Discrimination Tests With Sureness: Thurstonian and R-Index Analysis

Graham Cleaver
Consumer Perception & Behaviour
Unilever Food & Health Research Institute
Vlaardingen

Sensometrics
July 2008

Unilever Categories & Brands

Foods

- Savoury & Dressings
- Spreads
- Weight Management
- Tea
- Ice Cream

Home and Personal Care

- Skin
- Deodorants
- Laundry #1 in D&E
- Daily Hair Care # 1 in D&E
- Household Care
- Oral Care

World Number 1

World Number 2

Local Strength

Our 12 €1 billion brands

Knorr, SUNSILK, Lipton, Blue Band, Becel, Hellmann's, Dove, Surf, Rexona, LUX, dirt is good

Content

A / Not-A Test With Sureness

- *R-Index Calculation*
- *Thurstonian Analysis*

Power Analysis

- *Alternative R-Index Analyses*
- *vs Thurstonian*

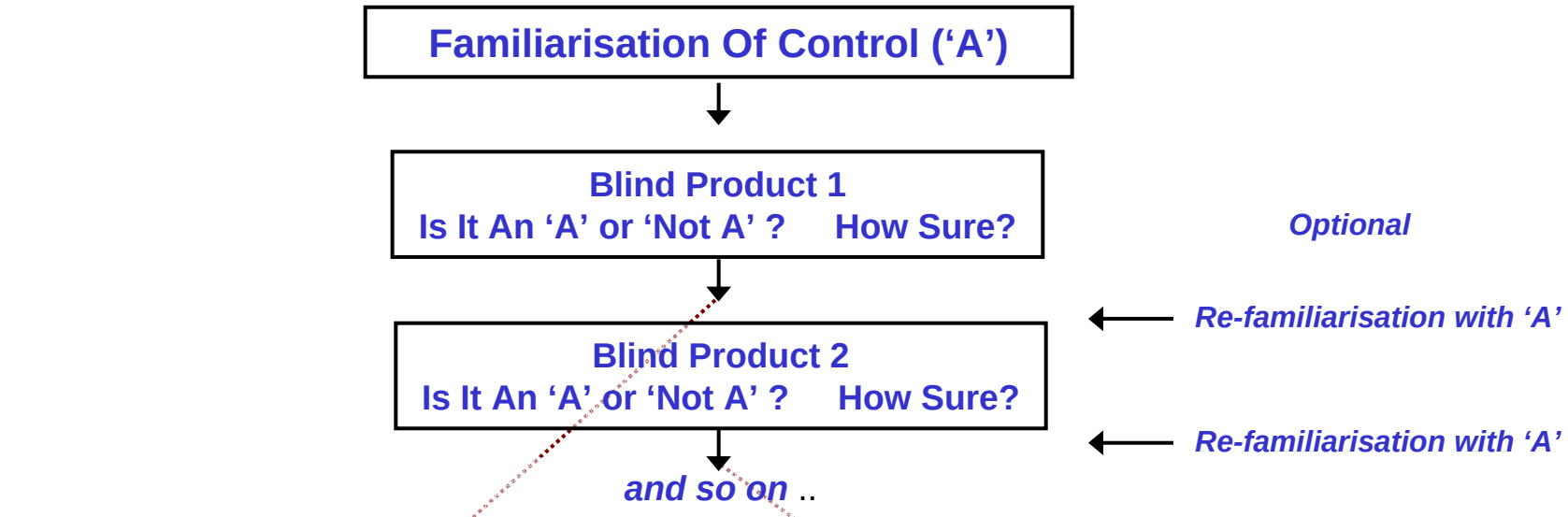
Power Charts

- *Difference Test*
- *Similarity Test*

Issues & New Developments

- *Difference Between Individuals*
- *Alternative Analysis*

A / Not-A Test Procedure



Product Identity	'A'			Response	'Not A'		Total
	Sure	Not Sure	Guess	Guess	Not Sure	Sure	
Control (A)	4	11	8	4	2	1	30
Test (Not A)	0	2	5	8	12	3	30

R Index is the percentage of all possible 'theoretical' pairwise comparisons between each Control and each Test product that would be classified correctly

Pure Chance: R Index = 50% Perfect Discrimination: R Index = 100%

In this example, R Index = 82%

R Index Alternative Analyses

1. **R-Index as binomial proportion**
   Bi J. & O'Mahony M. (1995) Table for testing the significance of the R-Index Journal of Sensory Studies 10 (4) , 341–347

2. **R-Index equivalent to Mann-Whitney U Statistic, with correction for ties**
   Mann H.B. & Whitney D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18, 50-56

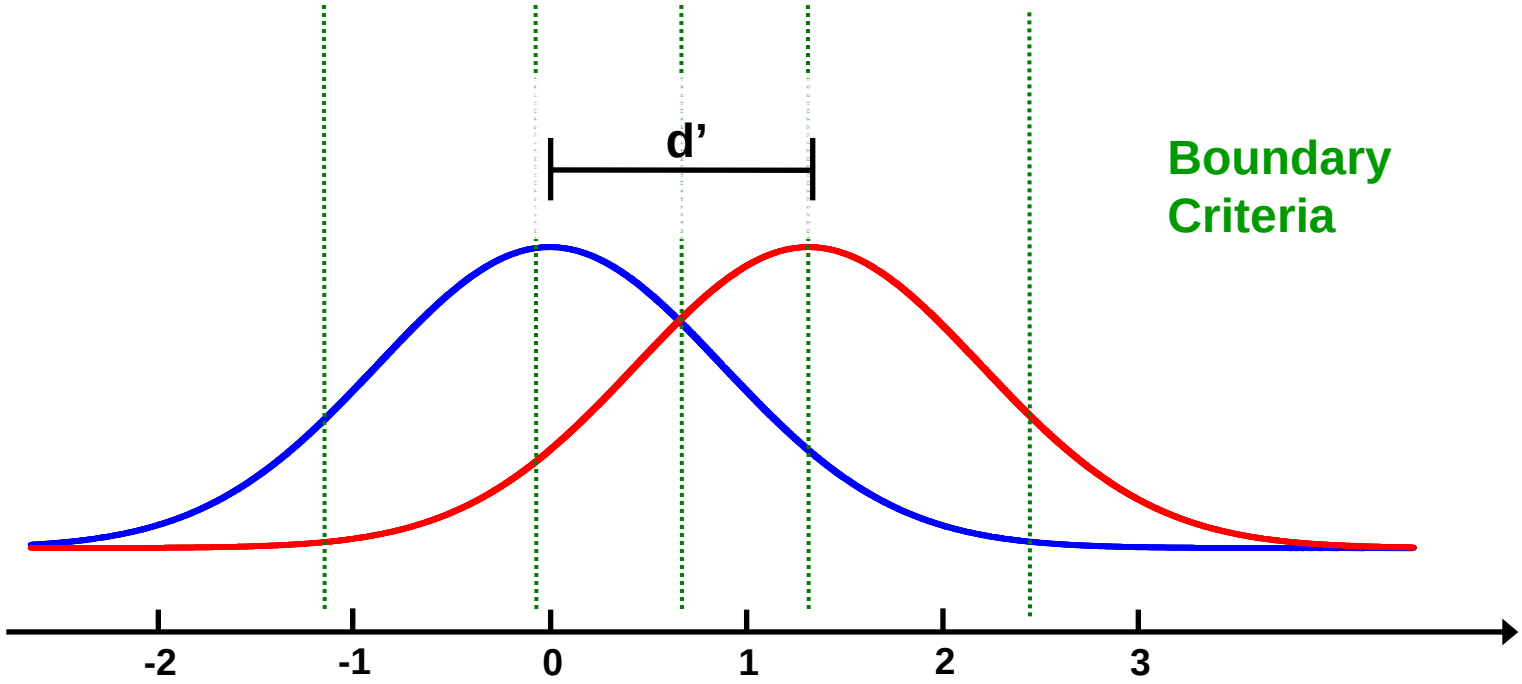
3. **R-Index with revised variance estimator**
   Bi J. & O'Mahony M. (2007) Updated and extended table for testing the significance of the R-Index Journal of Sensory Studies 22 (6) , 713-720

Aim:

- **Establish which is most appropriate / most powerful significance test for the R-Index**
- **Compare with Thurstonian analysis in terms of power**
- **Create power curves for difference testing and similarity testing**

Thurstonian Model

Product Identity	'A'		Response		'Not A'	
	Sure	Not Sure	Guess	Guess	Not Sure	Sure
Ref (A)	4	11	8	4	2	1
Test (Not A)	0	2	5	8	12	3



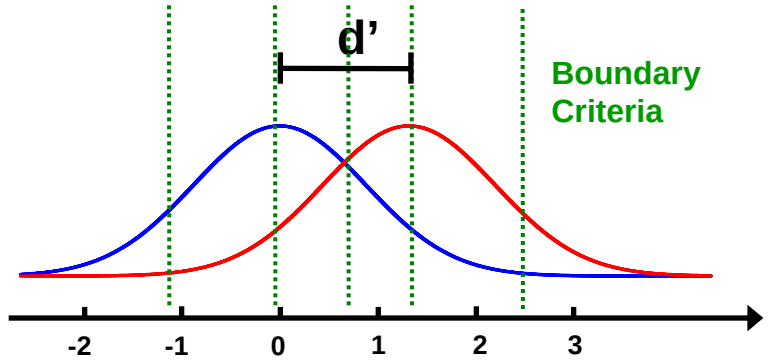
SAS Code Equal Variance

```
DATA DAT;
   INPUT PRODUCT SURENESS COUNT;
   LINES;
0 1  4
0 2 11
0 3  8
0 4  4
0 5  2
0 6  1
1 1  0
1 2  2
1 3  5
1 4  8
1 5 12
1 6  3
;
RUN;
```

```
PROC LOGISTIC DATA=DAT;
        WEIGHT COUNT;
        MODEL SURENESS=PRODUCT
                / AGGREGATE SCALE=NONE
                 LINK=PROBIT;
RUN;
```

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-1.1642	0.2832	16.9001	<.0001
Intercept	2	1	-0.0437	0.2175	0.0404	0.8407
Intercept	3	1	0.6879	0.2297	8.9681	0.0027
Intercept	4	1	1.3468	0.2595	26.9415	<.0001
Intercept	5	1	2.4484	0.3499	48.9596	<.0001
PRODUCT		1	-1.3401	0.2971	20.3437	<.0001

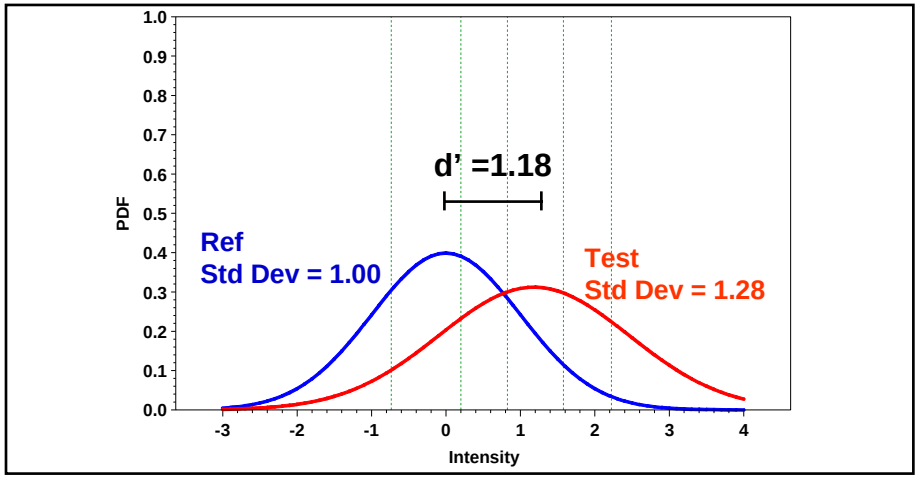


SAS Code Unequal Variance

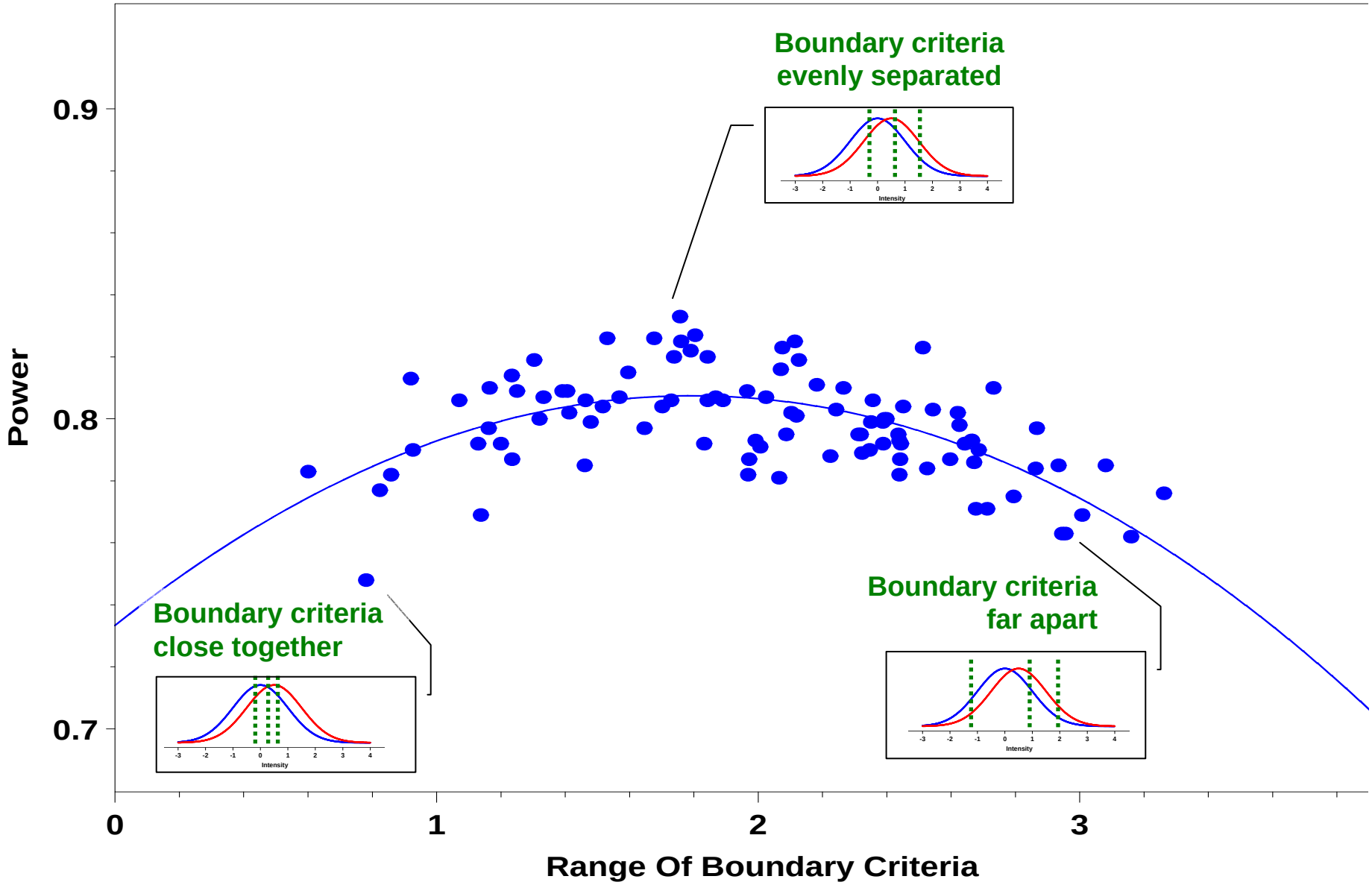
Obs	STIM	SURENESS	NR	NT	RP	NC
1	0	1	11	50	0.22	6
2	0	2	19	50	0.38	6
3	0	3	10	50	0.20	6
4	0	4	7	50	0.14	6
5	0	5	2	50	0.04	6
6	0	6	1	50	0.02	6
7	1	1	4	50	0.08	6
8	1	2	6	50	0.12	6
9	1	3	9	50	0.18	6
10	1	4	12	50	0.24	6
11	1	5	9	50	0.18	6
12	1	6	10	50	0.20	6

```
PROC NLIN DATA=DAT;
   PARMS C1=-2 D1=0.5 D2=0.5 D3=0.5
         D4=0.5 B=1.0 A=0.05;
   BOUNDS D1>0, D2>0, D3>0, D4>0;
   AX=EXP(A*STIM);
   IF SURENESS=1 THEN DO;
     ZJ=(C1-B*STIM)*AX;
     MODEL RP=PROBNORM(ZJ);
     END;
   IF SURENESS>1 AND SURENESS<NC
     THEN DO;
     IF SURENESS=2 THEN DO;
       KJ=C1+D1;
       K0=C1;
       END;
      IF SURENESS=3 THEN DO;
       KJ=C1+D1+D2;
       K0=C1+D1;
       END;
      IF SURENESS=4 THEN DO;
       KJ=C1+D1+D2+D3;
       K0=C1+D1+D2;
       END;
      IF SURENESS=5 THEN DO;
       KJ=C1+D1+D2+D3+D4;
       K0=C1+D1+D2+D3;
       END;
      ZJ=(KJ-B*STIM)*AX;
      Z0=(K0-B*STIM)*AX;
      PJ=PROBNORM(ZJ);
      P0=PROBNORM(Z0);
      MODEL RP=PJ-P0;
      END;
    IF SURENESS=6 THEN DO;
      KJ=C1+D1+D2+D3+D4;
      ZJ=(KJ-B*STIM)*AX;
      MODEL RP=1-PROBNORM(ZJ);
      END;
    _WEIGHT_=NT/MODEL.RP;
    DEV=-2*NR*LOG(MODEL.RP);
    IF RP>0 AND RP<1 THEN
       DEV=DEV+2*NR*LOG(RP);
    _LOSS_=DEV/_WEIGHT_;
    PR=MODEL.RP;
RUN;
```

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
C1	-0.7355	0.0884	-0.9627	-0.5083
D1	0.9346	0.0783	0.7334	1.1357
D2	0.6266	0.0641	0.4618	0.7914
D3	0.7520	0.0849	0.5337	0.9703
D4	0.6445	0.0988	0.3906	0.8983
B	1.1825	0.1281	0.8532	1.5117
A	-0.2451	0.0899	-0.4763	-0.0139

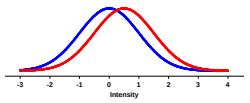


Power Depends On Scale Usage

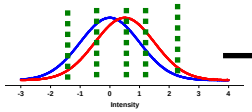


Power Determination: Simulation

Select 'true' underlying product difference



Simulate set of boundary criteria



Simulate data set of n replicates of each product based on Thurstonian model

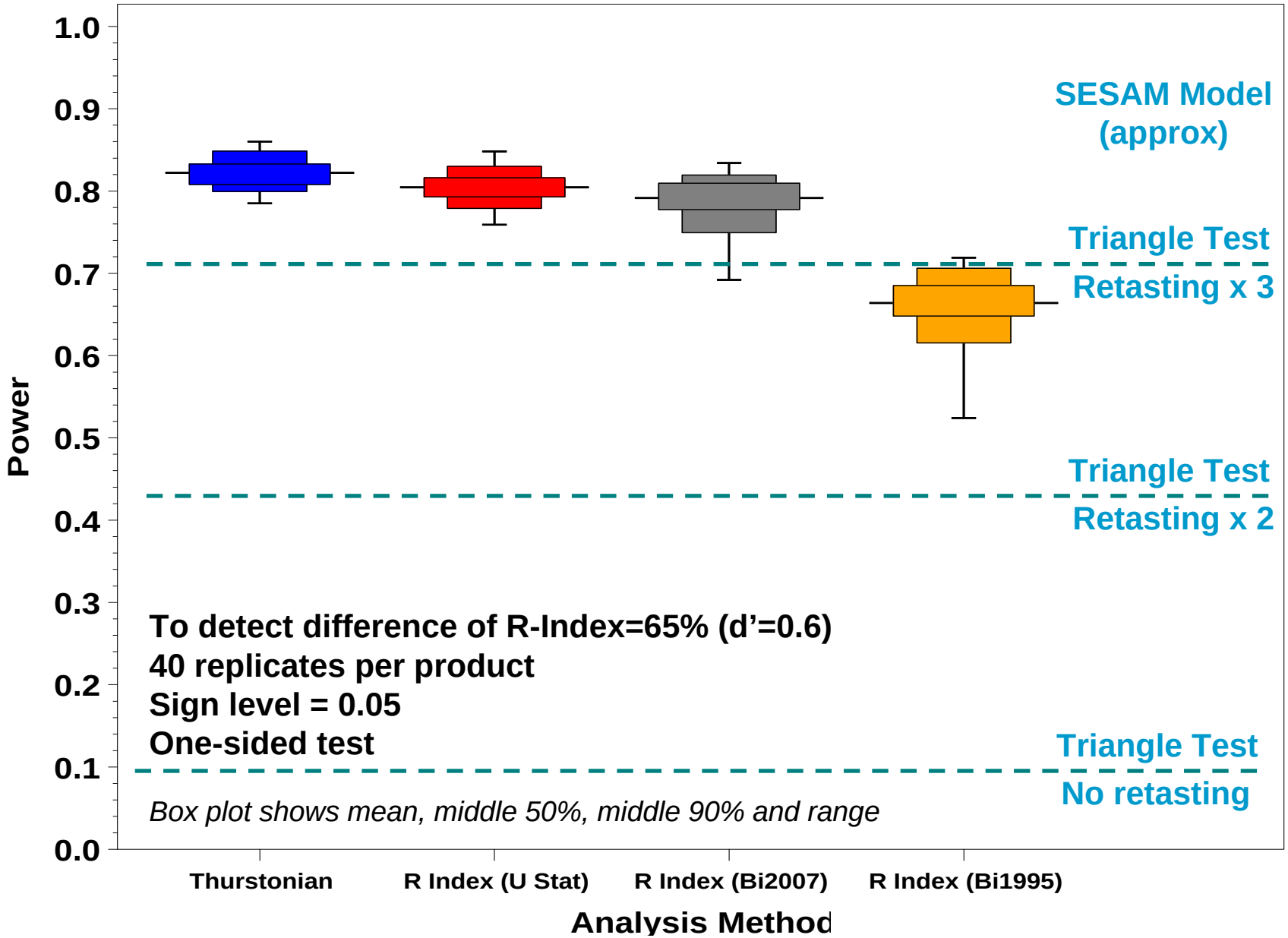
Thurstonian Analysis d'	R-Index Analysis 1	R-Index Analysis 2	R-Index Analysis 3
Significant?	Significant?	Significant?	Significant?
Power	Power	Power	Power
Distribution of Power	Distribution of Power	Distribution of Power	Distribution of Power
Power Curve	Power Curve	Power Curve	Power Curve

Repeat x1000

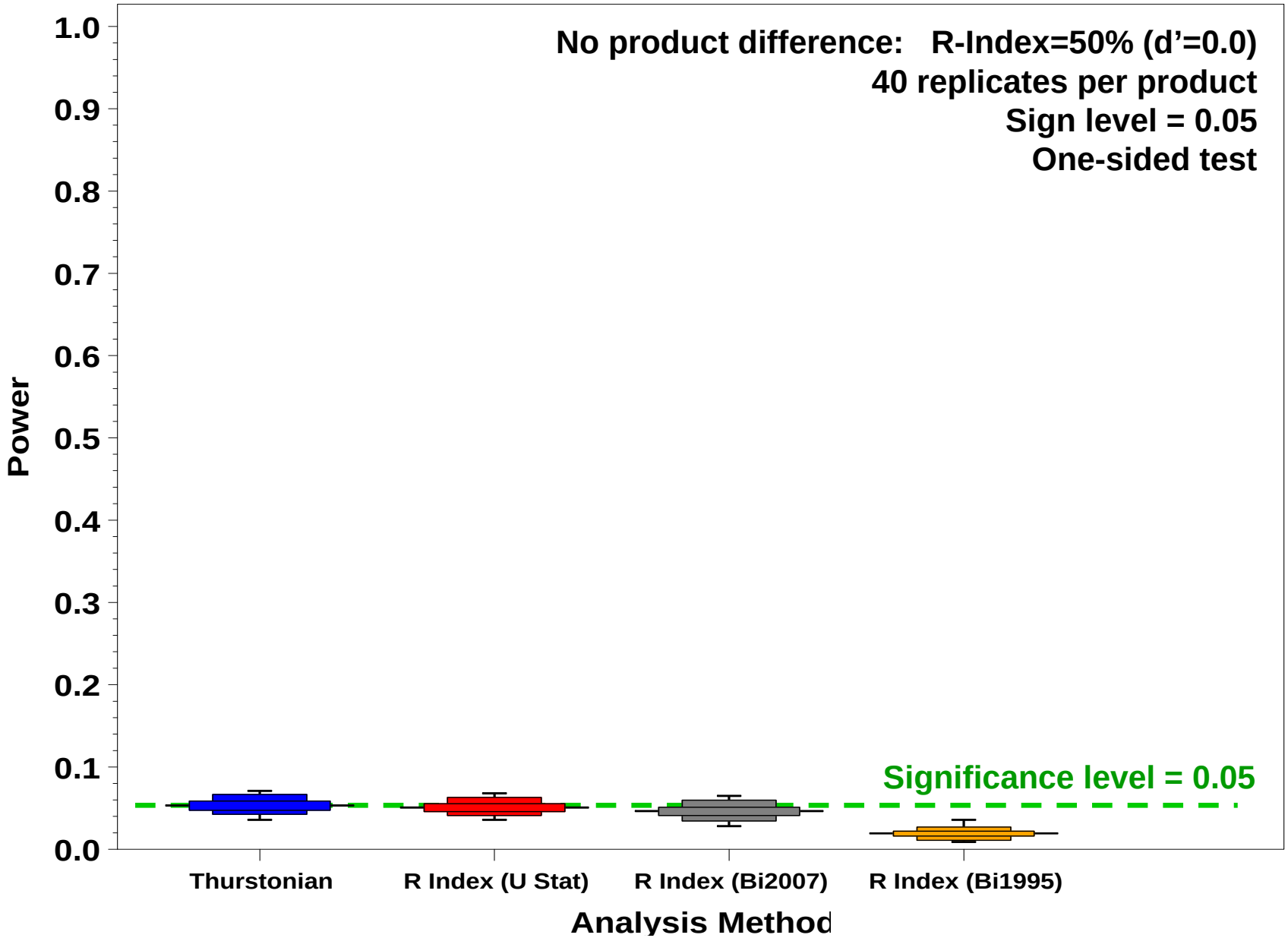
Repeat x100

Repeat Varying True Difference and Varying Replication n

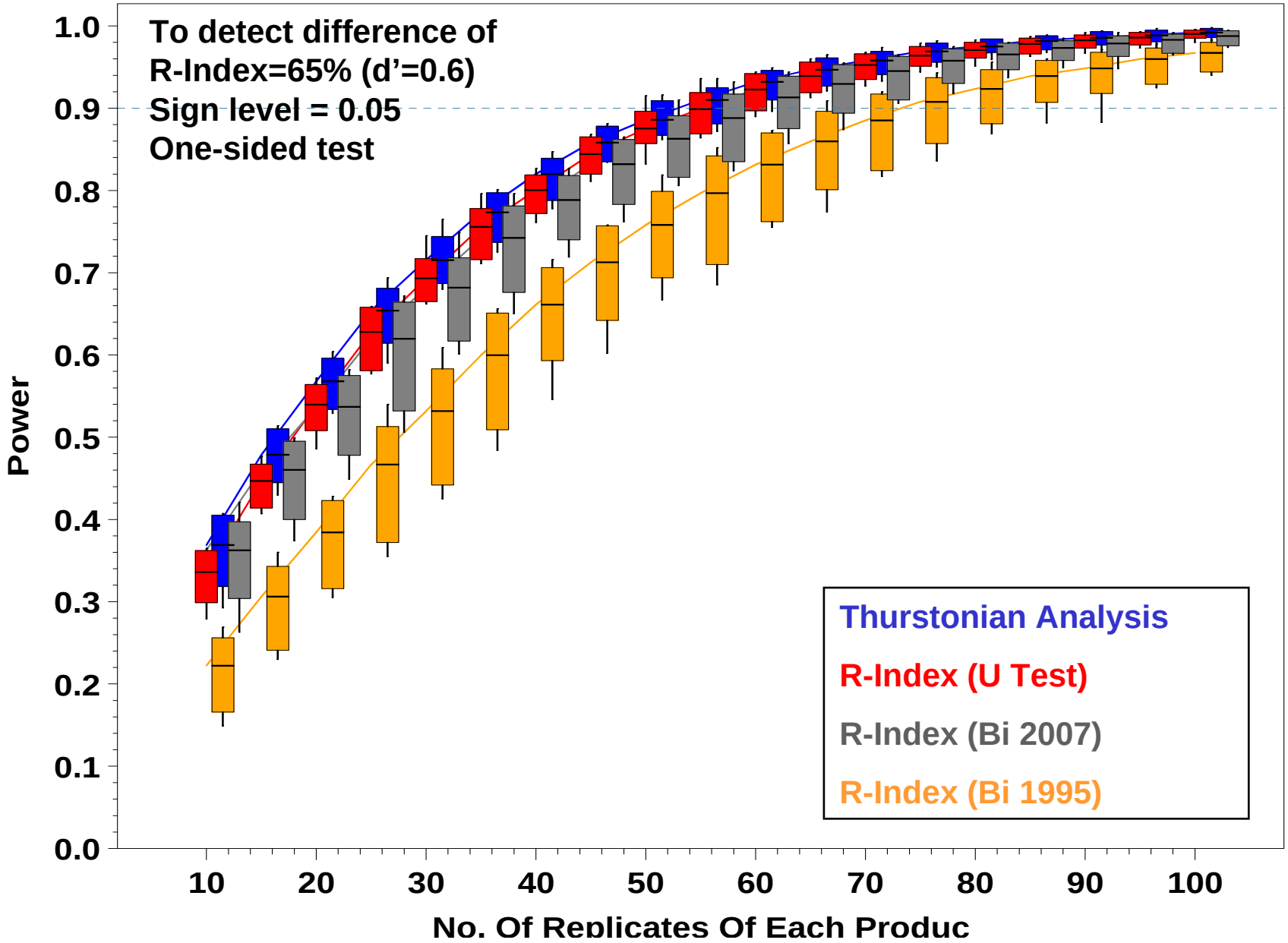
Power Comparison



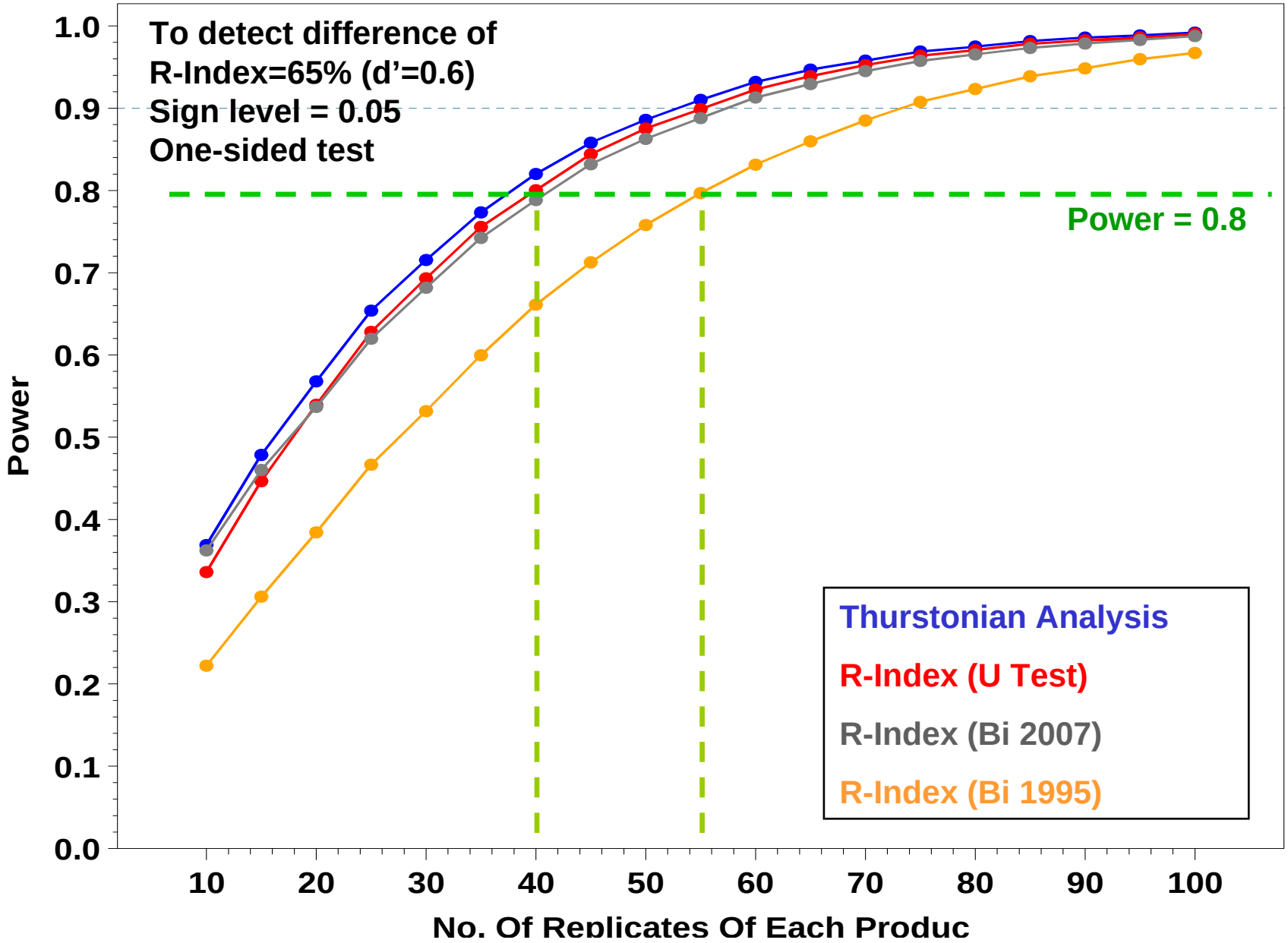
Power Comparison



Power Curve Comparison



Power Curve Comparison



Study Objective

Product Improvement

Sensory Differentiation vs Competition

Difference test

Aim
Demonstrate that the product difference is significant

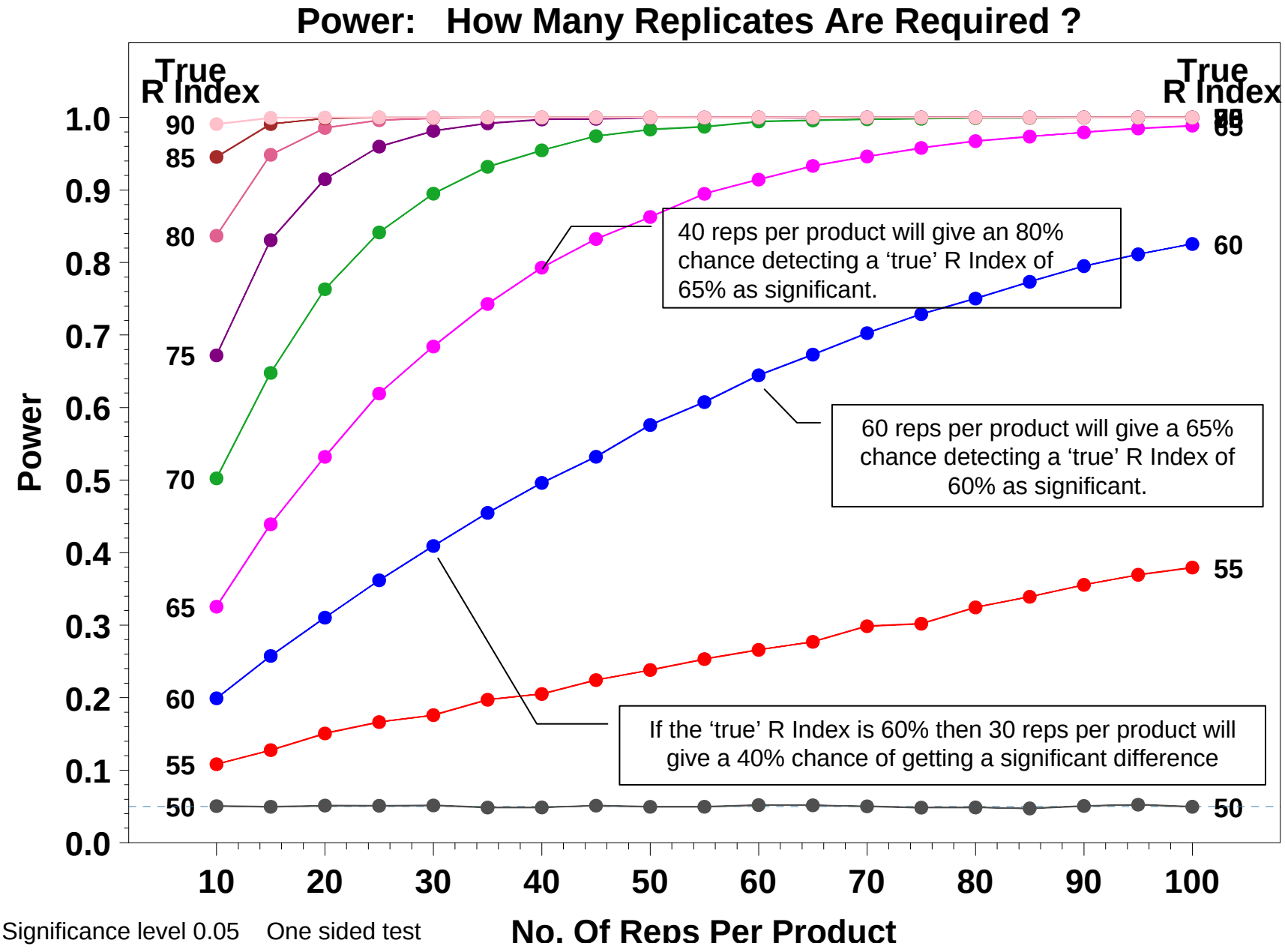
Changing the formulation for cost-savings

Changing suppliers of raw materials

Similarity test

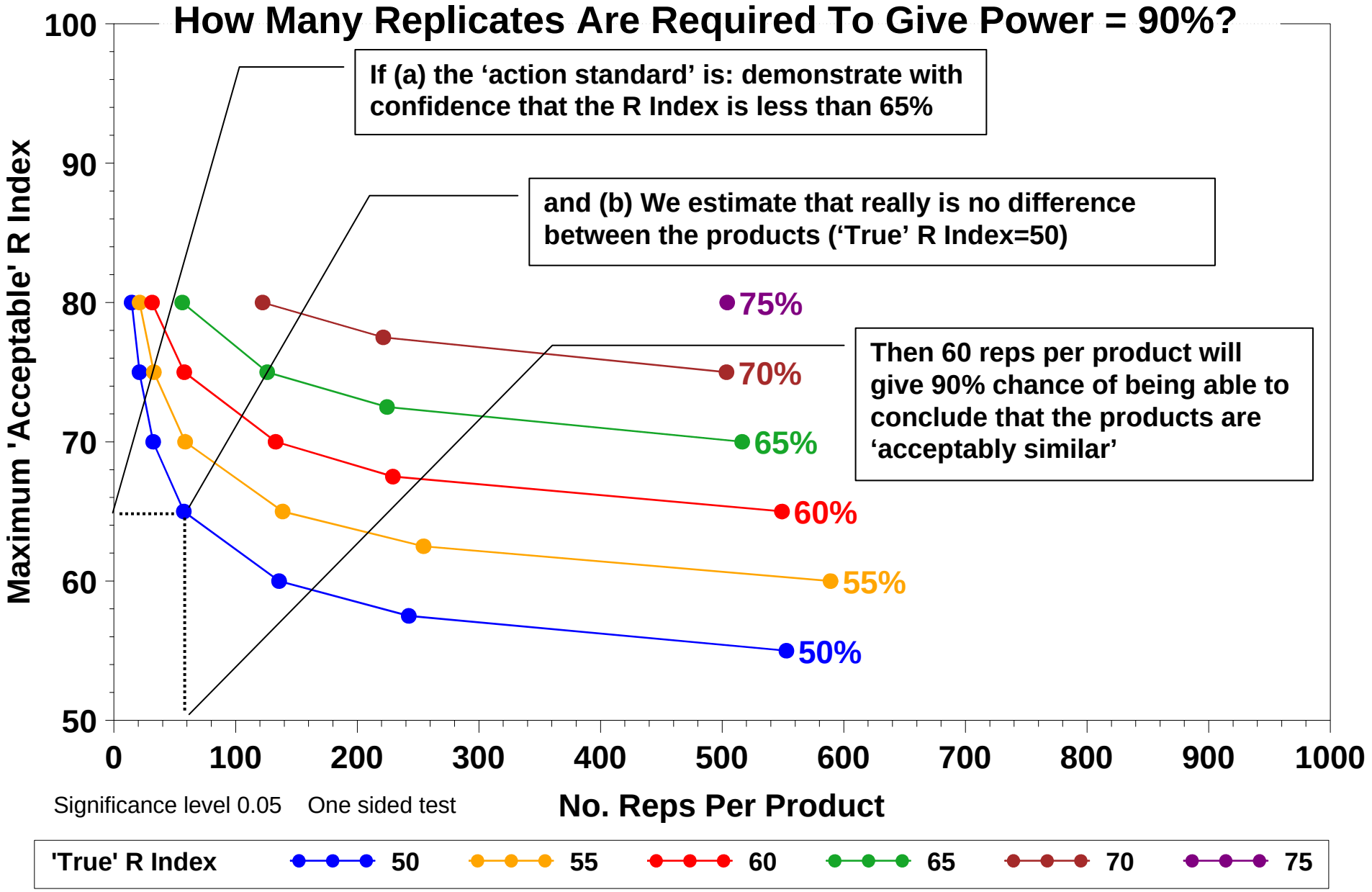
Aim
Demonstrate, with confidence, that the product difference is less than pre-specified value

Power: Difference Test

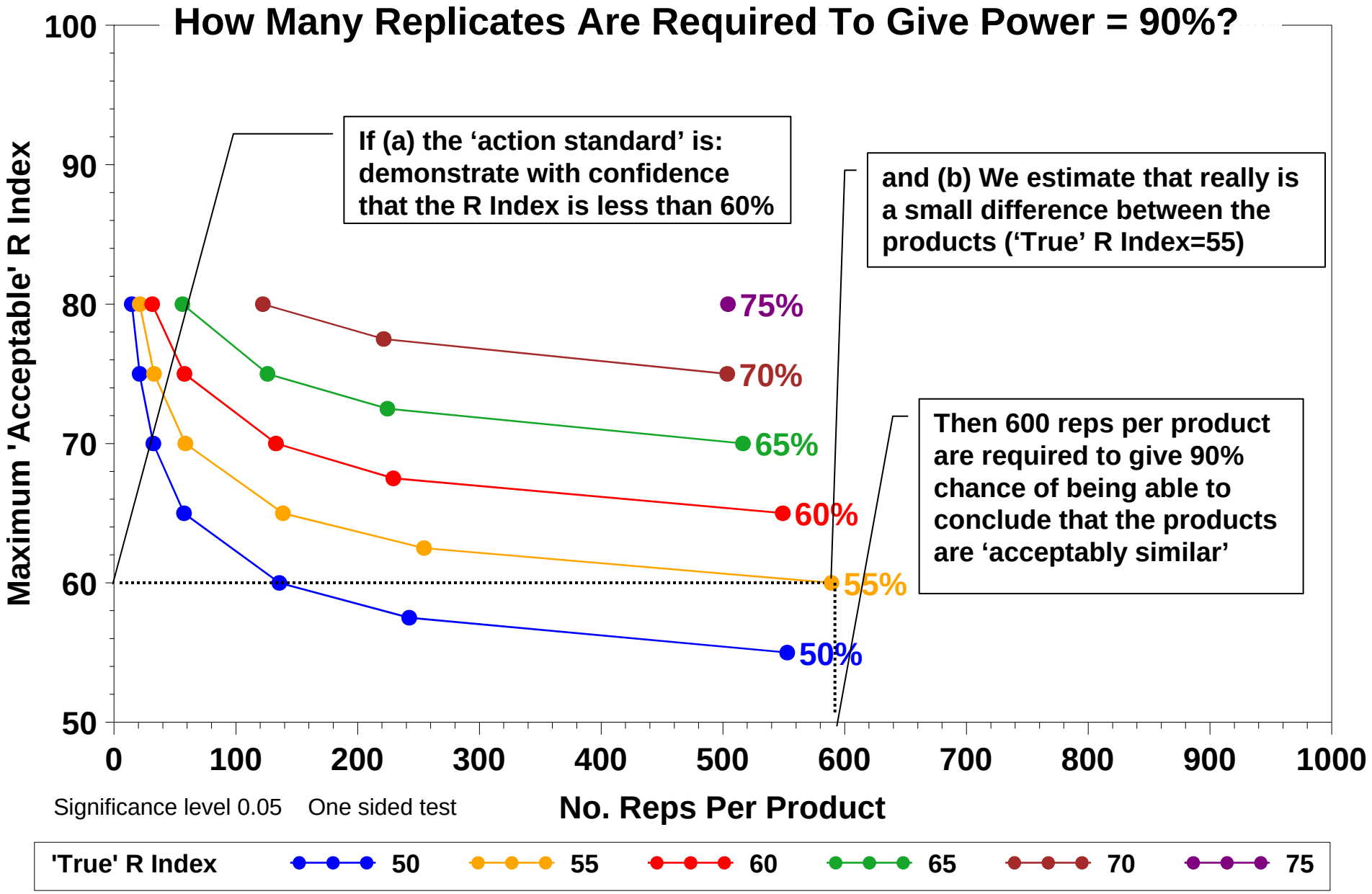


Significance level 0.05 One sided test

Power: Similarity Test



Power: Similarity Test



Individual vs Pooled Analysis

Subject		Same as Reference: Sure	Same as Reference: Not Sure	Same as Reference: Guess	Different to Reference: Guess	Different to Reference: Not Sure	Different to Reference: Sure	Total	R Index	Std Err
1	Ref	0	4	1	3	2	0	10	78.5	12.6
	Test	0	0	0	5	4	1	10		
2	Ref	0	0	4	3	2	1	10	76.5	12.7
	Test	0	1	0	0	6	3	10		
3	Ref	1	5	3	1	0	0	10	84.0	12.6
	Test	0	0	6	2	2	0	10		
4	Ref	3	5	0	0	2	0	10	77.0	12.9
	Test	0	2	5	2	0	1	10		
5	Ref	0	1	3	5	1	0	10	84.0	12.6
	Test	0	0	0	4	4	2	10		

Mean = 80.0 SEM = 5.67

z = 5.29 ***

Pooled	Ref	4	15	11	12	7	1	50	73.9	5.7
	Test	0	3	11	13	16	7	50		

z = 4.21 ***

Alternative Analyses

Pooled Analysis

- *Most suited to analysis of one individual, pooling over replicate assessments*
- *Or where it can be safely assumed that individual assessors are the same and using the scale in the same way*

Individual Analysis For Each Assessor Before Pooling

- *Protects against bias caused by differences in usage of scale*
- *Requires some replication of each product per assessor*
- *Initial results indicate 4+ reps per assessor*

New Directions

- *Aim: Overall analysis that allows for individual differences in sensitivity and boundary criteria, and influences of other factors*
  Poster P13: A statistical model for A-Not A data with and without sureness. R.H.B. Christensen G. J. Cleaver and P. B. Brockhoff

Summary

R-Index

- *Important to use the correct analysis for significance test*
- *Recommended: Test based on U-Statistic with ties*
- *Or test in Bi (2007) if table of critical values required*

Thurstonian Analysis

- *Based on assumption about underlying nature of perception*
- *More insights on perceptual difference and scoring*
- *Similar power to R-Index analysis*

Power Charts

- *To be used according to objective of study*
  - ** Difference Test*
  - ** Similarity Test*

New Directions

- *Potential bias with simple pooling of data*
- *Analysis based on individual R-Index values safer*
- *New Generalised NL Model – see Poster (Christensen et al)*