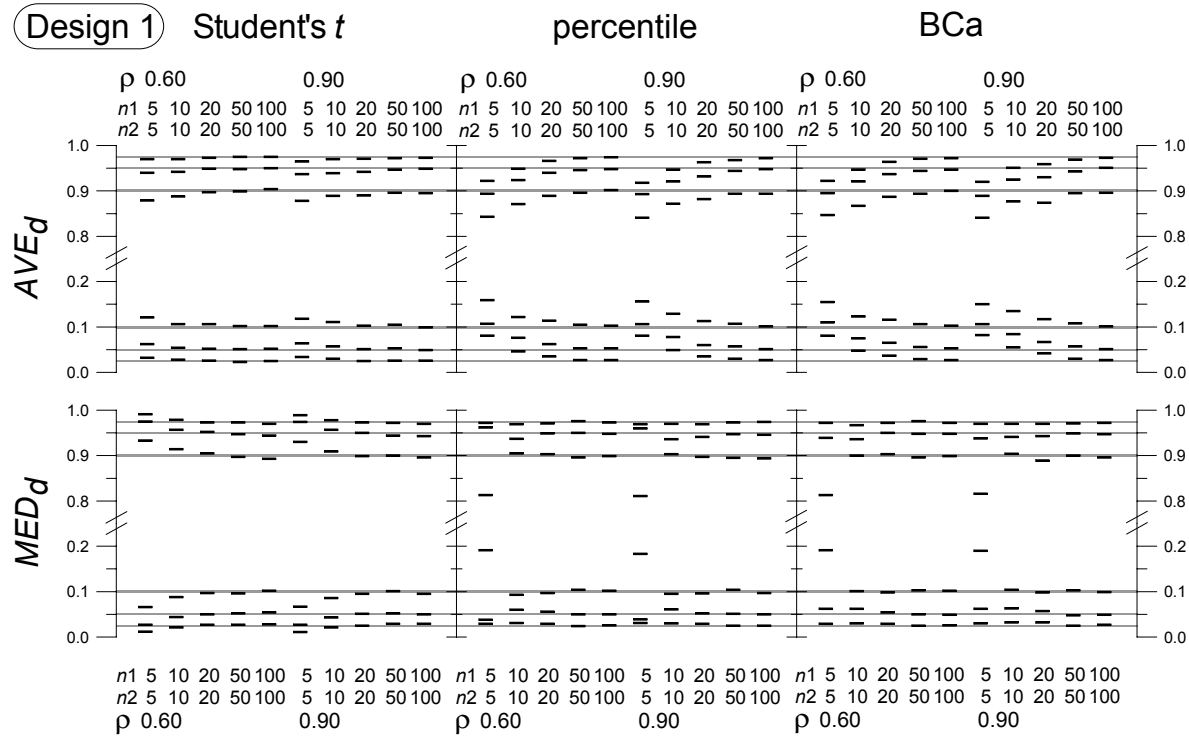
Result

- 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 $n1 \gtrsim 10$ and $n2 \gtrsim 10$.

- BCa confidence intervals of $\widehat{d_{MED}}$ as measure of difference in location offer good coverage performance in unpaired experiments for $n1 \gtrsim 20$ and $n2 \gtrsim 20$.

- BCa confidence intervals of $\widehat{d_{MAD}}$ as measure of difference in scale offer acceptably coverage performance in unpaired experiments for $n1 \gtrsim 50$ and $n2 \gtrsim 50$.

Helianthus annuus L. — source–sink ratio manipulations

$x =$ percentage of unfilled achenes

shading   defoliation
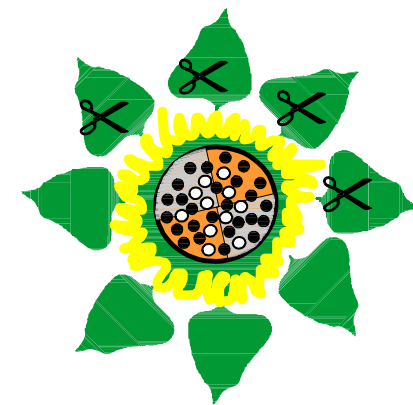
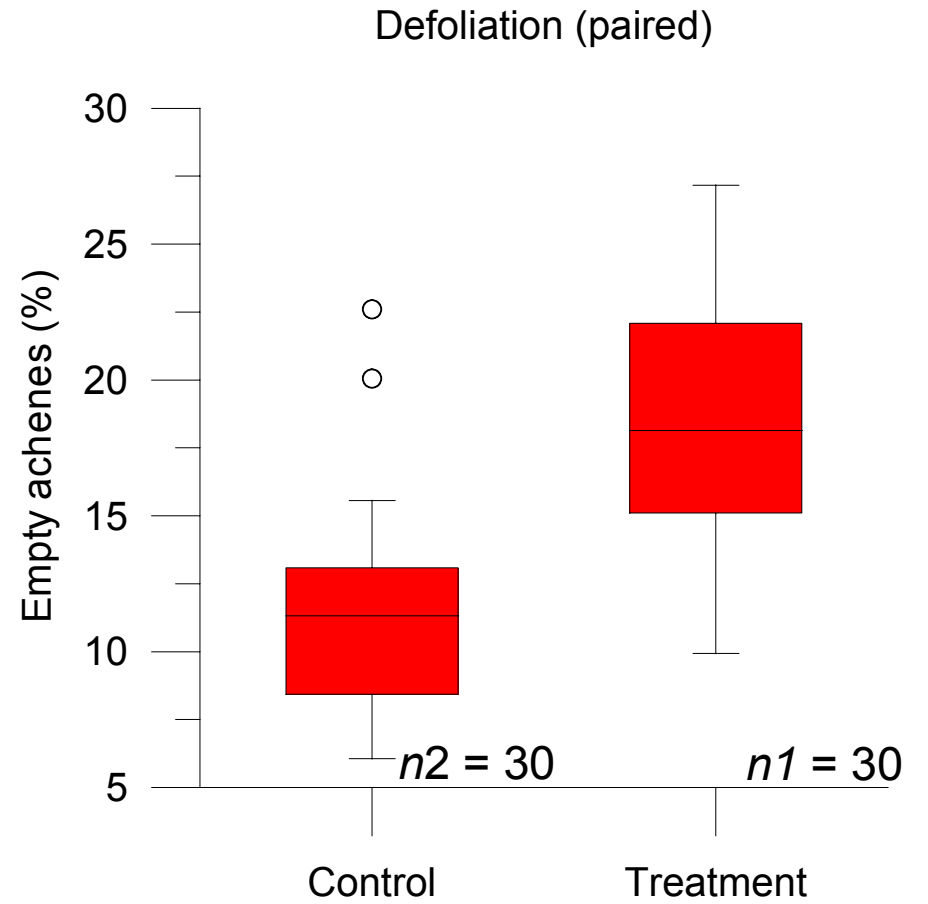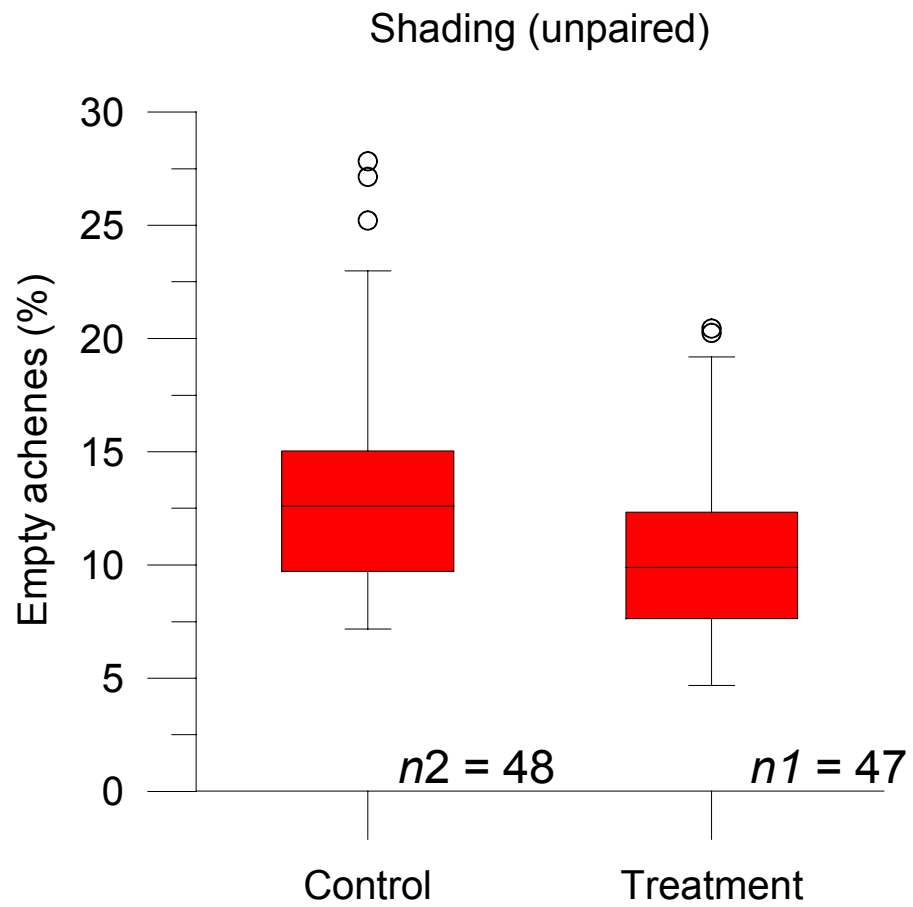ratio $\uparrow$ ratio $\downarrow$

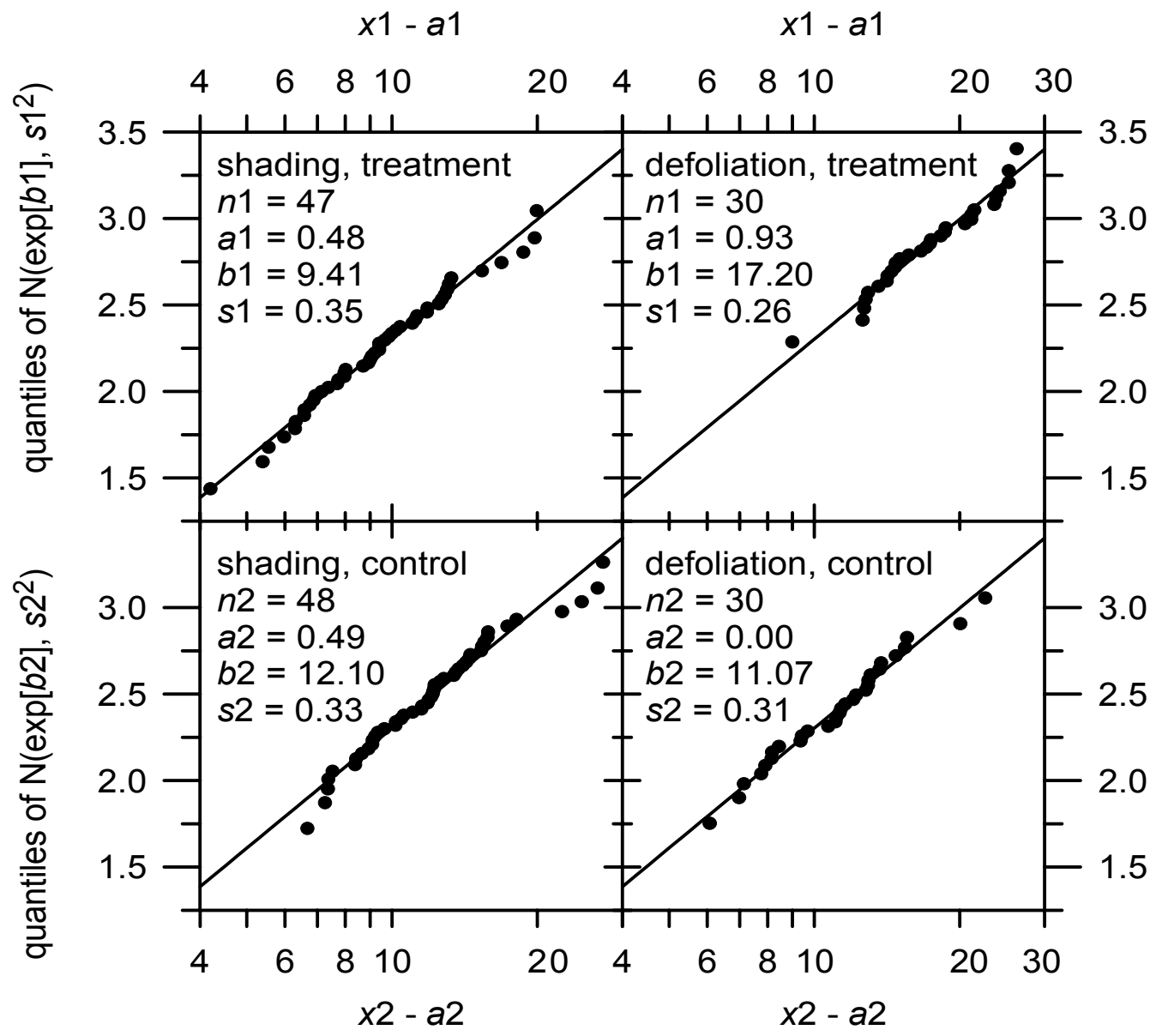unpaired experiment   paired experiment

Shading (unpaired)
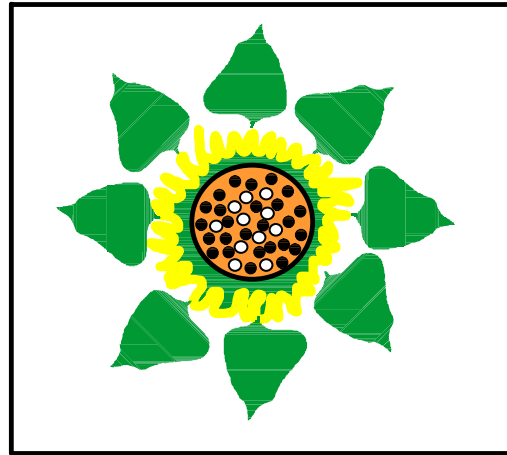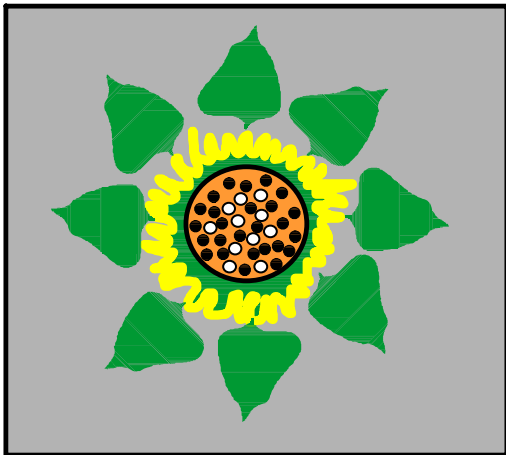
Defoliation (paired)
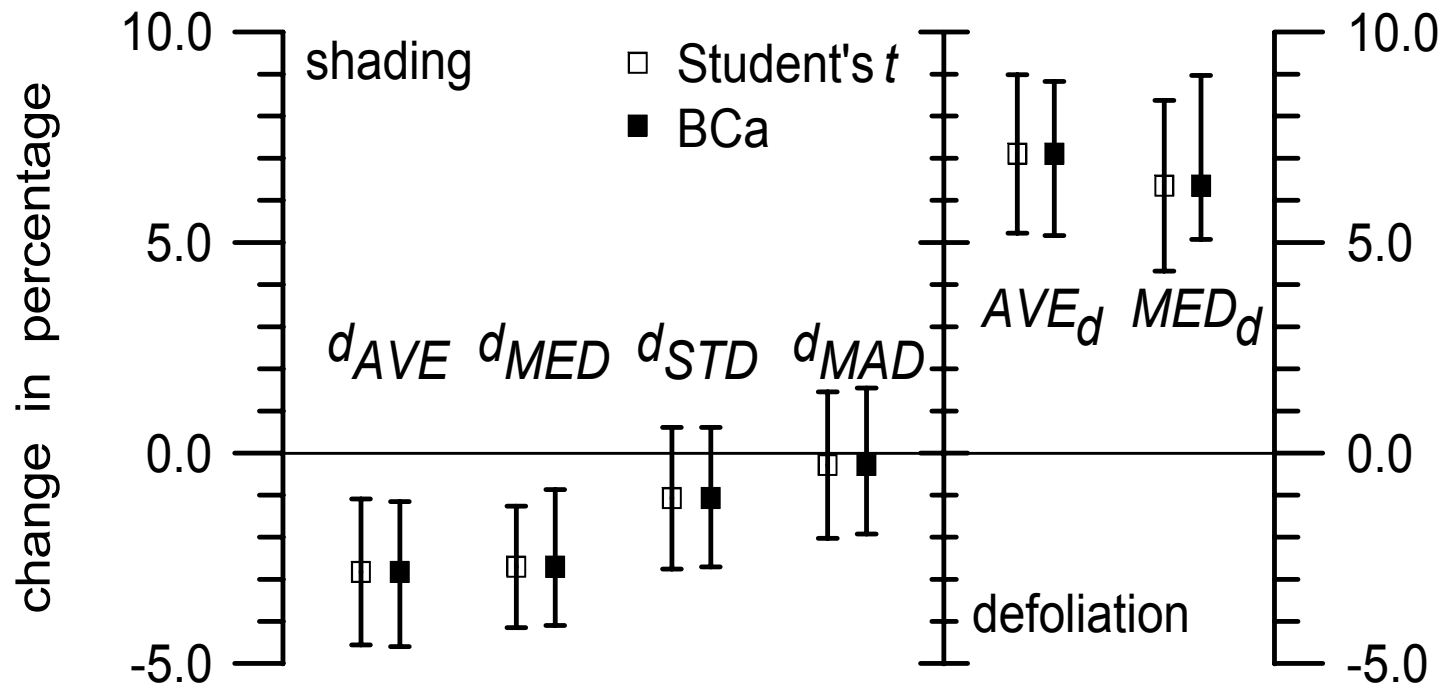
2SAMPLES demonstration

```
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 ?)**