Extending the Robust Means Modeling Framework

Alyssa Counsell, Phil Chalmers,
Matt Sigal, Rob Cribbie

One-way Independent Subjects Design

- Model: $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$, $j = 1, \ldots, J$
  - Y_{ij} = score of the ith subject in the jth group
  - μ = population grand mean
  - τ_j = fixed treatment effect for the jth group
  - ε_{ij} = random error component for the ith subject in the jth group

Assumptions of the One-Way Independent Groups ANOVA

- $\varepsilon_{ij} \sim NID(0, \sigma^2)$
  - Population distribution shapes are normal
    - Considerable evidence that this is rarely the case with behavioral science data
  - Populations variances are equal
    - Differences in group variances often exceed 2:1; ratios as large as 8:1 not uncommon

Effects of Assumption Violation on the ANOVA *F*

- When sample sizes and variances are unequal, empirical Type I error rates for the ANOVA *F* can deviate substantially from α
  - Positive Pairings (α =.05): Rates can be deflated below 1% with a 2:1 ratio of largest to smallest variances
  - Negative Pairings (α =.05): Rates can exceed 15% with a 2:1 ratio of largest to smallest variances

Effects of Assumption Violation on the ANOVA *F*

- When distribution shapes are moderately nonnormal, empirical Type I error rates do not deviate substantially from α, however power can be substantially lower than alternative procedures (e.g., nonparametric tests)
  - If distributions are extremely skewed, both Type I error rates and power are affected

Effects of Assumption Violation on the ANOVA *F*

- When the assumptions of normality and variance homogeneity are simultaneously violated, empirical Type I error rates are severely affected
  - Differences between the empirical Type I error rate and α depend on the distribution shape(s), pairings of group sizes and variances, etc.

Comparing Means under Assumption Violation

- Some alternatives to regular ANOVA-F test:
  - Transformations
  - Nonparametric approaches
  - Adjusted degrees of freedom tests
    - E.g., Welch statistic

Modifications to ANOVA

- Most popular alternative is Welch statistic
  - This statistic addresses heterogeneity of variance by using group weights based on a ratio of sample size to variance and adjusting the degrees of freedom

Comparing Means under Assumption Violation

- When distributions are nonnormal and variances are unequal, the recommended alternative is to use the Welch heteroscedastic *F* test in conjunction with trimmed means and Winsorized variances
  - 10-20% trimming is usually recommended

Welch on Trimmed Means

- Although the Welch test using trimmed means and Winsorized variances has been around since Yuen (1974), the test is still rarely used in practice
  - This lack of use may partly be due to the fact that until recently it was not available with popular statistical packages
- However, Rand Wilcox recently released an *R* package (WRS) that will compute the Welch on trimmed means, including multiple comparison tests

Welch on Trimmed Means

- Previous research (Wilcox, 2012) has found that even when distribution shapes are extremely nonnormal, empirical Type I error rates typically do not deviate by more than a couple percentage points from the nominal α level
- Further, power rates often exceed traditional parametric procedures

A New Alternative to ANOVA

- Fan and Hancock (2012) discussed the use of a modified structured means modeling approach, which they call robust means modeling (RMM)

Robust Means Modeling (RMM)

- RMM is incorporated within the framework of structural equation modeling, using simple constraints on means
  - i.e., the means are constrained to be equal; chi square test evaluates whether imposing the mean constraint on the model fits the data

Robust Means Modeling (RMM)

- RMM has important advantages for being able to compare the means of distributions under assumption violation
  - Variances can be explicitly modeled to be distinct across populations
  - Alternative estimation methods are available that help correct for the effects of nonnormality

Estimation

- Estimation is typically done through maximum likelihood as a weighted combination of fit across the k samples
- Other alternatives include:
  - Satorra-Bentler (1988,1992) corrected statistics
  - Asymptotically distribution free methods (Browne, 1984)
  - Modified ADF procedures (e.g., Satterthwaite, Yuan-Bentler (1997, 1999, 2000), etc.)

Fan and Hancock Results

- RMM found to perform quite well under extreme variance heterogeneity (e.g., 16:1 ratios) and severe nonnormality
- RMM outperformed the trimmed Welch across most conditions of variance heterogeneity and nonnormality
  - Trimmed-Welch performed very poorly, contradicting previous studies on the robustness of this procedure

The Current Study

- We aimed to extend the findings of Fan and Hancock (2012) to different conditions (e.g., differing distribution shapes between groups; Part One) and to multiple comparisons (Part Two)

Part One

Omnibus ANOVA

Method

- Monte Carlo Study
- RMM models done using the sem function in lavaan package in R, trimmed Welch using Wilcox's t1way function in the WRS package
- Data simulated to have a normal distribution, moderate skew or severe skew

Method (cont)

- Largest to smallest variance ratios of 1:1, 8:1
- Average $n = 20$, $n = 60$, and 200
- 4 groups
- Equal sample size, negative pairing, positive pairing
- All Means = 0 in Type I Error condition
- Means equally spaced in Power condition

Results

T1 Rates-Normal Distribution

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.050	0.050	0.051	0.055	0.056	0.055	0.057	0.051	0.053
Positive	0.027	0.048	0.050	0.055	0.057	0.055	0.057	0.050	0.052
Negative	0.143	0.052	0.054	0.058	0.061	0.058	0.062	0.055	0.059

T1 Rates-Moderate Skew

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.048	0.053	0.056	0.058	0.070	0.058	0.060	0.053	0.055
Positive	0.028	0.060	0.055	0.066	0.079	0.066	0.067	0.062	0.063
Negative	0.143	0.071	0.060	0.078	0.071	0.078	0.082	0.076	0.078

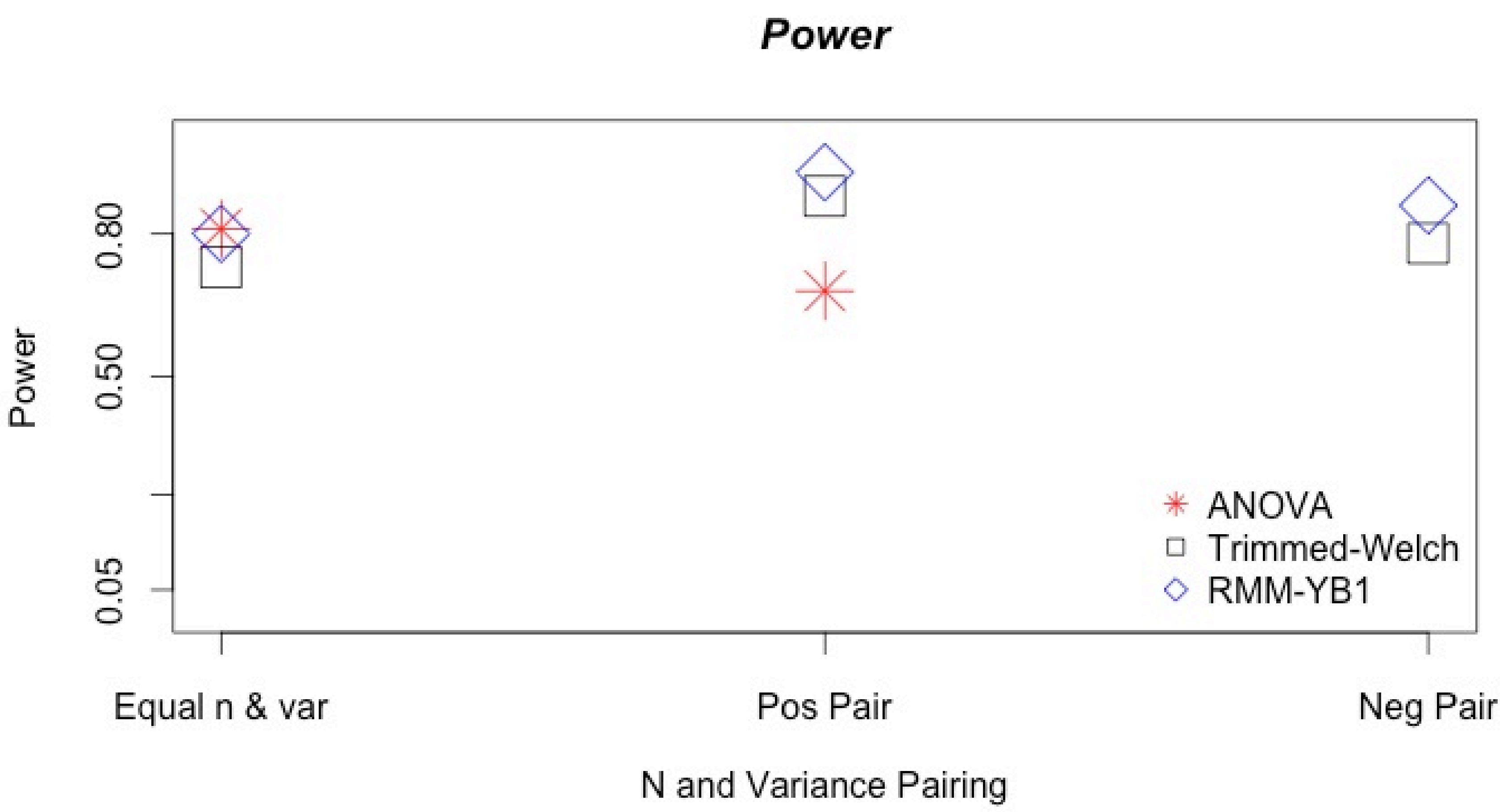
T1 Rates-Extreme Skew

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.041	0.056	0.051	0.060	0.124	0.060	0.061	0.056	0.057
Positive	0.032	0.070	0.045	0.074	0.136	0.074	0.076	0.071	0.073
Negative	0.135	0.119	0.063	0.123	0.119	0.123	0.128	0.121	0.123

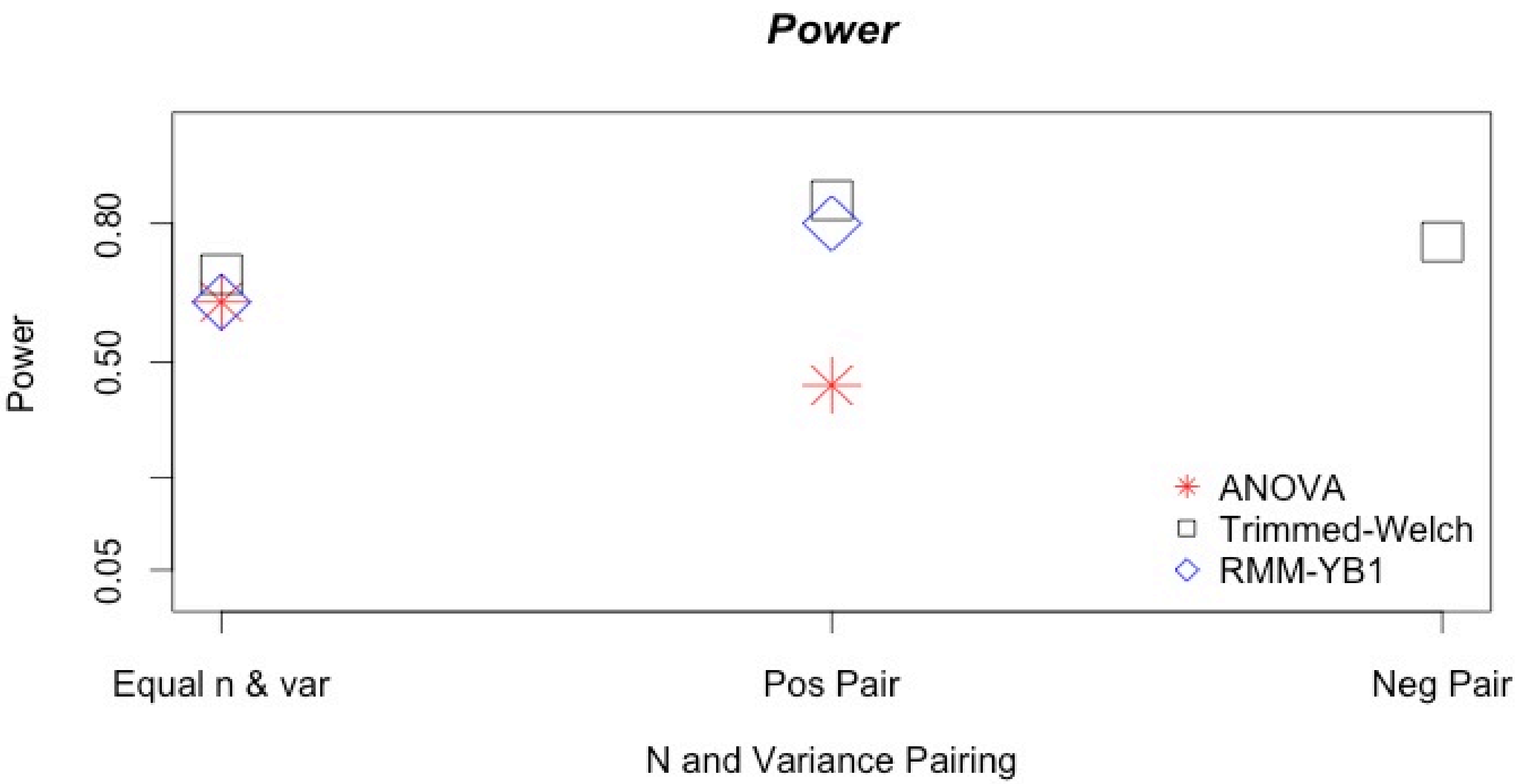
T1 Rates- Different Distributions

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.069	0.077	0.055	0.082	0.076	0.082	0.085	0.079	0.080
Positive	0.043	0.071	0.051	0.078	0.067	0.078	0.080	0.072	0.075
Negative	0.089	0.054	0.056	0.060	0.072	0.060	0.065	0.059	0.060

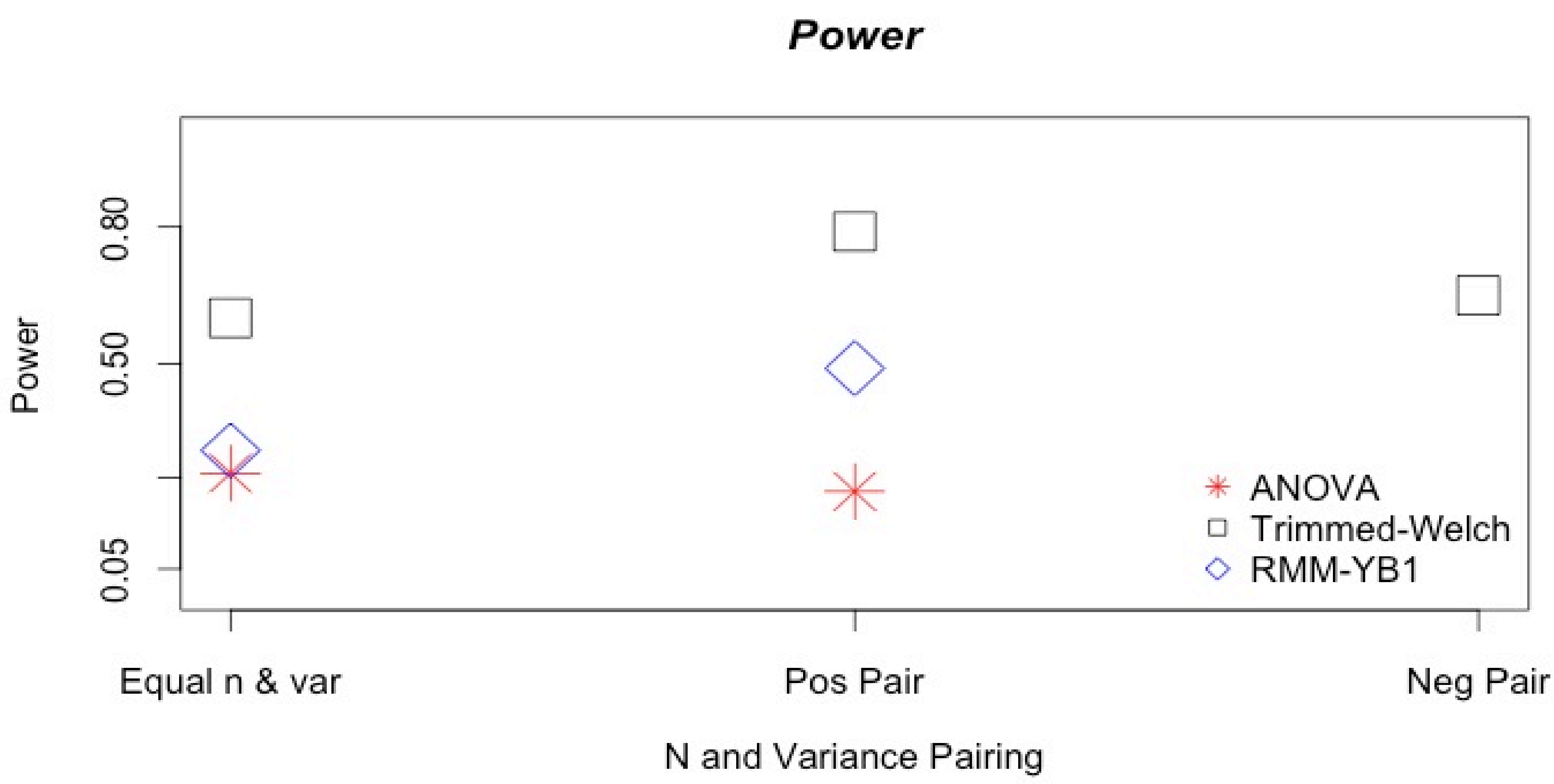
Power- Normal Distribution



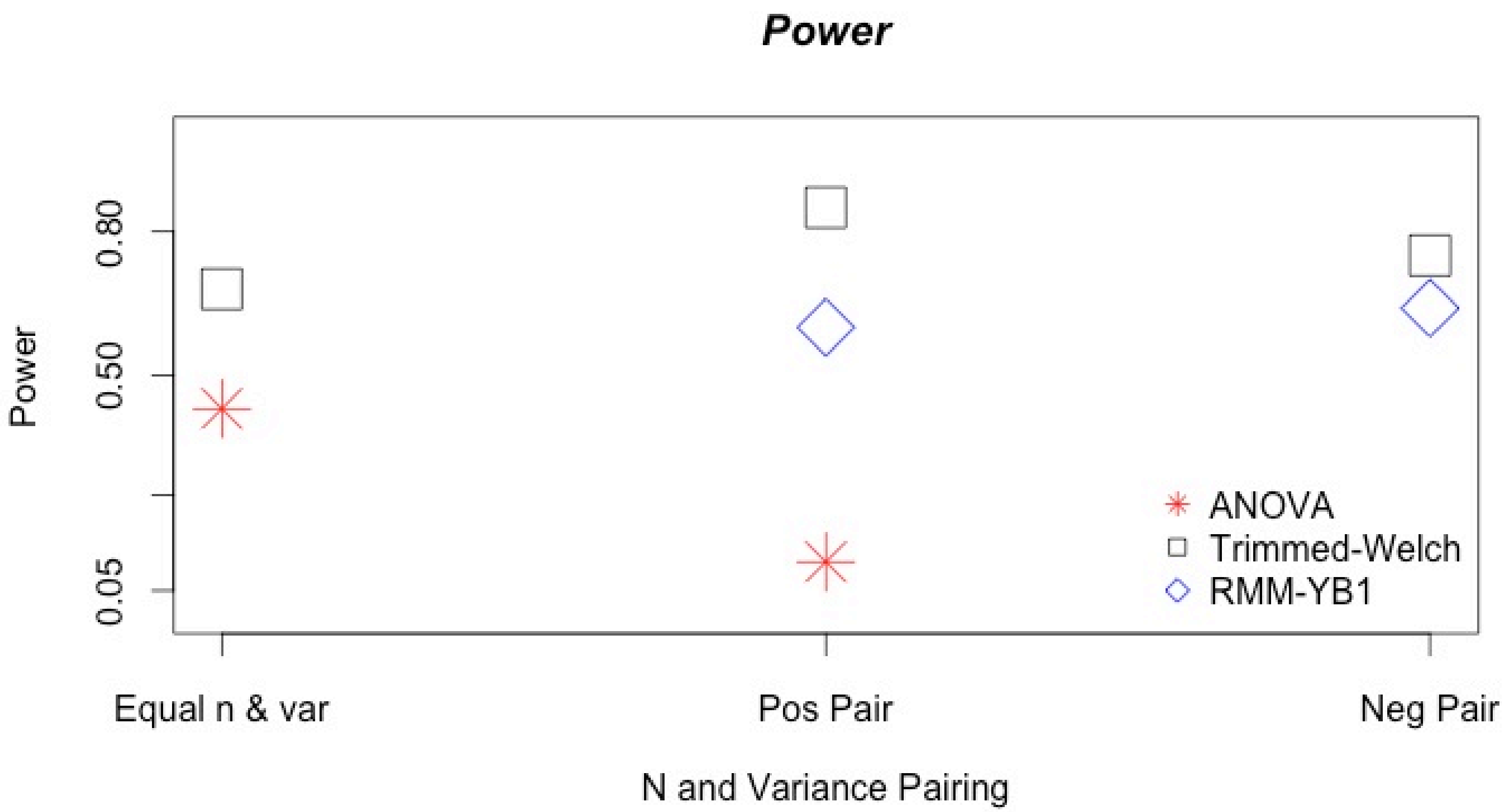
Power- Moderate Skew



Power- Extreme Skew



Power- Different Distributions



Conclusion

- Both the trimmed Welch and RMM methods displayed improved performance compared to the regular ANOVA-F
- Across the conditions, the trimmed Welch tended to outperform the RMM methods

Part Two

Multiple Comparisons

Pairwise Comparisons

- While the omnibus ANOVA tells researchers *whether* differences exist, researchers are more interested in *where* they exist
- Researchers are often interested in whether significant differences exist between any “pair” of treatment conditions

Multiplicity Control?

- For a one-way design, k(k-1)/2 pairwise comparisons are conducted with an appropriate test statistic

  (e.g., t, Welch t, Mann-Whitney)
- Question: What α should be adopted, what error rate should be adopted, and what multiple comparison procedure should be used?

What Error Rate Should Be Adopted?

- Comparisonwise error rate: the probability of falsely rejecting a null hypothesis for any one comparison is set at α
- Familywise error rate (FWE): the probability of falsely rejecting one or more null hypotheses in a family of hypotheses is set at α

How to conceptualize power?

- All pairs power
  - The probability of detecting all true pairwise differences
- Any pairs power
  - The probability of detecting at least one true pairwise difference
- Average per pair power***
  - The average probability of detecting a true pairwise difference

Study Goal

- We wanted to compare the FWE rates and power of pairwise comparisons based on RMM to rates based on traditional methods (ANOVA, trimmed Welch)

Method

- Data generation was identical to the omnibus methods
- We examined the FWE rates and per comparison power
- Procedures compared were:
  - No multiplicity control
  - Bonferroni
  - Holm’s sequential modified-Bonferroni

Results

FWE Rates- Normal Distribution

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.042	0.04	0.042	0.043	0.048	0.043	0.045	0.045	0.04
Positive	0.025	0.039	0.041	0.041	0.046	0.041	0.043	0.043	0.039
Negative	0.129	0.04	0.04	0.044	0.051	0.044	0.05	0.052	0.045

FWE Rates- Moderate Skew

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.035	0.032	0.035	0.034	0.082	0.034	0.037	0.0372	0.032
Positive	0.019	0.035	0.03	0.038	0.089	0.038	0.042	0.042	0.036
Negative	0.118	0.052	0.039	0.056	0.089	0.056	0.062	0.062	0.057

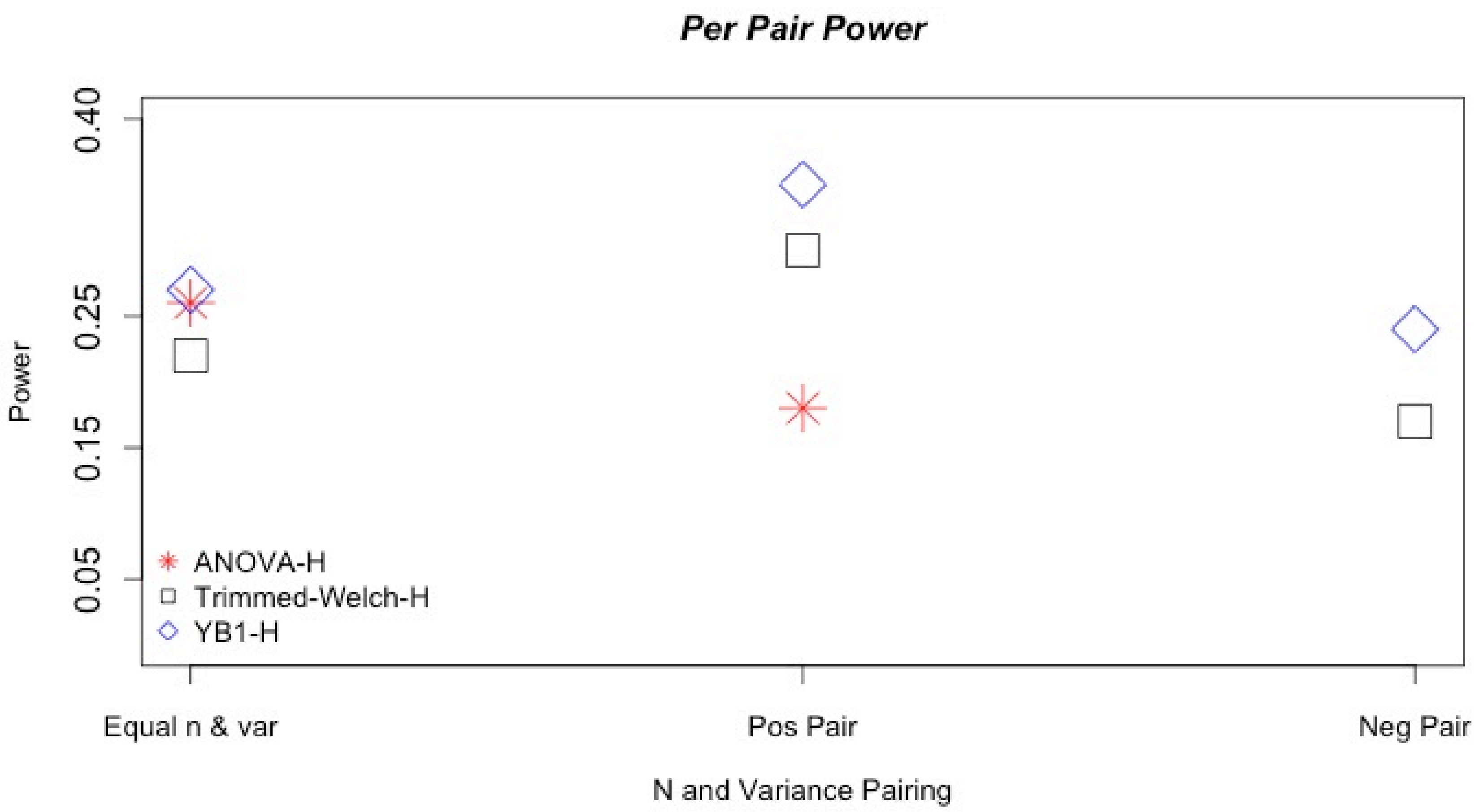
FWE Rates- Extreme Skew

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.032	0.022	0.031	0.025	0.235	0.025	0.028	0.028	0.024
Positive	0.029	0.038	0.031	0.041	0.24	0.041	0.045	0.045	0.04
Negative	0.125	0.102	0.05	0.106	0.212	0.106	0.113	0.113	0.106

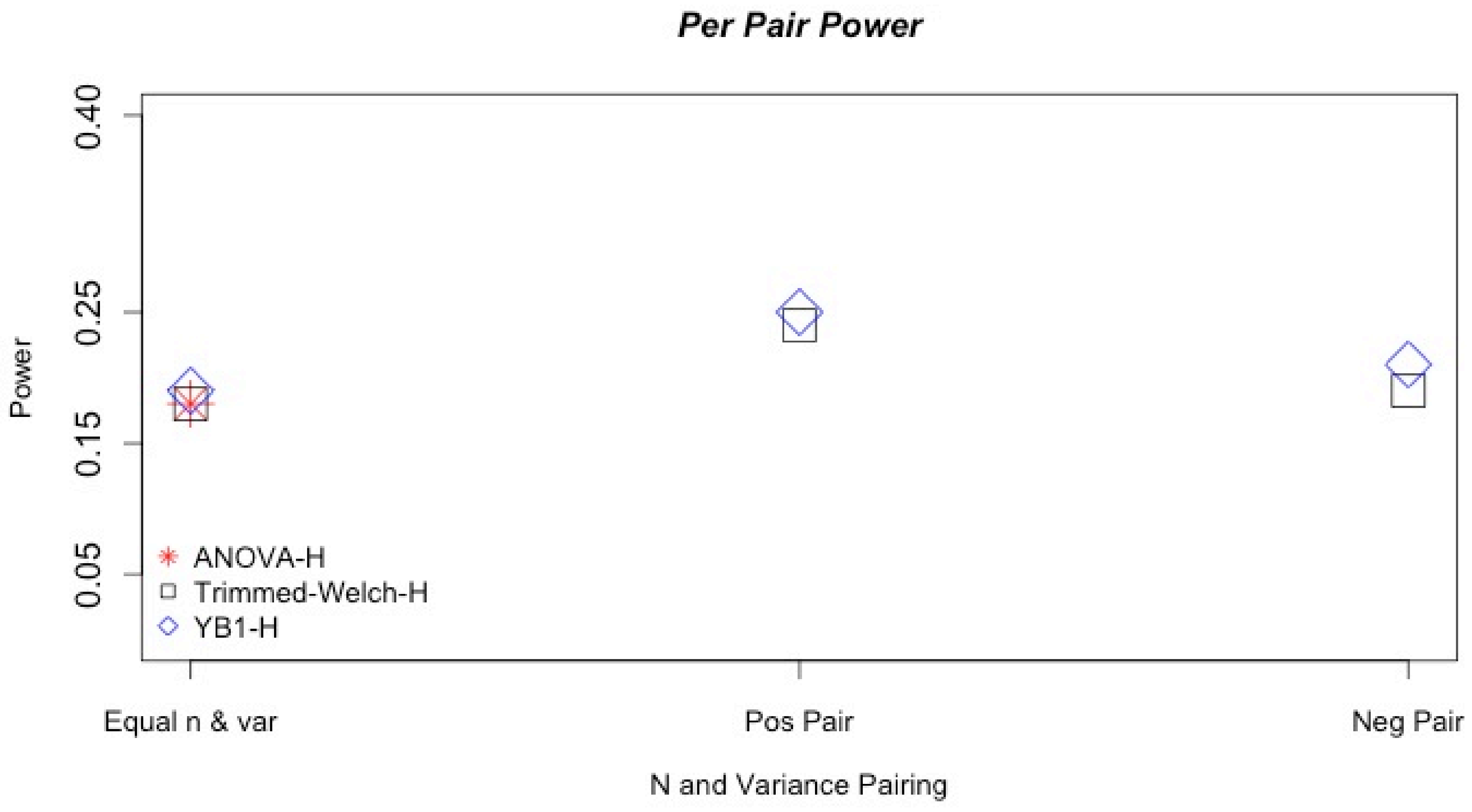
FWE Rates-Different Distributions

Pairing	ANOVA	Welch	Welch trim	RMM ML	MLR	Sat. Bent	WLS	YB1	YB2
Equal	0.062	0.064	0.04	0.067	0.117	0.067	0.07	0.07	0.064
Positive	0.036	0.057	0.039	0.061	0.105	0.061	0.063	0.063	0.057
Negative	0.077	0.038	0.041	0.043	0.089	0.043	0.049	0.049	0.043

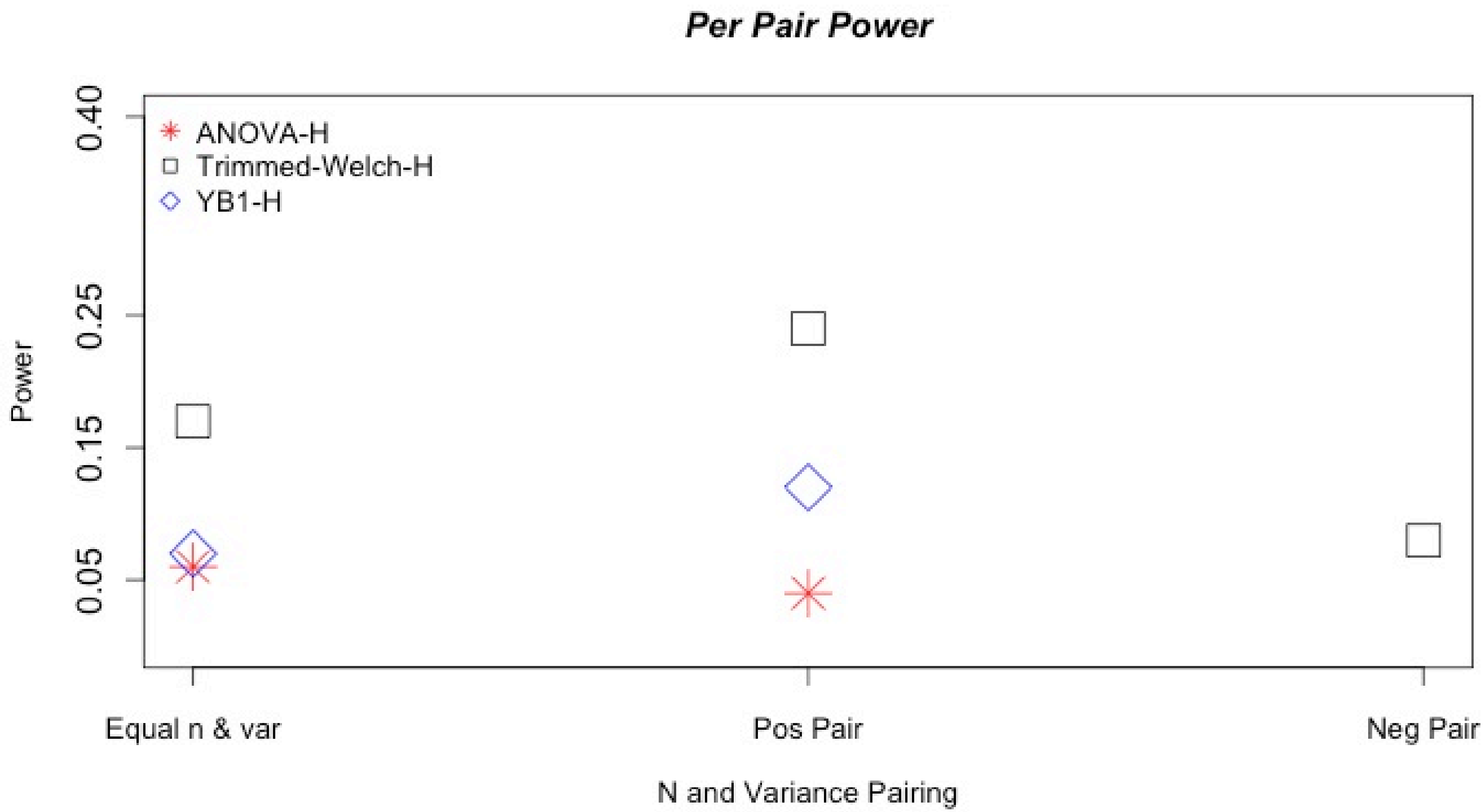
Per Pair Power – Normal Distribution



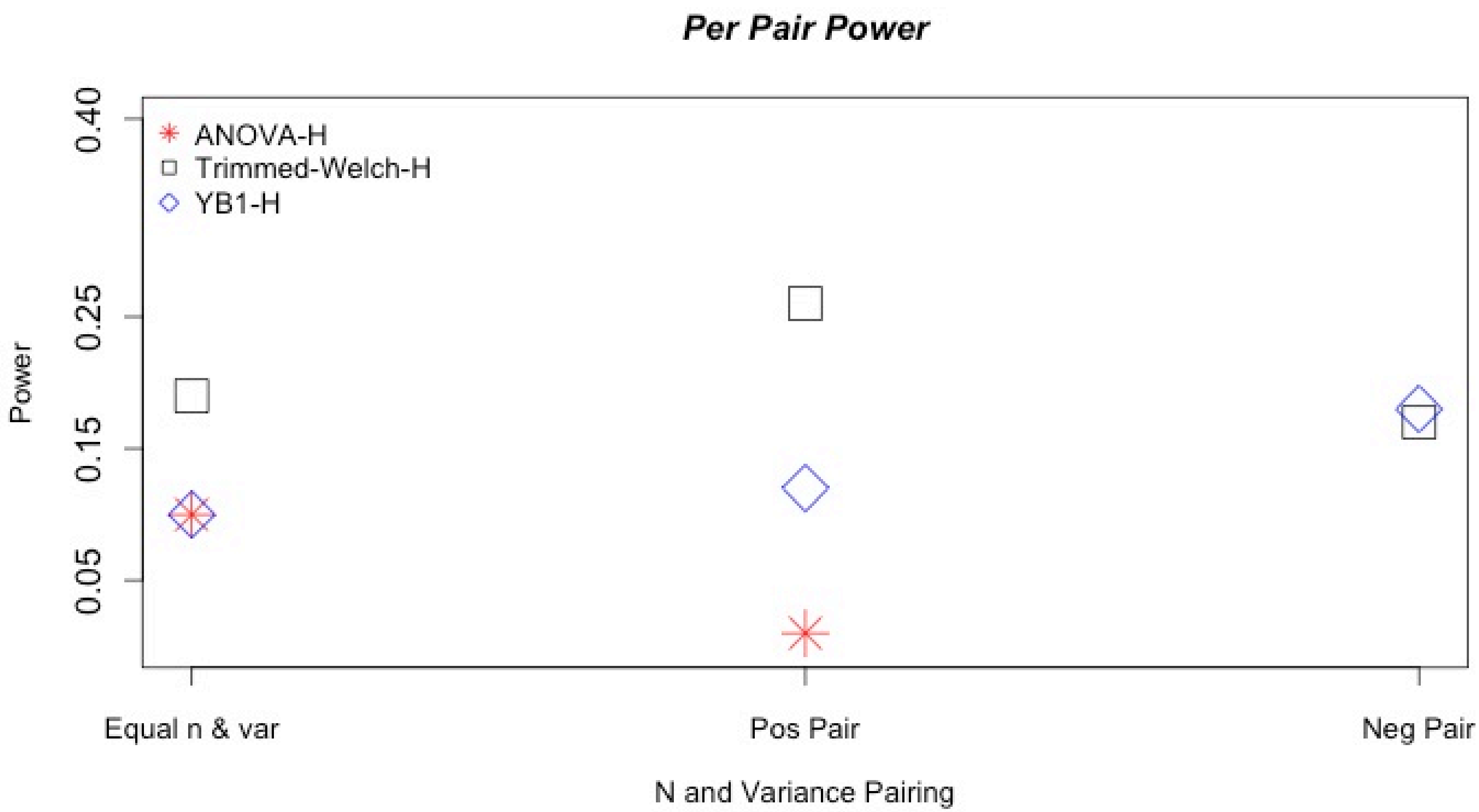
Per Pair Power – Moderate Skew



Per Pair Power – Extreme Skew



Per Pair Power-Different Distributions



Results

- Both the trimmed Welch and most of the RMM methods demonstrate accurate FWE rates across the conditions
  - The YB1 and YB2 procedures performed best for the RMM methods
- However, the trimmed Welch displays greater or comparable power than the RMM methods under all of the conditions

General Conclusion

- Given the power advantage and more accurate Type I error rates, we would recommend using the trimmed Welch over the RMM procedures investigated for comparing the means of independent groups

- However, the flexibility of the procedures (e.g., the ability to model unique group variances) makes it likely that future research will explore improvements to the performance of the RMM procedures that will make them more competitive with existing robust parametric solutions

Thank You!

- Questions?
- Email: counsell@yorku.ca

QUANTITATIVE METHODS
Dept. of Psychology - York University