

# Correlating Bootstrap with Clinical Trials

*By Mohd Tariq  
PhUSE SDE – Trivandrum  
30<sup>th</sup> July 2016*



CHILTERN  
Designed Around You®



# Bootstrap in Life



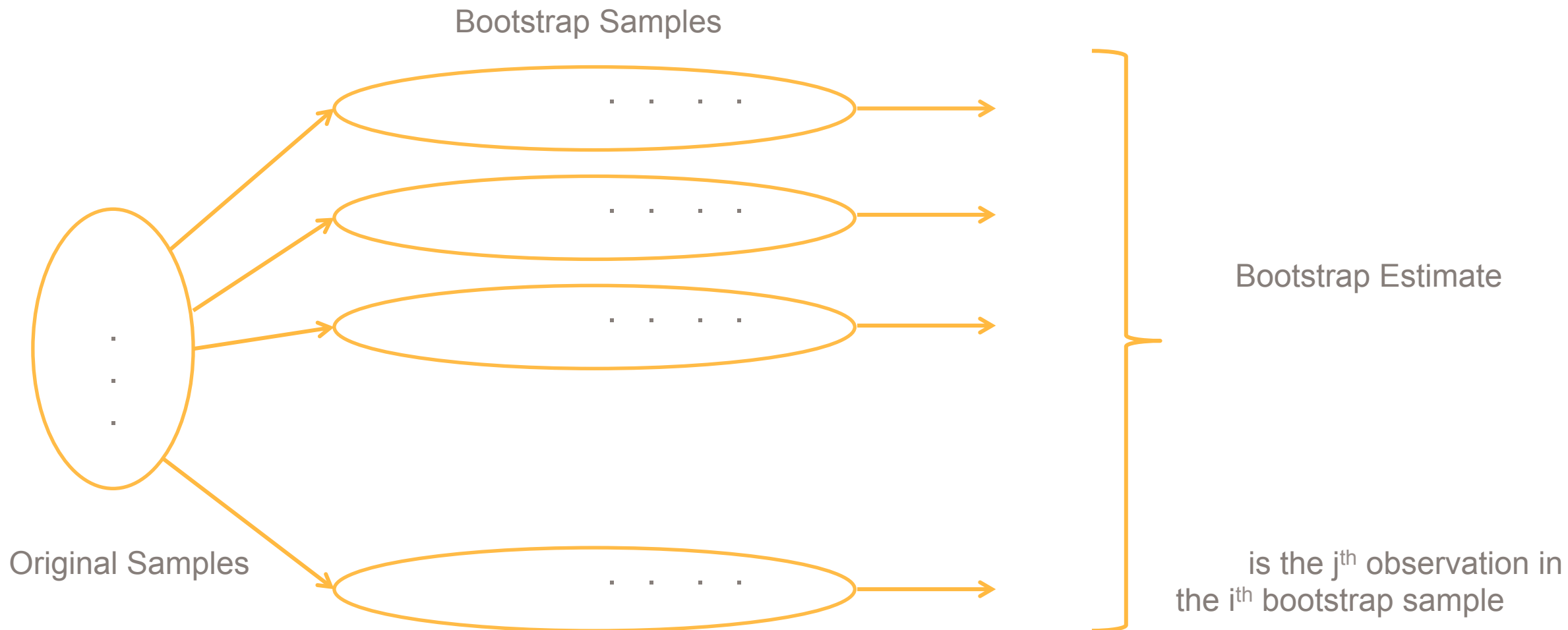
# Agenda of Presentation

- Introduction
- Reliable estimate for SE
- CI using Bootstrap
  - Bootstrap  $t$  – interval
  - Percentile Interval
  - BCa Interval
- Hypotheses Testing using Bootstrap

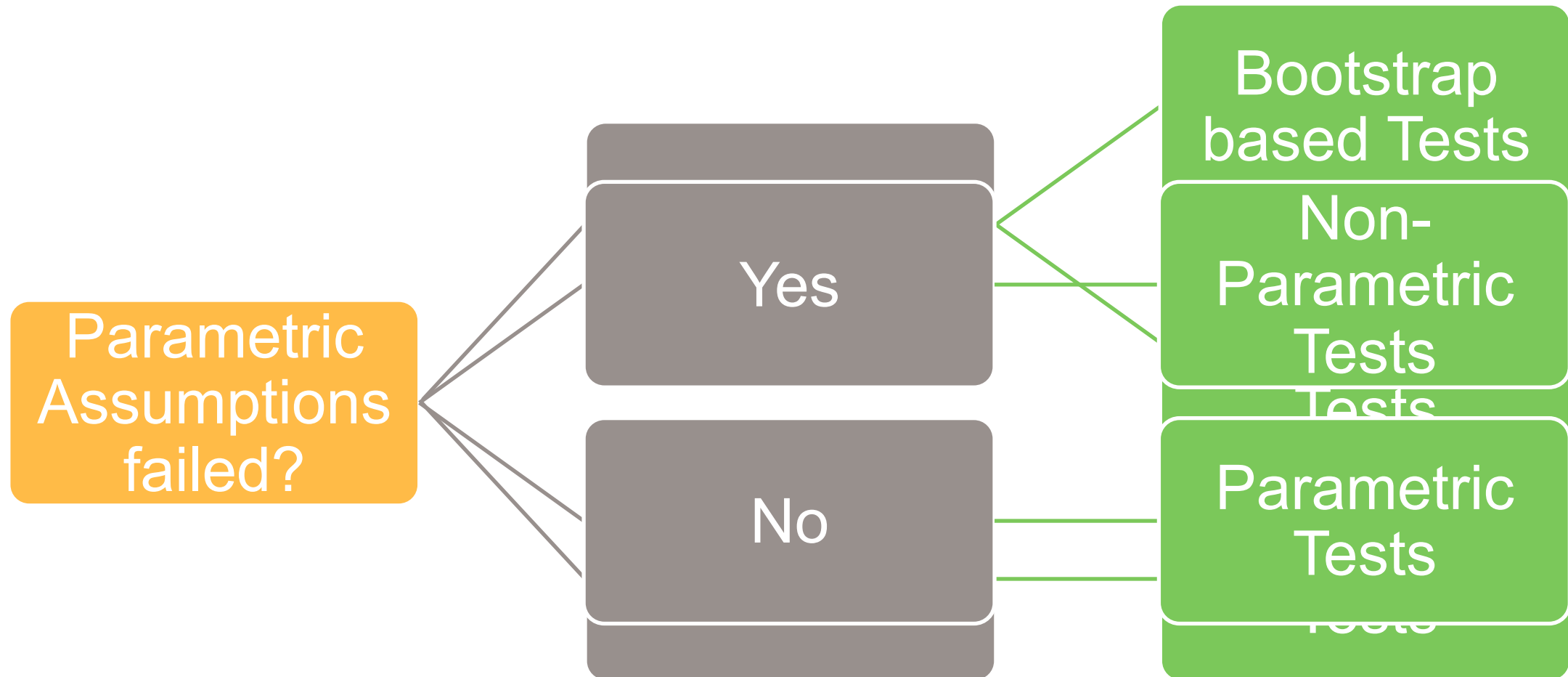
- Bootstrap is a resampling technique.
- The idea behind bootstrap is to use the sample at hand as a population to resample (with replacement) from the sample at hand and create a large number of “phantom samples” known as bootstrap samples.
- The method of bootstrapping was first introduced by Efron (1979).
- The use of the term 'bootstrap' comes from the phrase “To pull oneself up by one's bootstraps”.



# Simple Algorithm



# When we can use Bootstrap?



- Draw independent bootstrap samples from original sample.
- Compute the bootstrap estimate of mean (statistic of interest) for all the Bootstrap sample.
- The sample SD of these estimates will give the SE of the estimate (here mean).

For mean we have a neat formula to calculate the standard error. But for estimate other than mean (like median), there is no neat formula that enables us to compute the numerical value of the estimate exactly. Hence bootstrap provide a solution to this.

- It is recommend to use at least 200 replicates to get a reliable estimate of SE.



**CI using Bootstrap**

## Consider a case study

Consider a randomized, active control clinical trial in which the objective is to test for non inferiority of the experimental treatment compared to control.

Where  $\delta$  is the Non-inferiority Margin, which is equal to -3.

Sample Size = 10 for each treatment.

Number of Replicate = 2000.

| Data      |      |           |      |
|-----------|------|-----------|------|
| Treatment | AVAL | Treatment | AVAL |
| Active    | 42   | Control   | 51   |
| Active    | 50   | Control   | 42   |
| Active    | 40   | Control   | 40   |
| Active    | 45   | Control   | 53   |
| Active    | 47   | Control   | 36   |
| Active    | 57   | Control   | 41   |
| Active    | 39   | Control   | 42   |
| Active    | 61   | Control   | 49   |
| Active    | 53   | Control   | 31   |
| Active    | 46   | Control   | 45   |

The normal t – interval is:

$$[ \theta - t_{\downarrow(n-1)\uparrow(1-\alpha)} se, \theta - t_{\downarrow(n-1)\uparrow(\alpha)} se ]$$

- Does not adjust for skewness.
- Errors that can result when  $\theta$  is not a sample mean.

Bootstrap t-interval adjust for these errors.

- Generate B independent bootstrap samples from the original sample.
- Compute  $Z^*(b) = (\hat{\theta}^*(b) - \hat{\theta}) / se(\hat{\theta}^*(b))$ , for each bootstrap samples.
- Now take  $\alpha^{\text{th}}$  percentile &  $(1-\alpha)^{\text{th}}$  percentile value from all  $Z^*(b)$ .
- Construct the interval.

| LOWER_CI | UPPER_CI |
|----------|----------|
| -2.75830 | 12.5994  |

# Percentile Interval

- Draw B independent sample from the original sample.
- Compute the required statistics for each sample.
- The interval is given by the 2.5<sup>th</sup> percentile as lower limit and 97.5<sup>th</sup> percentile as the upper limit of the 95% CI.

| LIM2_5 | LIM97_5 |
|--------|---------|
| -1     | 11.8    |

- It is also a percentile based interval.
- The percentile used depend on two numbers  $a$  and  $z_{\downarrow 0}$ , called acceleration and bias correction.

$$z_{\downarrow 0} = \Phi^{-1} (\#\{\theta^* (b) < \theta\} / B) \dots\dots\dots (1)$$

$$a = \frac{\sum_{i=1}^j (\theta_{\downarrow}(\cdot) - \theta_{\downarrow}(i))^3}{6 \{ \sum_{i=1}^j (\theta_{\downarrow}(\cdot) - \theta_{\downarrow}(i))^2 \}^{3/2}} \dots\dots\dots (2)$$

Where  $\theta_{\downarrow}(i)$  = the statistics of interest calculated on the jackknifed sample with the i-th record removed.

$\theta_{\downarrow}(\cdot)$  = the mean of the j jackknife sample estimate.

j = number of jackknife samples.

B = number of bootstrap samples.

$$\alpha_{\downarrow 1} = \Phi(z_{\downarrow 0} + z_{\downarrow 0} + \Phi^{-1}(\alpha) / (1 - \alpha(z_{\downarrow 0} + \Phi^{-1}(\alpha)))) \dots \dots \dots (3)$$

and

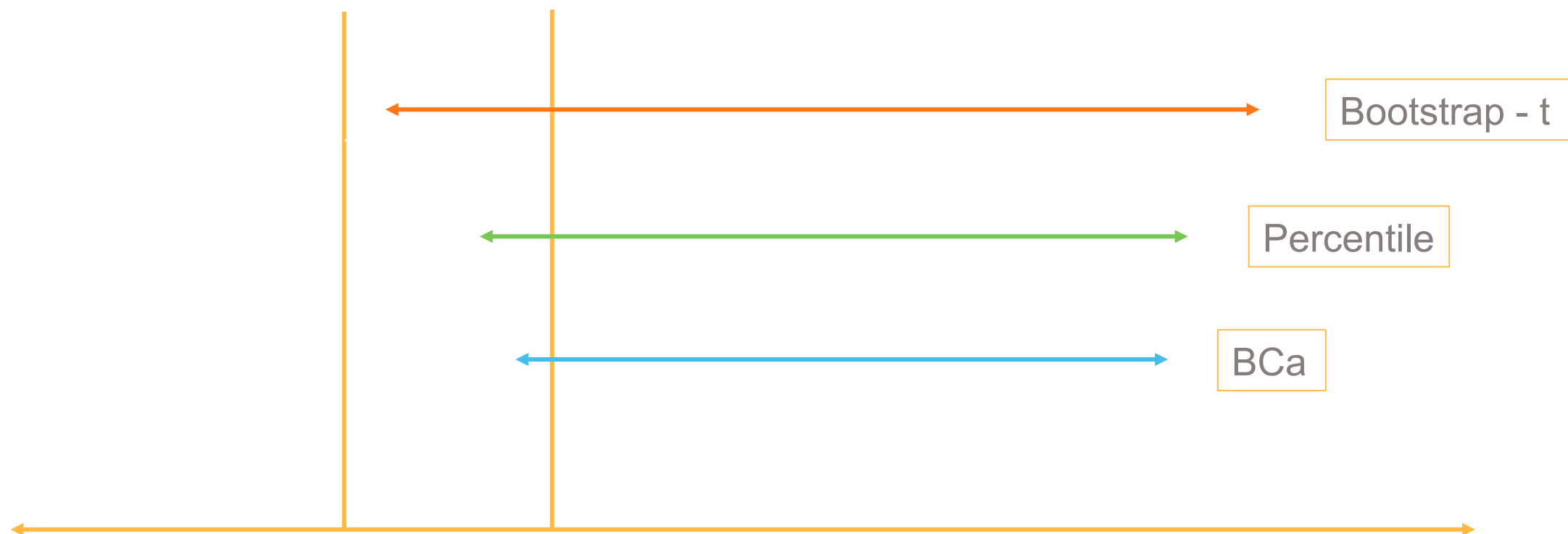
$$\alpha_{\downarrow 2} = \Phi(z_{\downarrow 0} + z_{\downarrow 0} + \Phi^{-1}(1 - \alpha) / (1 - \alpha(z_{\downarrow 0} + \Phi^{-1}(1 - \alpha)))) \dots \dots \dots (4)$$

- Calculate  $N_{\downarrow 1} = B \times \alpha_{\downarrow 1}$   
 $N_{\downarrow 2} = B \times \alpha_{\downarrow 2}$
- Then the CI is given by  
( $N_{\downarrow 1}$  th observation,  $N_{\downarrow 2}$  th observation)  
when arranged in ascending order.

# Results from BCa Interval

| Acceleration ( $\alpha$ ) | Bias ( $z_{\downarrow\alpha}$ ) | $\alpha_{\downarrow 1}$ | $\alpha_{\downarrow 2}$ | $N_{\downarrow 1}$ | $N_{\downarrow 2}$ |
|---------------------------|---------------------------------|-------------------------|-------------------------|--------------------|--------------------|
| ACCEL                     | BIAS                            | ALPHA1                  | ALPHA2                  | N1                 | N2                 |
| 0.012248                  | 0.010027                        | 0.029084                | 0.97876                 | 58.1682            | 1957.51            |

| Lower Limit | Upper Limit |
|-------------|-------------|
| LIM2_90     | LIM97_87    |
| -0.5        | 11.7        |



We can conclude Non-inferiority.



## Hypotheses Testing using Bootstrap

- In bootstrap hypothesis test for comparing the two means, there is no compelling reason to assume equal variances assumption, so we do not make that assumption.
- We need estimator of the two population that use only the assumption of a common mean.

Suppose  $\bar{x}$  be the mean of the combined sample. We can translate both samples and then resample each population separately to have common mean  $\bar{x}$ .

We can transform data as :

$$x_{\downarrow i} = x_{\downarrow i} - \bar{x} + \bar{x}$$

$$y_{\downarrow i} = y_{\downarrow i} - \bar{y} + \bar{x}$$

# Algorithm for Comparing two means

- Transform the data such that both the sample have same mean, which should be under  $H_0$ .

$$T \downarrow i = T \downarrow i - T + O$$

$$A \downarrow i = A \downarrow i - A + O$$

Where  $T$  &  $A$  are the group mean &  $O$  is the overall mean (Pooled mean).

- From B bootstrap dataset ( $T^*$ ,  $A^*$ ) where  $T^*$  is sampled with replacement for  $T \downarrow 1$ ,  $T \downarrow 2$ , .....,  $T \downarrow n$  and  $A^*$  is sampled with replacement for  $A \downarrow 1$ ,  $A \downarrow 2$ , .....,  $A \downarrow m$ .

- Evaluate on each dataset following

$$t \downarrow obs = T - A / \sqrt{\sigma \downarrow 1 \uparrow^2 / n + \sigma \downarrow 2 \uparrow^2 / m}$$

$$t(x \uparrow^* b) = T \uparrow^* - A \uparrow^* / \sqrt{\sigma \downarrow 1 \uparrow^{*2} / n + \sigma \downarrow 2 \uparrow^{*2} / m}$$

Where  $t \downarrow obs$  is the observed value of statistic.  $t(x \uparrow^* b)$  is the bootstrap value of statistics.

B is the number of bootstrap samples.

- Approximate  $ASL_{boot}$  by :

$$ASL \downarrow boot = \# \{ t(x \uparrow * b) \geq t \downarrow obs \} / B$$

- Result

| COUNT | ASL   |
|-------|-------|
| 239   | 0.120 |

Since  $ASL > 0.05$ , we don't have enough evidence to reject Null Hypothesis. We can say that both the treatment means are equal.

In a recent CPMP guidance document “Guidance on Clinical Trials in Small Populations” (released for consultation in March 2005), it is stated that “some forms of Bootstrap methods make no assumptions about data distributions and so can be considered a ‘safe’ option when there are too few data to test or verify model assumptions. They prove particularly useful where very limited sample data are available”.

- Can be use to create a confidence interval for median survival.
- Can be use to detect the outlier in the data.
- Can be used in Regression when error are not normal.

- An Introduction to the Bootstrap - Bradley Efron, Robert J. Tibshirani - Springer US
- Introductory Biostatistics for the Health Sciences Modern Applications Including Bootstrap  
Michael R. Chernick, Robert H. Friis - Wiley Interscience
- A Bootstrap-Based Test for Establishing Non-inferiority in Clinical Trials - Michael Chen ,  
Farid Kianifard & Sunil K. Dhar (Journal of Biopharmaceutical Statistics)



Questions?

# Thank you



CHILTERN  
Designed Around You®