JSM2005 August 7-11 Minneapolis — Using Our Discipline to Enhance Human Welfare

DataMineIt℠

A Nonparametric Statistic for Joint Mean-Variance Quality Control

J.D. Opdyke, DataMineIt

JSM Activity #29: New Research in Quality Control and Monitoring – Contributed - Papers

Contents

1. The Quality Control Problem:

- One-sided, two-sample hypothesis test of

$$H_0: \mu_s \leq \mu_c \text{ and } \sigma_s^2 \leq \sigma_c^2$$

vs.

$$H_A: \mu_s > \mu_c \text{ or } \sigma_s^2 > \sigma_c^2$$

- SPC research has been addressing joint location-scale effects more than has research on quality control-related hypothesis testing

2. Previous & Related Work

2.a) the ‘modified’ t (Brownie et al. 1990)

- 7 years of expert testimony in the regulatory telecommunications arena (implementing the Telecom Act of 1996) supported its use for testing the defined joint hypotheses to compare quality of service provided by Bell companies to CLEC vs. ILEC customers
- Opdyke (2004) showed that the 3 statistical claims supporting its use – i) asymptotic standard normality, ii) minimizing the possibility of ‘gaming’ (manipulation), and iii) appropriateness for the defined joint hypotheses – are false

2.a.i)

$$t_{mod} \sim N\left(0, \frac{\dfrac{\sigma^2_{study}}{n_{study}} + \dfrac{\sigma^2_{control}}{n_{control}}}{\dfrac{\sigma^2_{control}}{n_{study}} + \dfrac{\sigma^2_{control}}{n_{control}}} \right) \text{ asymptotically}$$

- **NOT standard normal – relative size of population variances determines t_{mod} variance:**
  - **if $\sigma^2_{study} / \sigma^2_{control} > 1$, $Var(t_{mod}) > 1$**
  - **if $\sigma^2_{study} / \sigma^2_{control} < 1$, $Var(t_{mod}) < 1$**
  - **if $\sigma^2_{study} / \sigma^2_{control} = 1$, $Var(t_{mod}) = 1$**

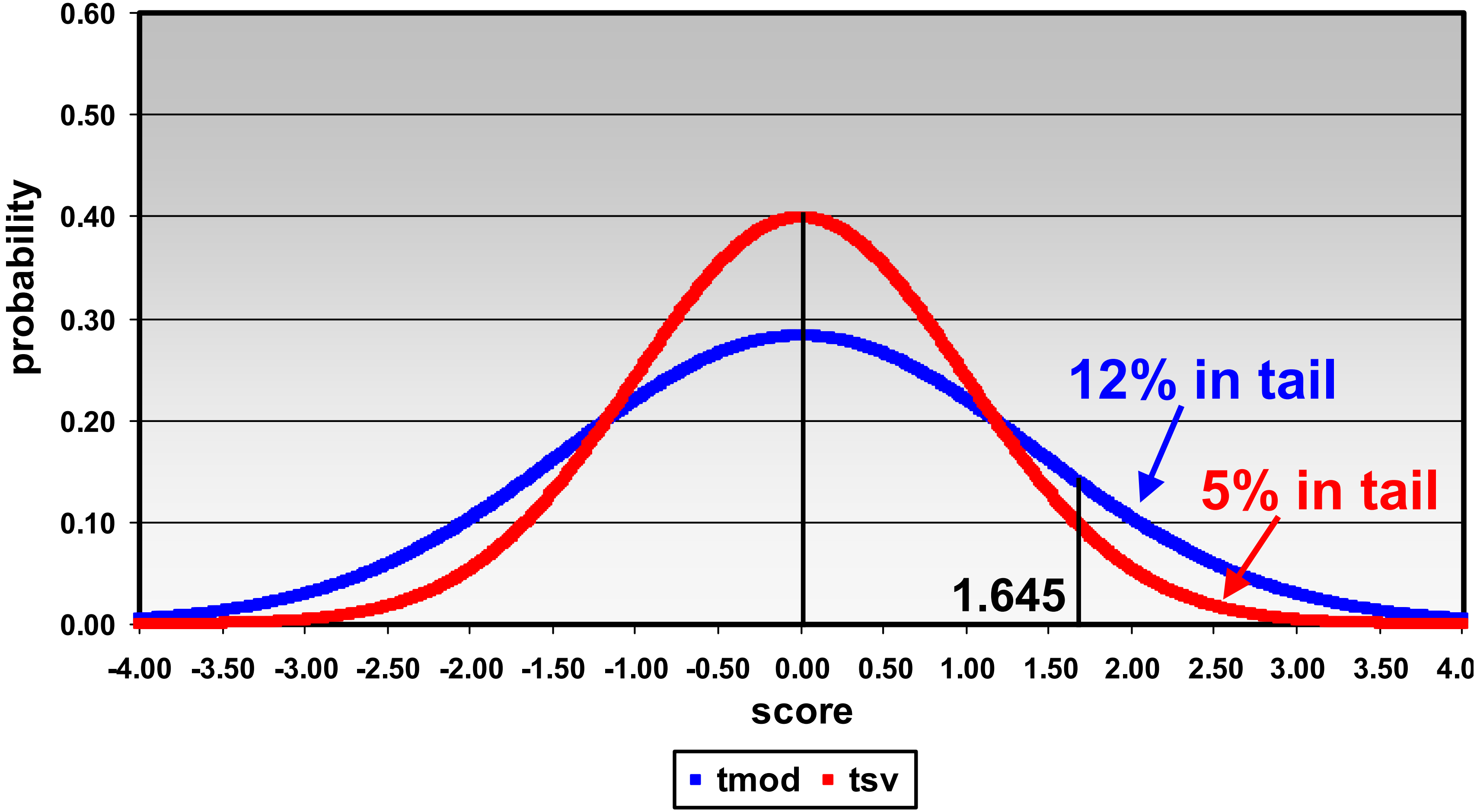
JSM2005 August 7-11 Minneapolis
Using Our Discipline to Enhance Human Welfare

DataMineIt[SM]
©J.D. Opdyke

GRAPH 1

t_{mod} vs. t_{sv}, with $\sigma^2_{study} / \sigma^2_{control} = 2.0$

($n_{control}$ = 10,000, n_{study} = 100)



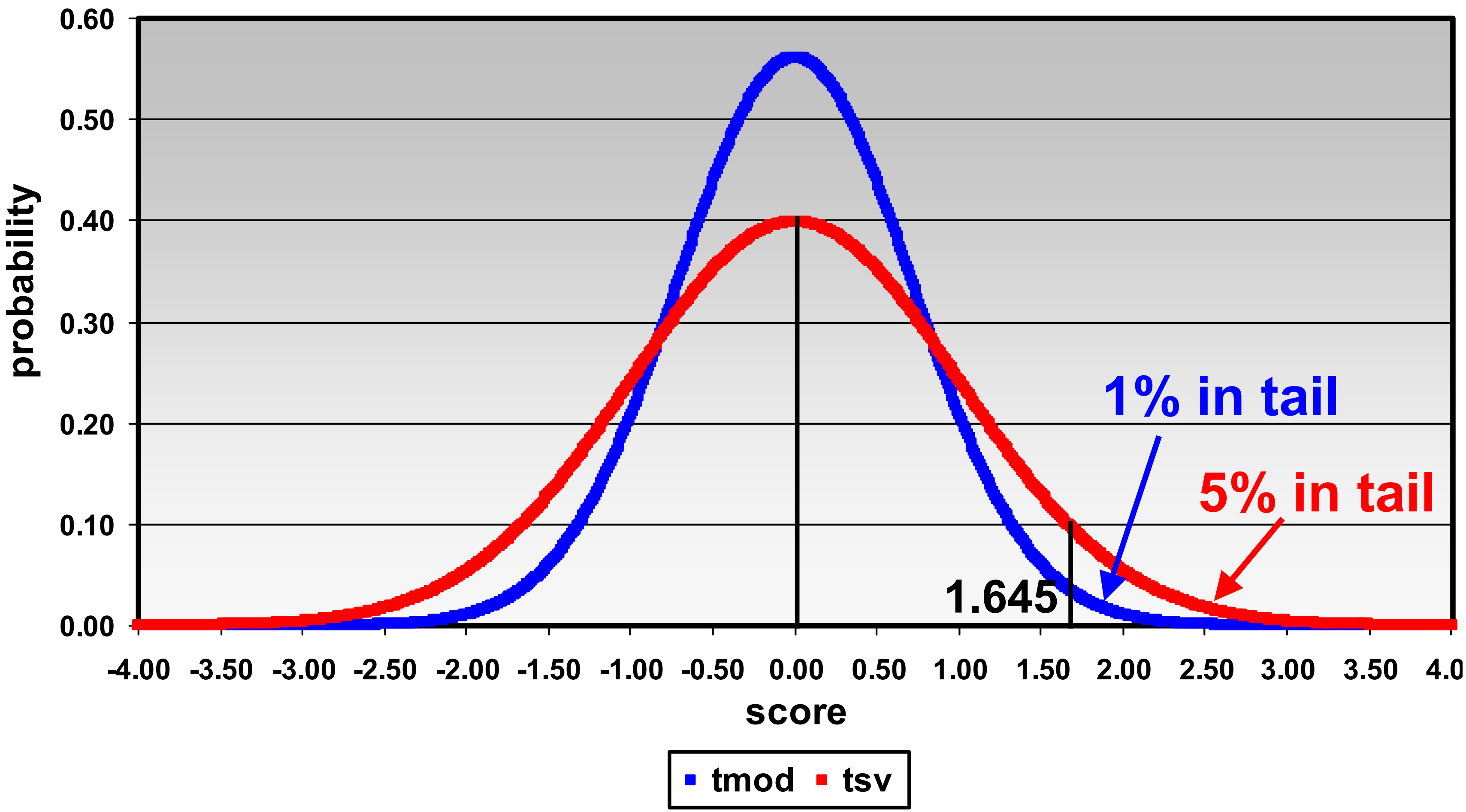
Minneapolis JSM2005 August 7-11
Using Our Discipline to Enhance Human Welfare

DataMineIt[SM]
©J.D. Opdyke

GRAPH 2

t_{mod} vs. t_{sv}, with σ^2_{study} / $\sigma^2_{control}$ = 0.5

($n_{control}$ = 10,000, n_{study} = 100)

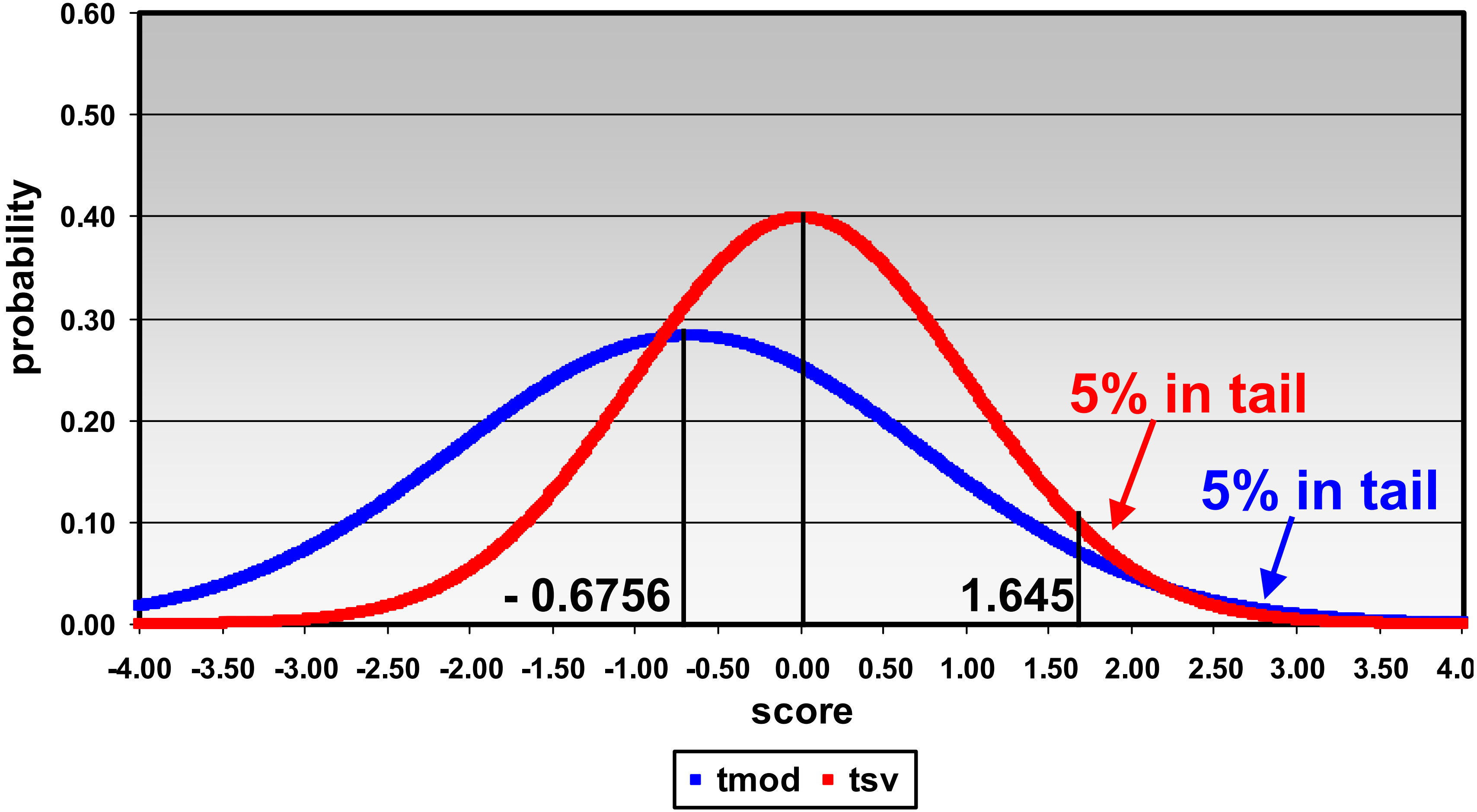


- Consequences for t_{mod} tail probabilities (when using fixed critical value from standard normal table, as advised):
  - if $\sigma^2_{study} / \sigma^2_{control} > 1$, $Pr(t_{Mod} > z_{crit_\alpha}) > \alpha$
  - if $\sigma^2_{study} / \sigma^2_{control} < 1$, $Pr(t_{Mod} > z_{crit_\alpha}) < \alpha$
  - if $\sigma^2_{study} / \sigma^2_{control} = 1$, $Pr(t_{Mod} > z_{crit_\alpha}) = \alpha$
- This allows a “trade-off” of better/worse average performance for worse/better variability with Pr(reject Ho | Ha) = constant

GRAPH 3

t_{mod} shifted vs. t_{sv}, with σ^2_{study} / $\sigma^2_{control}$ = 2.0

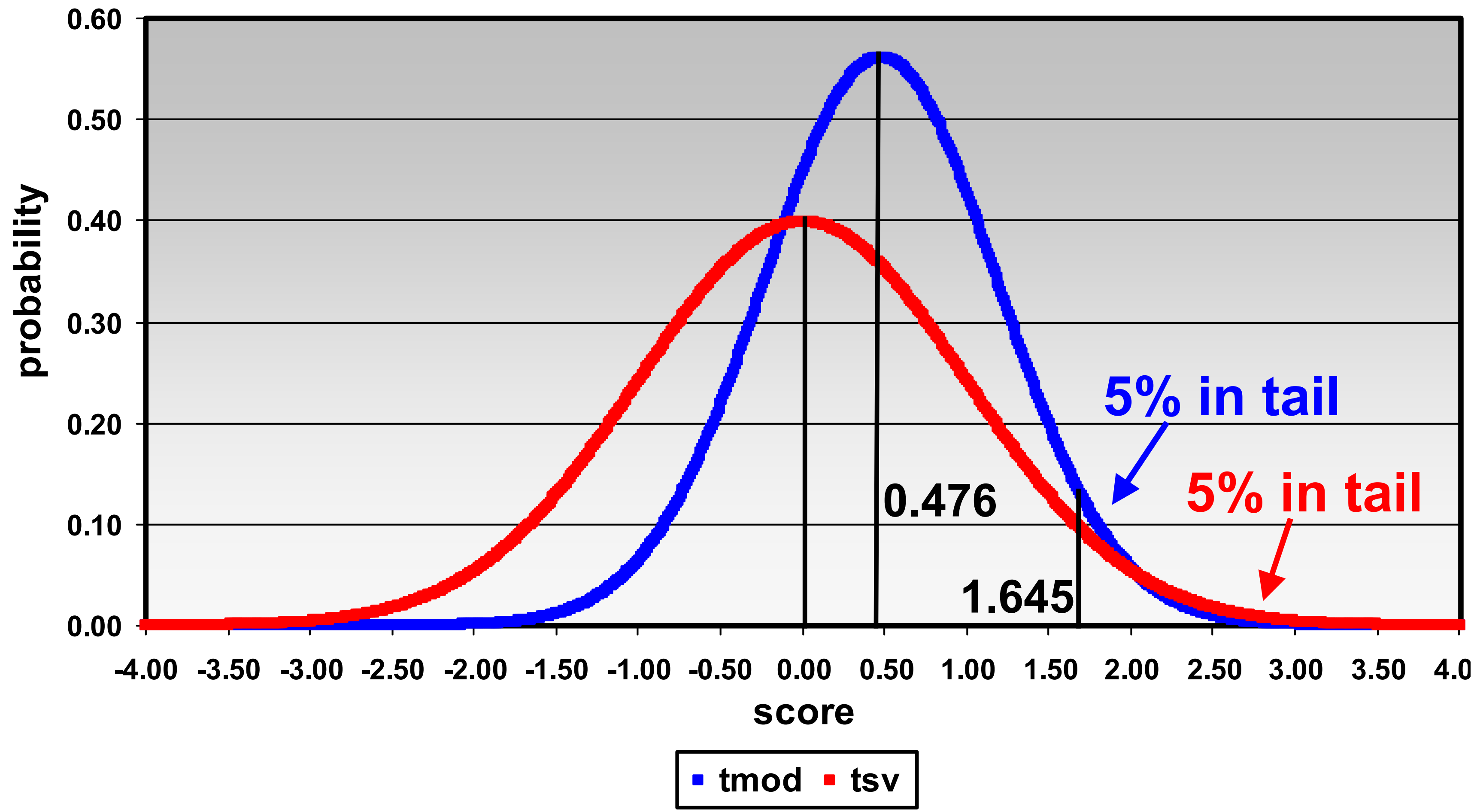
($n_{control}$ = 10,000, n_{study} = 100)



GRAPH 4

t_{mod} shifted vs. t_{sv}, with $\sigma^2_{study} / \sigma^2_{control} = 0.5$

($n_{control} = 10{,}000$, $n_{study} = 100$)



2.a.ii) Therefore, t_{mod} could actually *encourage* 'gaming' by allowing worse average service with greater precision!

2.a.iii) Inappropriate for the defined joint hypotheses because Graph 1 ⇒ very low asymptotic power for variance violations, & Graphs 3-4 ⇒ NO power! – the "trade-off" masks the Ha:

<u>NOTE</u>: Brownie et al. (1990) proposed t_{mod} only for concurrent increases in BOTH mean and variance, NOT increases in one, the other, OR both ("AND" in Ha: instead of "OR" in Ha:)

2.b) Bűning and Thadewald (2000) present an asymptotically X^2, adaptive Lepage-type test (BT) of

$$G(z) = F\left(\left[z-\theta\right]/\tau\right)$$ **(location and/or scale shifts only)**

for symmetric, and even slightly asymmetric, distributions.

- **BT is two-sided, but it is adapted in Opdyke (2004) for the defined, one-sided joint hypotheses**
- **BT is not as powerful as OBMax2 under almost all of the tested data conditions**

2.c) Tests of Stochastic Dominance:

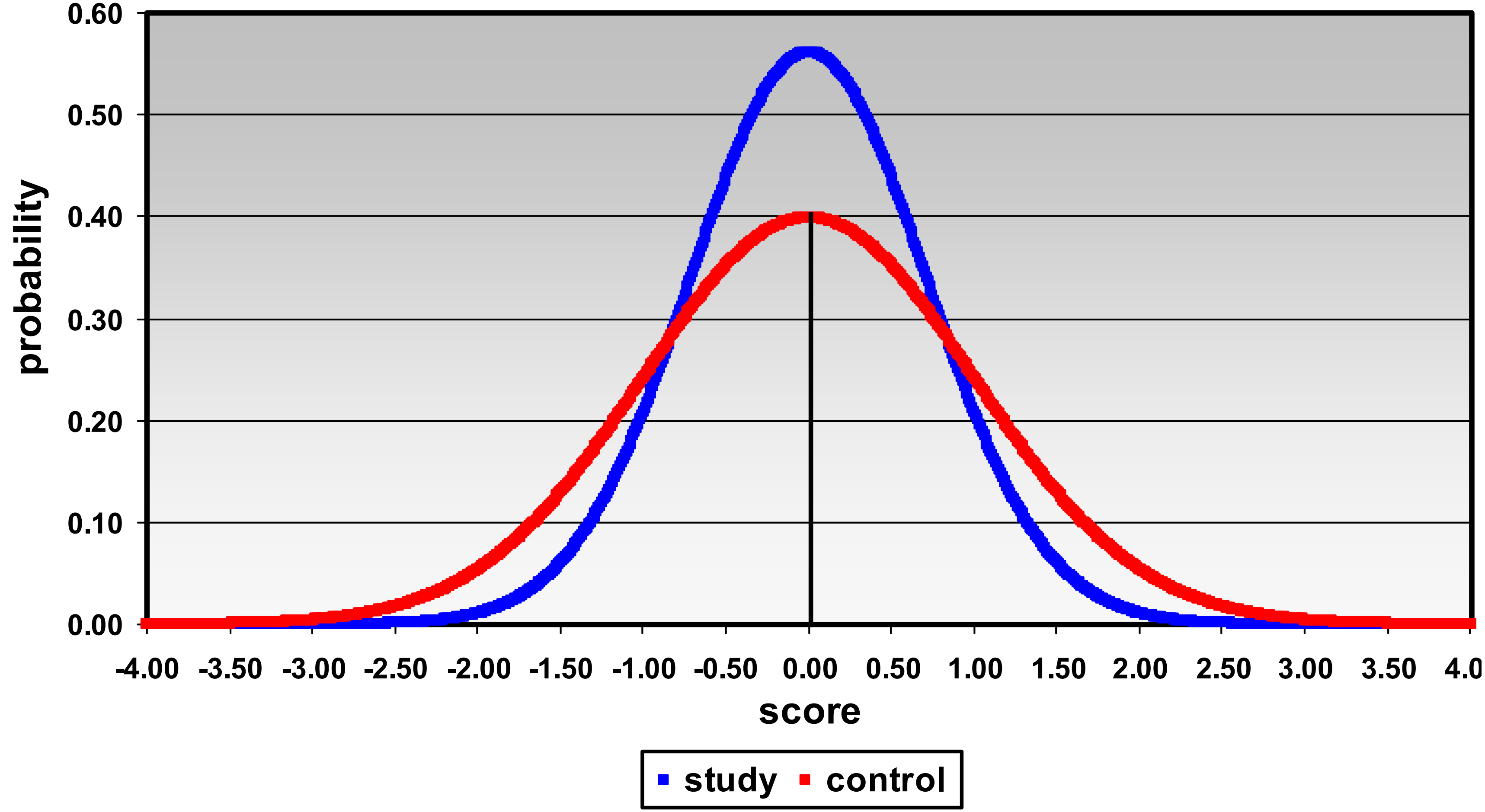
- **Rosenbaum (1954) was shown in Opdyke (2004) to have very little power for these joint hypotheses**

- **Kolmogorov-Smirnov was shown in Opdyke (2004) to sometimes have more power than OBMax2's constituent statistics, but KS severely violates the nominal level when means are equal and the study-group variance is SMALLER**

JSM2005 August 7-11 Minneapolis
Using Our Discipline to Enhance Human Welfare

DataMineIt℠
©J.D. Opdyke

GRAPH 5
KS violates nominal level under Ho: below, which is consistent with its design



- Reiczigel et al. (2005) propose a bootstrap of a rank Welch test (BRW): not yet compared to OBMax2 due to its computational intensity (a drawback for non-small samples)

2.d) Pesarin's (2000) joint location-scale permutation test: not yet compared to OBMax2 due to its computational intensity, but indications from Opdyke (2004) are that the "trade-off" nature of the p-value combining functions used will make it less powerful than OBMax2, at least for non-small samples.

2.e) Opdyke (2004, 2005):

- **Opdyke (2004) proposed choosing between several conditional statistical procedures based on the relative sample sizes and kurtoses of the (symmetric) data being tested**

Sample Size		Normal and Platykurtotic	Leptokurtotic	Skewed
		OBt	OBG	
Balanced	Shoemaker	**OBtShoe**	**OBGShoe**	transform
Unbalanced	Levene	**OBtLev**	**OBGLev**	transform

where…

Conditional Statistical Procedure	if $\bar{X}_s > \bar{X}_c$, use...	if $\bar{X}_s \le \bar{X}_c$, use...
OBtShoe	OBt	Shoemaker's F_1
OBtLev	OBt	'modified' Levene
OBGShoe	OBG	Shoemaker's F_1
OBGLev	OBG	'modified' Levene

OBt = O'Brien's (1988) generalized t test (a simple conditional OLS regression)

OBG = O'Brien's (1988) Generalized rank sum test (OBt on pooled ranks)

F_1 = Shoemaker's (2003) variance ratio test, ~ F

Levene = Levene's (1960) one-way ANOVA variance test modified as per Brown & Forsythe (1974) ~ F

- **Opdyke (2005) combined these four tests into a single statistic via a maximum test (or minimum p-value)**

$$p_{OBMax} = \min\begin{pmatrix} \beta_{OBtShoe} \cdot p_{OBtShoe}, \\ \beta_{OBGShoe} \cdot p_{OBGShoe}, \\ \beta_{OBtLev} \cdot p_{OBtLev}, \\ \beta_{OBGLev} \cdot p_{OBGLev}, \\ \beta_{tsv} \cdot p_{tsv}, \\ 1.0 \end{pmatrix} \quad \text{where } \beta_{OBtShoe} = 2.8,\ \beta_{OBGShoe} = 2.8,\ \beta_{OBtLev} = 2.8,\ \beta_{OBGLev} = 2.8,\ \&\ \beta_{tsv} = 1.8$$

and

$$p_{OBMax3} = \min\begin{pmatrix} \beta_{OBtShoe} \cdot p_{OBtShoe}, \\ \beta_{OBtLev} \cdot p_{OBtLev}, \\ \beta_{tsv} \cdot p_{tsv}, \\ 1.0 \end{pmatrix}$$

$$\text{where } \beta_{OBtShoe} = \beta_{OBtLev} = 3.2, \text{ and } \beta_{tsv} = 1.6$$

- **The betas account for inflated levels due to the conditional nature of the combined test (see O'Gorman (1997) for a similar adjustment)**
- **Use OBMax under symmetry, and OBMax3 if symmetry cannot be assured, because OBMax violates the test level under a combination of a) large, equal sample sizes, and b) skewed data, *and* c) a smaller study sample variance (the OBG rank-based tests do not perform well under these conditions).**

3. Proposed Statistic: OBMax2

- **This current work further develops Opdyke (2005) by combining OBMax and OBMax3 into OBMax2 – a single statistic under skewed or symmetric data. If $s_s^2 \le s_c^2$ & $\left(\bar{X}_s - 0.5 s_c\right) \le \bar{X}_c$, and either sample is skewed as per the test of D'Agostino et al. (1990), OBMax2 = OBMax3; otherwise, OBMax2 = OBMax.**

- **Also, OBMax2 accounts for NStudy > NControl, which is not relevant in regulatory telecom since NClec << NIlec virtually always.**

$$\text{If } n_s > n_c, \text{ then } \beta_X = \beta_X + \min\left[2.5, \log_{2.7}\left(n_s / n_c\right)\right]$$

4. Monte Carlo Simulations

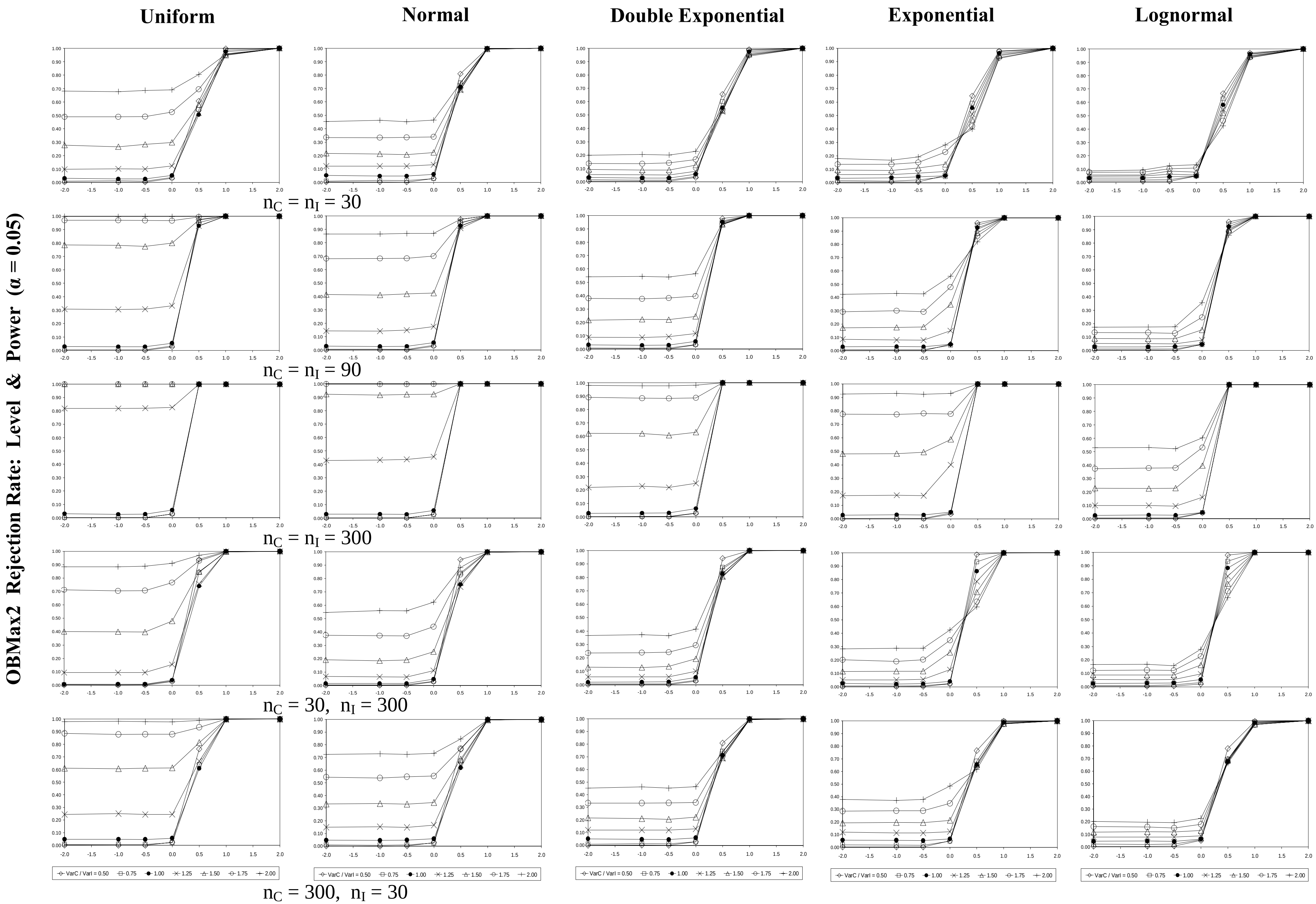
- **5 distributions:**
  Double Exponential, Exponential, Lognormal, Normal, Uniform
- **5 sample size pairings: 30-30, 90-90, 300-300, 30-300, 300-30**
- **49 mean-variance configurations:**

$$\mu_s = \mu_c - 2.0\sigma_c,\ \mu_c - 1.0\sigma_c,\ \mu_c - 0.5\sigma_c,\ \mu_c,\ \mu_c + 0.5\sigma_c,\ \mu_c + 1.0\sigma_c,\ \mu_c + 2.0\sigma_c$$

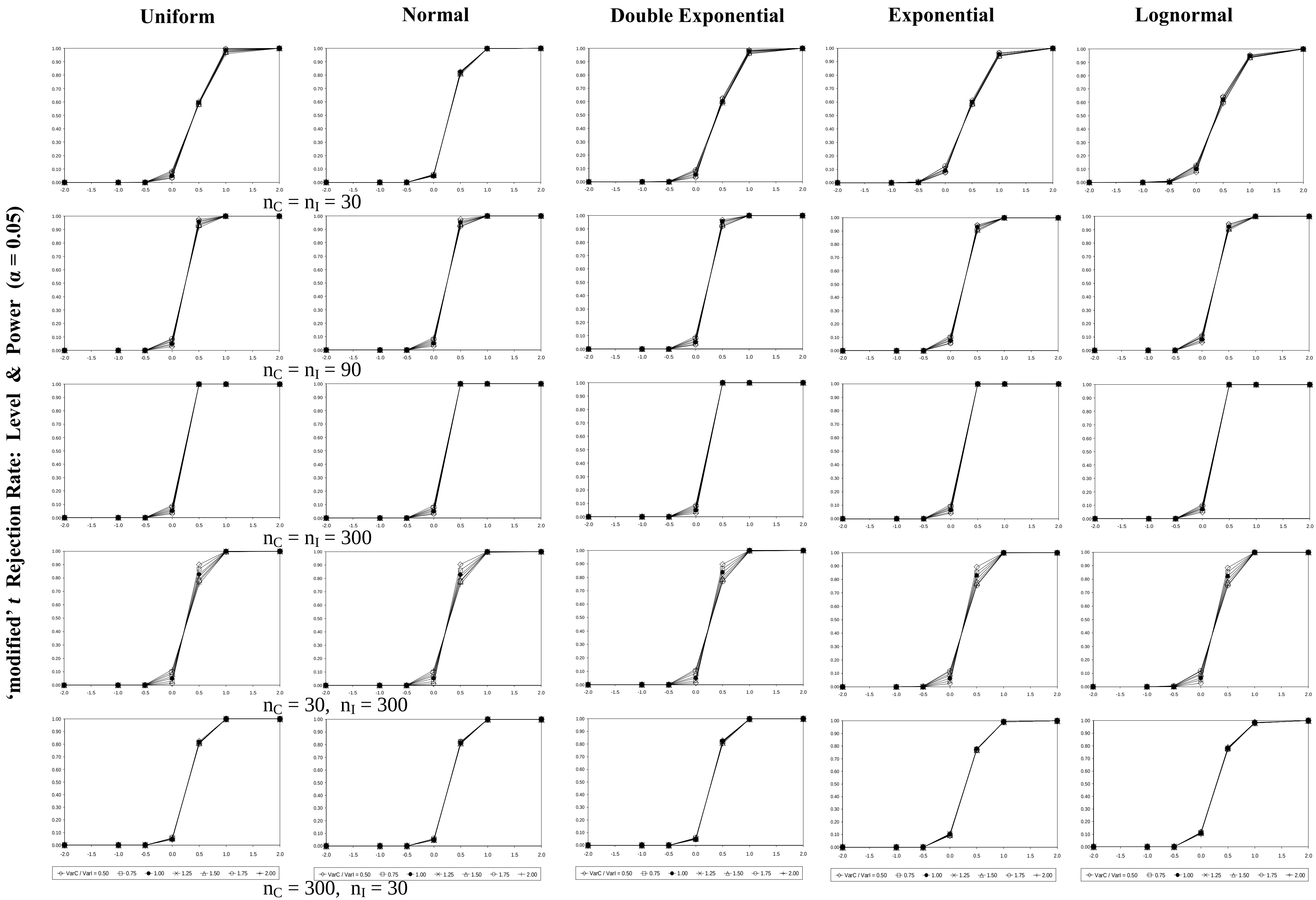
$$\sigma_s^2 / \sigma_c^2 = 0.5,\ 0.75,\ 1.00,\ 1.25,\ 1.50,\ 1.75,\ 2.00$$

- **2 levels: α = 0.05, α = 0.10**
- **TOTAL: 2,450 scenarios – 600 null, 1,850 alternate**

5. Results

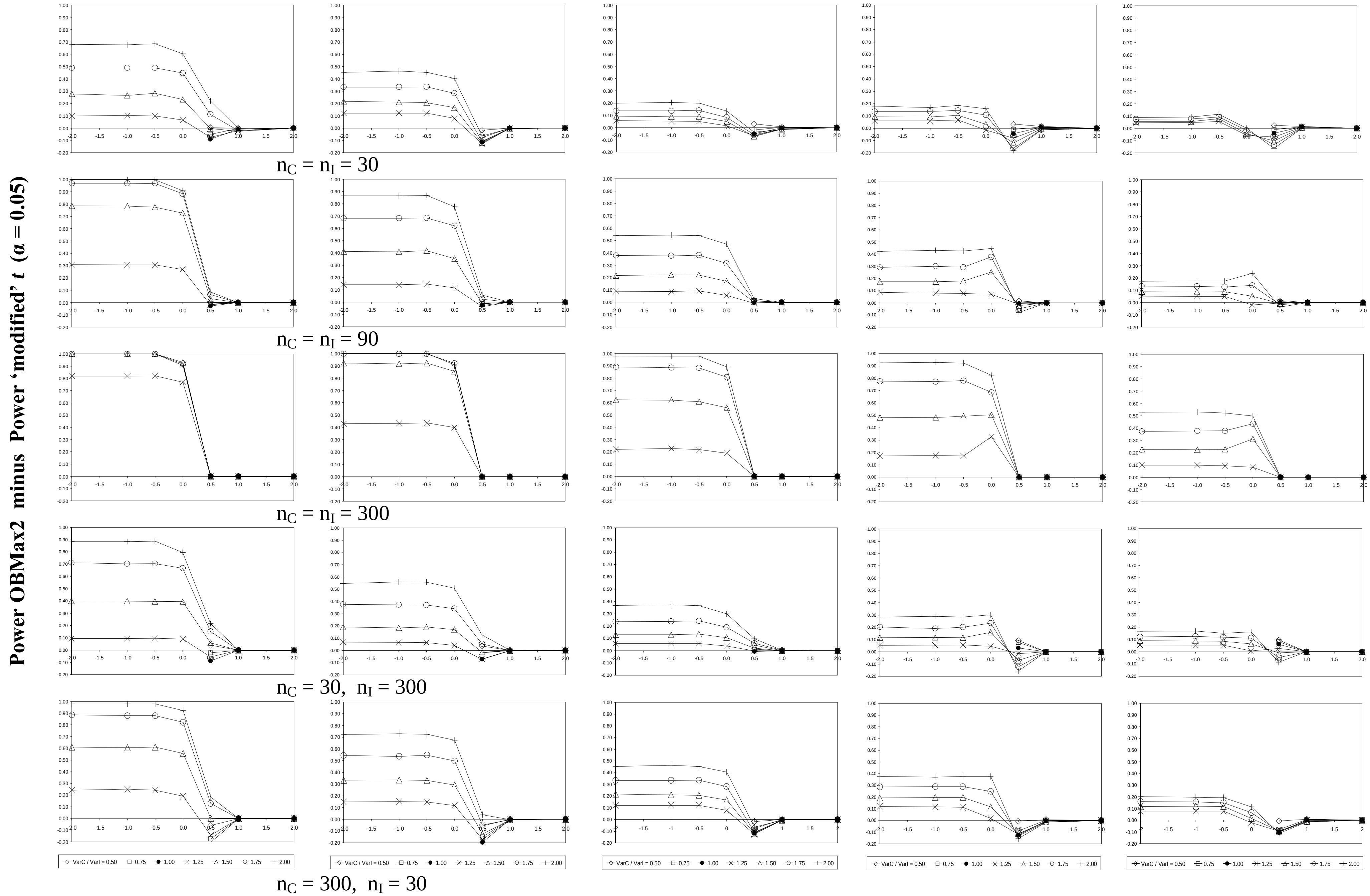

Uniform
Normal
Double Exponential
Exponential
Lognormal
OBMax2 Rejection Rate: Level & Power (α = 0.05)
$n_C = n_I = 30$
$n_C = n_I = 90$
$n_C = n_I = 300$
$n_C = 30,\ n_I = 300$
$n_C = 300,\ n_I = 30$
VarC / VarI = 0.50
0.75
1.00
1.25
1.50
1.75
2.00

5. Results

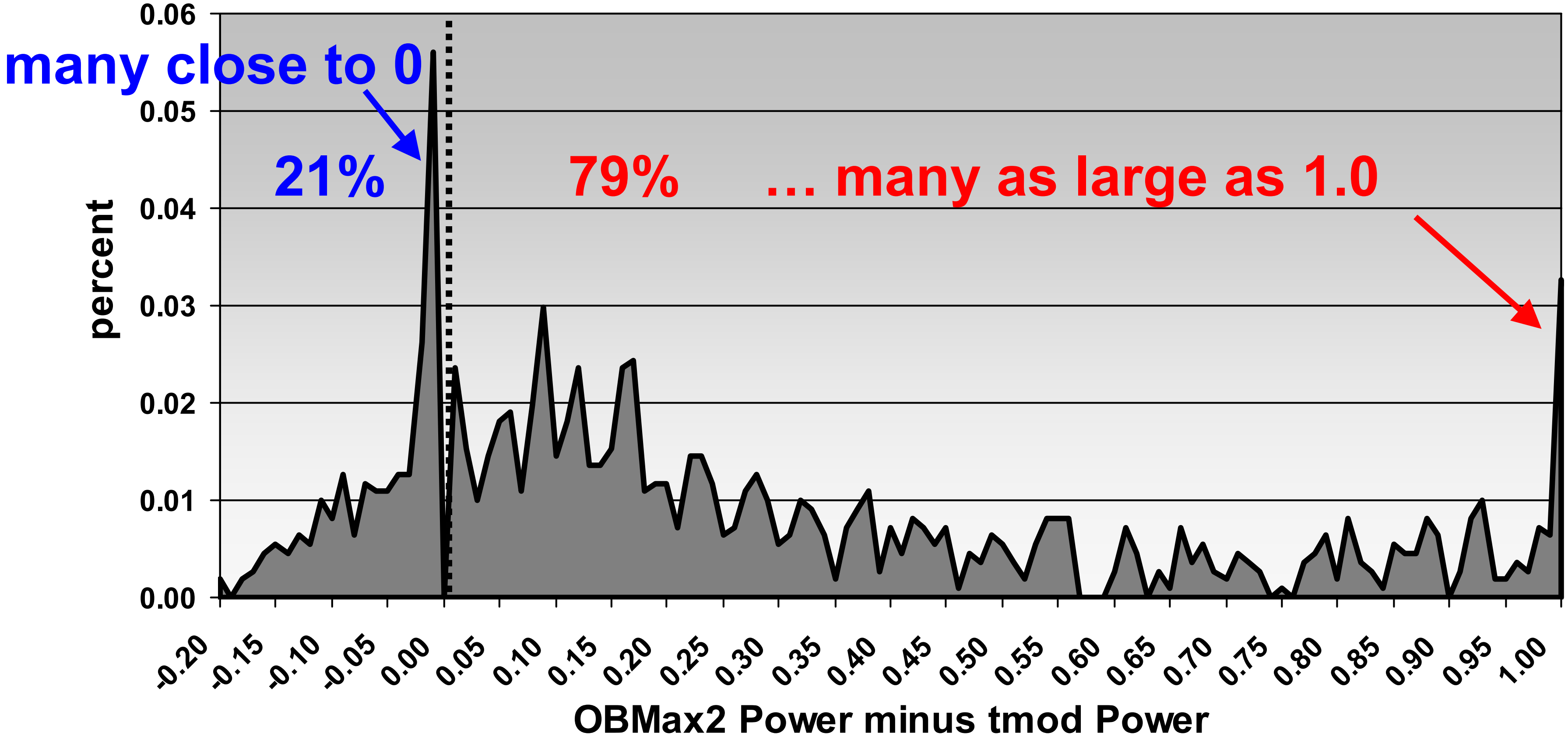

Uniform
Normal
Double Exponential
Exponential
Lognormal
'modified' t Rejection Rate: Level & Power (α = 0.05)
nC = nI = 30
nC = nI = 90
nC = nI = 300
nC = 30, nI = 300
nC = 300, nI = 30
VarC / VarI = 0.50
0.75
1.00
1.25
1.50
1.75
2.00

5. Results

Uniform | Normal | Double Exponential | Exponential | Lognormal

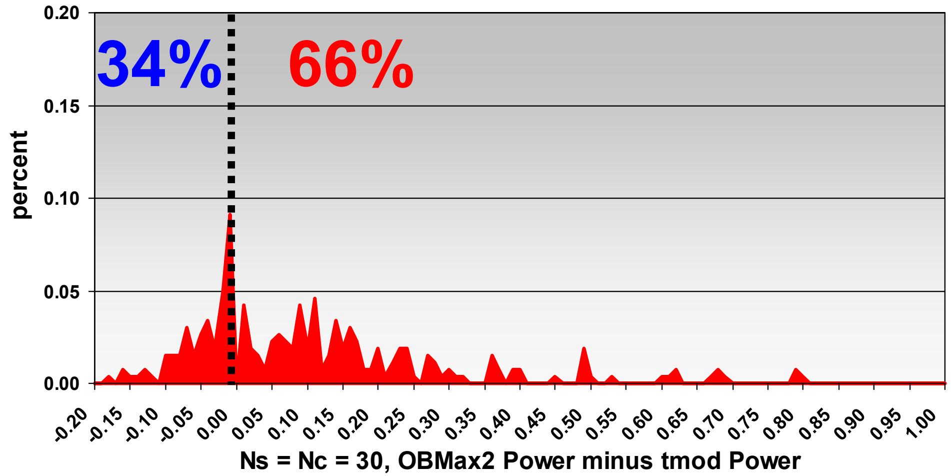


GRAPH 5
OBMax2 Power minus tmod Power
(differences in 1,106 of 1,850 alternate hypotheses)

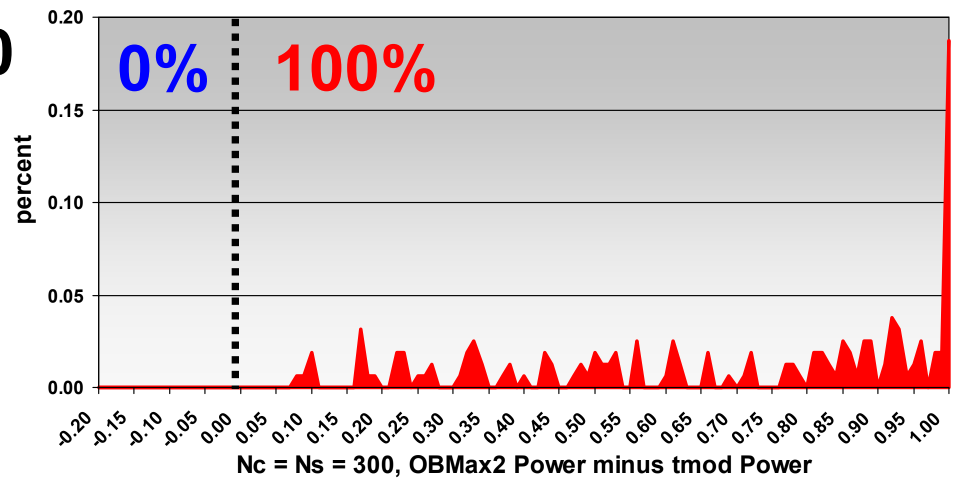


GRAPH 6
OBMax2 Power minus tmod Power by Sample Sizes

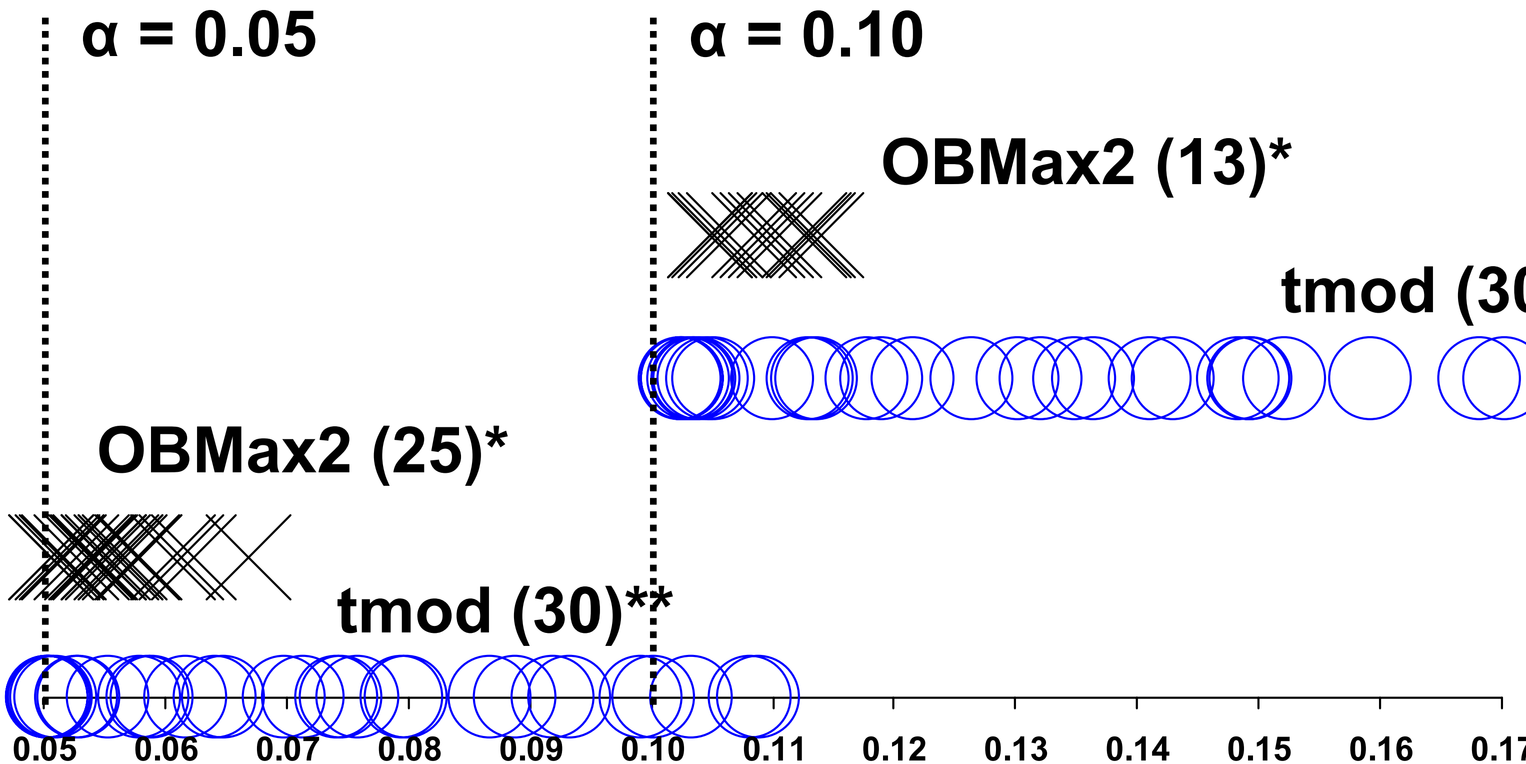
Ns = Nc = 30



Ns = Nc = 300



GRAPH 7
Level Violations: OBMax2 vs tmod (of 600 null hypotheses)



actual level of test

* 15/38 from Ns = 300, Nc = 30

** 47/60 from skewed distributions

6. Conclusions

Strengths:

- **It works! OBMax2 maintains validity under a range of representative distributions where the only proposed competitor cannot.**
- **OBMax2 often has DRAMATICALLY greater power than the only proposed competitor.**
- **OBMax2 is easy to implement – it is not computationally intensive, which is very important for larger sample sizes.**

Limitations:

- OBMax2 adjustments are ad hoc, not based on an analytic derivation of an asymptotic distribution, which would be preferrable
- Although better than the only proposed competitor, OBMax2 still has low power under skewed data
- OBMax2 violates the test level for cross-distributional comparisons (e.g. Study = Uniform vs. Control = Lognormal). However …

- **However…**
  - **Many test statistics are used under this assumption (e.g. most conditions of exchangeability for permutation tests)**
  - **… often because, even with a demonstrated asymptotic distribution, they violate the test level under these conditions with non-asymptotic sample sizes**
  - **And this is not necessarily a limitation for quality control if the processes generating the data are the same or similar and the concern is that undeteted or unaccounted for factors are affecting the mean and/or variance of one of the populations (e.g. two plants performing same process - exponential wait times will not suddenly become uniform)**
  - **Visual inspection can indicate if very different populations are being compared, in which case the question is less about the first two moments and more about completely dissimilar processes.**

Continuing Research:

- Derive the asymptotic distribution of OBMax2, although Yang et al. (2005) have shown that even under stronger assumptions (normality, and no conditioning), this is a non-trivial problem.
- Short of such a derivation, perform more in-depth distributional analyses of OBMax2's p-values.
- Increase OBMax2's power under asymmetry.
- Explore different functional forms for OBt & OBG
- Make OBMax2 more robust to cross-distributional comparisons.
- Compare OBMax2's power against that of BRW and Pesarin's combined permutation test.

References:

- Brown & Forsythe, 1974
- Brownie et al., 1990
- Bűning & Thadewald, 2000
- D'Agostino et al., 1990
- Levene, 1960
- O'Brien, 1988
- O'Gorman, 1997
- Opdyke, 2004
- Opdyke, 2005
- Pesarin, 2000
- Reiczigel et al., 2005
- Rosenbaum, 1954
- Shoemaker, 2003
- Yang et al., 2005

Acknowledgments:

- Sincere appreciation is expressed to Geri S. Costanza, M.S., for numerous and valuable insightful discussions.

Contact:

- J.D. Opdyke, President

DataMineIt SM
WHEN YOU NEED TO KNOW SM

JDOpdyke@DataMineIt.com

www.DataMineIt.com

203-249-4837 (ph) • 781-639-6463 (ph/fax)