

UnifyPow: A SAS Macro for Sample-Size Analysis

Ralph G. O'Brien, Cleveland Clinic Foundation, Cleveland, OH

ABSTRACT

UnifyPow is a freeware SAS[®] macro that performs statistical power analysis and other matters related to sample-size choice. Its functionality covers an extensive set of methods. One-group tests include the t and Wilcoxon for $H_0: \mu = \mu_0$; the binomial and Z approximation for $H_0: \pi = \pi_0$; t and Wilcoxon for paired means; McNemar's for paired proportions; and Fisher's r-to-Z for $H_0: \rho = \rho_0$. Two-group tests include t and Wilcoxon-Mann-Whitney for $H_0: \mu_1 - \mu_2 = \delta_0$; χ^2 , likelihood ratio (LR), and Fisher's exact for $H_0: \pi_1 = \pi_2$ (association in 2×2 tables); r-to-Z for $H_0: \rho_1 = \rho_2$. J-group tests include ANOVA via the cell-means model with general linear contrasts; χ^2 and LR tests for proportions ($2 \times J$ tables) with general linear contrasts on J logits. Also covered are the test of $H_0: \beta_j = \beta_{j0}$ in a multiple regression model predicting Y from q Xs ("Y | q Xs") in which X_j has tolerance $Tol(X_j) = 1 - R^2(X_j | \text{other } q - 1 \text{ Xs})$; the comparison of $R^2(Y | q_{\text{full}} \text{ Xs})$ versus $R^2(Y | q_{\text{reduced}} \text{ Xs})$ from nested linear models; and G^2 or $-2\ln L$ from full vs. reduced logistic, log-linear, or Cox survival models. All methods handle unequal n designs. Any set of alpha-levels may be specified and results for directional (one-tailed) tests are given when potentially appropriate. Both tabular and graphical output are available. A simple syntax unifies concepts and specifications across methods. For the current status of this ongoing project and information on downloading, see

<http://www.bio.ri.ccf.org/power.html>

INTRODUCTION

Collaborating statisticians should provide sound technical planning long before data are collected. Essential to this is choosing appropriate sample sizes and assessing statistical power. Nevertheless, sample-size planning has for too long been given short shrift by authors of texts on statistical methods, by statistics teachers, and by software developers, all of whom focus almost exclusively on methods related to analyzing data already collected. When attention *is* given to determining sample size and power, it is too often limited to providing crude approximation formulas or tables to handle a few elementary situations. As a result, a large proportion of research protocols still have inadequate or erroneous sample-size considerations. Nothing will address this problem better than the availability of good, affordable, comprehensive software, especially if this is functionally part of general statistical packages, such as the SAS System.

I have been addressing this matter sporadically for over a number of years. In a series of SUGI presentations (Lohr and O'Brien, 1984; O'Brien, 1984, 1986a, 1986b; O'Brien and Lohr, 1984; Wright and O'Brien, 1988), my colleagues and I discussed this idea and proposed implementations. Indeed, JMP[®] contains such tools for linear models. In a book chapter (O'Brien and Muller, 1993), I described a set of SAS %include modules

centered around one called OneWyPow, which handled power analysis for many types of problems. These modules were distributed as freeware via anonymous ftp. OneWyPow continued to expand in functionality and was later renamed UnifyPow. The O'Brien-Muller chapter appeared in a fairly obscure book. Until now, I have not promoted the freeware via other outlets.

UnifyPow is now a macro that has functionality far beyond that of OneWyPow, and it continues to undergo active development. It is still offered as freeware. Herein, I summarize what UnifyPow can do, give some examples, and outline how to download the latest release.

WHAT CAN UnifyPow DO?

This project is a dynamic one. This section describes what is being released by 10 March 1997 in time for SUGI 22 (San Diego), what I hope to release by 1 August 1997, as well as some things that will appear later.

Key

- * In March 1997 release.
- o Planned for August 1997 release.
- May not make August 1997 release.
- P Finds power for specified total sample size.
- N Find total sample size for specified power.
- CI Finds N to assure sufficiently narrow confidence interval. *CI functionality will not begin to appear until August 1997.*

General Functionality

- * Automatic tabling of results using PROC TABULATE.
- o Options for automatic graphing using SAS/GRAPH[®].
- * Results are collected in a SAS dataset, so user can customize their report, if needed.
- * All methods handle unbalanced sample sizes.
- * Both one-tailed and two-tailed tests are considered, unless impossible.
- o Both one-tailed and two-tailed confidence intervals are considered.

Specific Methods Supported

MEANS (LOCATION)

- * One-sample t test of $H_0: \mu = \mu_0$. Default: $\mu_0 = 0$. [P, N, CI]
- * One-sample Wilcoxon signed-rank test. [P, N]

- * One-sample matched-pairs t test of $H_0: \mu_1 - \mu_2 = \delta_0$. Default: $\delta_0 = 0$. [P, N, CI]
- * Two-sample t test of $H_0: \mu_1 - \mu_2 = \delta_0$. Default: $\delta_0 = 0$. [P, N, CI]
- * Two-sample Wilcoxon-Mann-Whitney test. [P, N]
- * One-way ANOVA overall F test on G independent means. [P, N]
- * Cell-means model ANOVA: the general linear hypothesis on G independent means:

$$H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0},$$
 where \mathbf{C} is $q \times G$ and full row rank; $q \geq 1$. This handles virtually any common test on G cell means in a fixed effects design, including regular and special tests for factorial designs. [P; N; CI, for $q = 1$]
- Factorial ANOVA: tests of general main effects and interactions. Note: This can already be done by properly setting up $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$. [P, N]
- * General F test for linear models. First, outside of UnifyPow construct a set of *exemplary* data whose values conform to the expected values defined by some conjectured true model. That is, begin with an $N_e \times q$ exemplary predictor (design) matrix, \mathbf{X}_e , and a conjectured vector of regression coefficients, $\boldsymbol{\beta}$. Then compute the exemplary outcome vector, $\mathbf{y}_e = \mathbf{X}_e\boldsymbol{\beta}$. Next, use any ordinary linear models routine to fit and ‘test’ the exemplary dataset of N_e cases defined by \mathbf{y}_e and \mathbf{X}_e . (Note: the linear models routine must handle data with $SSE = 0$. PROC GLM does.) Let SSH_e be the obtained sums of squares hypothesis value. Using theory described in O’Brien and Muller (1993), UnifyPow will accept N_e and SSH_e values to drive a sample-size analysis. This technique is cumbersome, but it will handle situations that fall outside of UnifyPow’s menu of common designs and tests. [P, N]

PROPORTIONS, LOGIT ANALYSIS, AND LOG-LINEAR MODELS

- * One-sample binomial test of $H_0: \pi = \pi_0$. Default: $\pi_0 = .50$ (sign test). [P, N, CI].
- * Unconditional test of two independent proportions, $H_0: \pi_1 - \pi_2 = \delta_0$. For default of $\delta_0 = 0$, this conforms to the usual χ^2 test and likelihood ratio test of association in a 2 (groups: A vs. B) \times 2 (outcome: yes vs. no) contingency table. [P; N; CI, for both the log odds ratio, $\ln(\psi) = \ln(\pi_1/\pi_2)$, and $\delta = \pi_1 - \pi_2$.]
- * Fisher's exact conditional test of two independent proportions. This parallels the functionality for the unconditional version of the test described immediately above.

- * Test of two correlated proportions. Let π_{ij} be the true proportion of cases falling into cell $\{i,j\}$ of a 2×2 contingency table. Then $H_0: \pi_{1+} - \pi_{+1} = \delta_0$ compares the row and column marginal probabilities. This is identical to $H_0: \pi_{12} - \pi_{21} = \delta_0$. Conditioning on the discordant frequencies, f_{12} and f_{21} , gives McNemar's test. [P, N, CI, for both log odds ratio, $\ln(\psi) = \ln(\pi_1/\pi_2)$, and $\delta = \pi_1 - \pi_2$.]
- * Overall likelihood ratio test on G independent proportions or logits, $\ln(\psi_j) = \ln[\pi_j/(1 - \pi_j)]$. [P, N]
- * Wald-type tests of the general linear hypothesis over G independent logits:

$$H_0: \mathbf{C}\boldsymbol{\psi} = \mathbf{0},$$
 where \mathbf{C} is $q \times G$ and full row rank; $q \geq 1$. Used properly, this handles log-linear models effects applied to contingency tables in which one of the variables is a dichotomous “response” and the others are predictors. [P; N; CI, for $q = 1$]
- * General likelihood ratio test. Accepts $-2\ln L(\text{full})$ and $-2\ln L(\text{reduced})$ —or $G^2(\text{full})$ and $G^2(\text{reduced})$ —the log likelihood ratio statistics from two nested logistic regression or log-linear models (or other generalized linear model) that were fit to an “exemplary dataset” of N_e artificial cases constructed to give estimates identical to the user's conjectured population parameters. For log-linear models, see the theory of the strategy, the results of the Monte Carlo work, and the example in O'Brien (1986), which is reprised in Agresti (1990). [P, N]
- o Overall likelihood ratio test for multiple logistic regression to predict a binary $Y = 0$ or 1 from q X 's. Let $\pi = \text{Prob}[Y = 1] = E[Y]$. An R^2 -type measure (see Agresti, 1990, p. 110) is

$$D(q \text{ Xs}) = [-2\ln L(\text{null}) - 2\ln L(q \text{ Xs})] / -2LL(\text{null})$$
 where $-2\ln L(\text{null})$ and $-2\ln L(q \text{ Xs})$ are log-likelihood statistics from the null model (intercept only) and the one with q predictors (“ $q \text{ Xs}$ ”). Let $D'(q \text{ Xs})$ be the population counterpart to $D(q \text{ Xs})$, technically, the limit of $D(q \text{ Xs})$ as N_{total} increases. Power is computed by specifying π and $D'(q \text{ Xs})$. [P, N]

REGRESSION AND CORRELATION

- * One-sample test of the Pearson correlation, $H_0: \rho = \rho_0$. Default is $\rho_0 = 0$ and the t test is used. When $\rho \neq 0$, the test based on Fisher's r-to-Z transform is used. [P, N, CI]
- * Overall F test for an ordinary least squares multiple regression model with q X s. Letting ρ^2 be the population counterpart to the common R^2 statistic, this tests $H_0: \rho^2(q \text{ Xs}) = 0$. [P, N]

- Usual t test of $H_0: \beta_j = \beta_{0j}$ based on user's conjectures for β_j ; $SD(X_j)$, the standard deviation of X_j ; $Tol(X_j) = 1 - R^2(X_j | \text{other } q - 1 \text{ Xs})$, the tolerance of X_j in the model with q Xs; and $SD(\epsilon)$, the standard deviation of the residual variation term. [P, N, CI]
- * Multiple partial correlation of q additional predictors given p Xs already in OLS regression model, i.e., $H_0: \rho^2(q \text{ Xs} | p \text{ Xs}) = 0$. [P, N]
- * Linear contrast hypotheses on r-to-Z transformed independent Pearson correlations. Let the j^{th} element of \mathbf{z} be

$$Z(\rho_j) = .5 * \ln[(1 + \rho_j)/(1 - \rho_j)].$$

Then we can test

$$H_0: \mathbf{Cz} = \mathbf{0},$$

where \mathbf{C} is $q \times G$ and full row rank; $q \geq 1$. By far, the most common use of this test is to compare two independent correlations,

$$H_0: \rho_1 - \rho_2 = 0,$$

using

$$H_0: Z(\rho_1) - Z(\rho_2) = 0.$$

[P, N]

SURVIVAL ANALYSIS

- Log-rank tests comparing two groups. [P, N]
- Will offer greater functionality, but plans are not complete.

AGREEMENT

- Test on κ statistic for 2×2 table in one sample, $H_0: \kappa = \kappa_0$. [P, N]
- Test on Lin's (correlation) coefficient of concordance between new continuous assay and gold standard, one sample, $H_0: K = K_0$. [P, N]

EXAMPLES

Limited space prevents a detailed introduction to all of UnifyPow's syntax and methods. I thought it would be most helpful if I just presented some examples. By doing so, I hope to show its ease-of-use and some of its depth. *These examples illustrate only a fraction of UnifyPow's complete functionality.*

Example 1: Comparing Two Independent Proportions

BeeBop Athletic Equipment is testing an experimental prototype of a running shoe being designed to be the successor to its popular XDM model. Does the XDM-X shoe reduce injuries? High-mileage runners will be randomly assigned to run either in the XDM-X or the XDM-S, which is just a standard XDM altered cosmetically to make it also look experimental. To get ample durability data on the XDM-X's, 2/3 of the runners will get the XDM-X. They will run at least 5 times/week,

50 miles/week for 26 weeks. One basic outcome measures will be: Did a serious running-related injury occur? (yes/no)

Let $\pi = \text{Pr}[\text{serious injury}]$. Ample experience with the XDM shoe suggests that about 6% of these runners would experience such an injury in this time period. BeeBop believes that the XDM-X is a breakthrough in biomechanical engineering that could cut this rate in half, to 3%. Thus, the scenario is $\pi_S = .06$ vs. $\pi_X = .03$. BeeBop plans to study about 200 runners, but could recruit as many as 270. What is the statistical power if $\alpha = .05$ and $\alpha = .01$?

The core UnifyPow statements are

```
PI .06 .03 .
WEIGHT .333 .667 .
NTOTAL 201 270 .
ALPHA .05 .01 .
```

(Input 1)

The problem statement (here, PI) must come first. This one sets the design (two independent proportions) and the scenario (.06 vs. .03). If the design was one testing three proportions, then the problem statement would have been something like

```
PI .06 .03 .02 .
```

Note that statements terminate with a lone period (.) and each statement must begin on new line. (I may improve this syntax at some point.) The remaining lines may come in any order. The WEIGHT statement specifies that 1/3 of the cases will be in the first group. NTOTAL calls for power to computed on 201 and 270 total cases (67+134 and 90+180). The ALPHA statement calls for the standard .05 and .01 test sizes. Any α -level may be used, thus allowing for easy analysis of Bonferroni tests, such as $\alpha = .05/3 = .0167$ if a family of 3 tests was being co-protected.

Running UnifyPow with these simple statements generates the results tabulated as Output 1 (next page). Although comparing two independent proportions seems like a simple problem, gifted statisticians have been debating the fine points for years and now almost 25 different methods have been suggested just to get good p values. Obtaining power probabilities for these tests is an even tougher research problem. UnifyPow gives approximate powers corresponding to four tests. The first method given ("Pooled Approximated Unconditional") corresponds to the ordinary Pearson χ^2 test, which can be approximated well by doing an ordinary t test (i.e., using a pooled variance) on $Y = 0$ (no) or 1 (yes) data; see D'Agostino, Chase, and Belanger (1988). The second set of values is also based on the t, but uses the unpooled variance term and approximates the unconditional exact test (Suissa and Shuster, 1985). Purists should favor this, because it carries the "exact" moniker but does not rely on the questionable conditioning restriction of Fisher's exact test, which UnifyPow also handles. Pragmatists should understand that Fisher's exact test generally has lower power than the others, as demonstrated here by UnifyPow. Finally, the last approximation corresponds to testing this hypothesis via a standard logit analysis.

Output 1. Power for testing two independent proportions.

Scenario: pi .06 .03

		ALPHA			
		0.05		0.01	
		Total N		Total N	
		201	270	201	270
		Pow-er	Pow-er	Pow-er	Pow-er
Pooled Approx. Uncondit'1*	2-tail t	.175	.220	.060	.082
	1-tail t	.267	.323	.096	.126
Unpooled Approx. Uncondit'1**	2-tail t	.151	.186	.049	.065
	1-tail t	.234	.281	.079	.103
Fisher's exact conditional	2-tld appr	.090	.129	.025	.040
	1-tld appr	.153	.207	.044	.067
Log Odds Ratio Likhd Ratio	2-tail z	.169	.211	.057	.077
	1-tail z	.258	.311	.091	.120

*The Pooled Approximate Unconditional is just a regular t test using Y = 0 (no) or 1 (yes). It is very similar to the ordinary Pearson chi-square (unconditional) test for a 2 x 2 table, but the t form seems to offer some improvement.

**The Unpooled Approximate Unconditional t test corresponds the Suissa-Shuster unconditional exact test. See references in UnifyPow.guide.

Students and clients understand such tables. They can see directly how increasing α , increasing N_{total} , and using a one-tailed test can increase power. In this case, BeeBop sees power values that are much too small.

Example 2: Testing Two Means (Locations)

In their study of the XDM-X running shoe, BeeBop will analyze a second outcome measure: the proportion of days a runner is injured, including days when he/she runs with the injury. Their data on the current XDM model suggests that this will have a median of about 9% and that 95% of these runners will have rates between 1% and 23%. BeeBop expects the XDM-X to improve on this, perhaps reducing the median injury-day rate to 7%, a 22% reduction. What is the statistical power for this outcome measure?

When the outcome measure itself is a proportion, it is common to transform it using an arcsin function before analysis by Normal-theory methods. We shall use $Y_i = \arcsin(P_i^{1/2})$, where P_i is the injury-day rate for the i^{th} runner. In the Y scale, the 9% median for P becomes a mean of $\mu_S = 0.30$ and the 95% limits on P (1% - 23%) become 0.10 and 0.50. Taking Y as Normal, this range covers about 4σ units, so we conjecture that $\sigma = 0.10$. A median of 7% for P transforms to $\mu_X = 0.27$.

Consider the following UnifyPow statements:

```
MU .30 .27 .                               (Input 2)
WILCOXON
WEIGHT .333 .667 .
SIGMA .08 .10 .125 .
ALPHA .05 .
NTOTAL 201 270 .
```

Here we have a MU problem, that is, we are comparing means with ANOVA methods. If we were comparing G means, then G values would have been given. The SIGMA statement directs UnifyPow to examine the power over a range of three possible standard deviations. Each value for σ gives a separate set of power results based on the

Output 2. Power for testing two means.

Scenario: mu .30 .27
AND Alpha: 0.05

			Std Dev					
			0.08		0.1		0.125	
			Total N		Total N		Total N	
			201	270	201	270	201	270
			Pow-er	Pow-er	Pow-er	Pow-er	Pow-er	Pow-er
Wilcoxon Mann-Whitney (Lehmann power appr)	2-tail W	Normal	.679	.806	.491	.615	.341	.437
		logistic	.737	.856	.546	.675	.382	.488
		Laplace	.847	.935	.669	.796	.486	.610
	1-tail W	Normal	.785	.882	.617	.731	.464	.564
		logistic	.832	.917	.669	.781	.507	.613
		Laplace	.911	.967	.776	.875	.612	.725
t test of two means	2-tail t	Normal	.703	.825	.514	.639	.358	.457
	1-tail t	Normal	.803	.895	.638	.749	.481	.583

ANOVA's assumption of a common standard deviation for the two groups. In this case, our target conjecture of $\sigma = 0.10$ is being "roughened" $\pm 25\%$, to $\sigma = 0.08$ and $\sigma = 0.125$.

The `WILCOXON` statement directs UnifyPow to also give powers for a Wilcoxon-Mann-Whitney test, the standard nonparametric alternative to the two-group `t` test. Approximate powers are computed assuming that the parent distribution for `Y` is either Normal ("light" tailed: $\gamma_2 = 0.0$), logistic ("slightly-heavy" tailed: $\gamma_2 = 1.2$), or Laplace ("moderately-heavy" tailed: $\gamma_2 = 3.0$). It is known that as kurtosis (γ_2) increases, the WMW test becomes more powerful relative to the `t` test. UnifyPow is able to compute power using three different approximations, but only the default one—the one that proved best in some Monte Carlo research I did—is given here.

The output for this example is given on the next page. Students and clients are often surprised to see how just a "minor" change in σ can substantially affect power.

There is also another way to specify the power scenario for the WMW test. Following Lehmann (1975), Hettmansperger (1984), and Noether (1987), the primary parameter for the effect size is

$$p_1 = \text{Prob}(Y_{1i} > Y_{2i'})$$

where Y_{1i} and $Y_{2i'}$ are simply any random pairing of observations from groups 1 and 2, respectively. Now we have $H_0: p_1 = .50$. While p_1 can be found from μ_1, μ_2, σ , and the shape of the parent distribution, in some studies it is easier to just work with p_1 directly. For example, it might be easiest to just ask Beebop researchers, "Consider pairing a random XDM-S runner and a random XDM-X runner. What is the probability that the XDM-S runner will have a greater proportion of injury days? If their conjecture is $p_1 = .60$ and the parent distribution of `Y` is thought to be moderately-heavy tailed, then the following UnifyPow commands would be used:

```
2WILCOXON .60 . (Input 3)
PARENT LAPLACE
NTOTAL 201 270 .
ALPHA .05 .
WEIGHT .333 .667 .
```

Other values for `PARENT` are `NORMAL` and `LOGISTIC`.

Related notes. Instead of the `NTOTAL` statements, we could have used statements such as

```
POWER .80 .90 .
```

This causes UnifyPow to find minimum values for N_{total} to achieve the stated powers.

The statement `2WILCOXON` has the counterpart `1WILCOXON` to handle the one-sample Wilcoxon test, otherwise known as the signed-rank test. Here, we are testing whether the median, δ , exceeds some value, δ_0 . The effect size is

$$p_1 = \text{Prob}(Y > \delta_0).$$

Most applications of Wilcoxon's signed rank test assess the difference between correlated means in a "matched pairs" situation. Here, `Y` is the difference score between

pairs of related observations, so $\delta_0 = 0$. Again, we have $H_0: p_1 = .50$. If you believed that 70% of the pairs would show a "positive" difference score and that the data would be a little more tail-heavy than the Normal, then you would specify

```
1WILCOXON .70 .
PARENT LOGISTIC
```

Example 3: One-Way ANOVA with Complex Contrasts

What if BeeBop had 3 variations of their experimental shoe, XDM-X₁, XDM-X₂, XDM-X₃? The design would have four groups. Expanding on Input 2, consider the statements

```
MU .300 .275 .270 .265 . (Input 4)
WEIGHT .334 .222 .222 .222 .
SIGMA .08 .10 .125 .
ALPHA .05 .
NTOTAL 207 270 .
NOOVERALL
CONTRASTS
"(1) XDM-S vs. XDM-X (all versions)"
  3 -1 -1 -1 .
"(2) Variation among XDM-X shoes"
  0 1 -1 0 .
> 0 0 1 -1 .
```

If the `NOOVERALL` statement had not been used, UnifyPow would have computed the power for the overall `F` test with 3 degrees of freedom, a test of questionable value in this situation. The `CONTRASTS` statement focuses on the two main questions in this study: (1) Is the XDM-X (averaged over the three variations) better overall? (2) Is there any difference among the XDM-X variations? These are separate questions and do not need to be handled as a family of contrasts. Persons thinking more conservatively are free to use a Bonferroni adjustment, i.e. $\alpha = 0.05/2 = 0.025$.

Being able to handle cell-means contrasts with multiple degrees of freedom gives UnifyPow exceptional flexibility.

Other Examples

Other examples can be found in the downloadable files.

GETTING THE FREeware

UnifyPow is freeware distributed primarily via the anonymous ftp site at the Department of Biostatistics and Epidemiology at the Cleveland Clinic Foundation. Downloading new versions periodically ensures that you are getting the latest version I feel is safe for public distribution.

The relevant ftp address is

```
ftp.bio.ri.ccf.org
```

and the directory is

```
UnifyPow.all
```

in the root directory (not in `/pub`). If you know how to obtain documents via anonymous ftp, this is all the information you need. Start by obtaining and reading the file `ReadMe.power`.

You can also use Netscape to get the files. I recommend this, but... beware of "cutting-edge" glitches. Go to

<http://www.bio.ri.ccf.org/power.html>

Clicking on the obvious hypertext should give you a list of the files in the UnifyPow.all directory. Clicking on a filename (Start with ReadMe.power.) should bring contents directly onto the screen. Doing a "Save As..." (specifying either "text" or "source") should create an ASCII file on your machine. I have been able to read these with many different editors and word processors, but not all.

Legal Disclaimer

UnifyPow and its related files are freeware. You may give them to others at no charge, and you may not charge a "service fee" for their distribution. I distribute this work *pro bono*, in the spirit of collaborative science.

Use these modules at your own risk.

THIS FREWARE COMES WITHOUT ANY WARRANTY WHATSOEVER. RALPH O'BRIEN AND THE CLEVELAND CLINIC FOUNDATION DO NOT AND CANNOT WARRANT THE PERFORMANCE OR RESULTS YOU MAY OBTAIN BY USING THIS FREWARE AND ITS DOCUMENTATION. IN NO EVENT WILL RALPH O'BRIEN AND THE CLEVELAND CLINIC FOUNDATION BE LIABLE TO YOU FOR ANY CONSEQUENTIAL, INCIDENTAL, OR SPECIAL DAMAGES, INCLUDING LOST PROFITS OR LOST SAVINGS, EVEN IF RALPH O'BRIEN OR THE CLEVELAND CLINIC FOUNDATION HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES, OR FOR ANY CLAIM BY ANY THIRD PARTY.

PROBLEMS AND SUGGESTIONS

One reason this project has remained active is because I really do appreciate knowing if you encounter problems or have suggestions for improvements. Most of the additions to UnifyPow come about this way. On the other hand, I cannot promise to address everyone's queries, especially those that are mostly related to consulting advice. I correspond best via email.

REFERENCES

- Agresti A (1990), *Categorical Data Analysis*, New York, John Wiley and Sons.
- D'Agostino RB, Chase W, Belanger A (1988), "The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Proportions," *American Statistician*, 42, 198-202.
- Hettsmansperger TP (1984), *Statistical Inference Based on Ranks*, New York: John Wiley and Sons.
- Lehmann EL (1975), *Nonparametrics: Statistical Methods Based on Ranks*, New York: John Wiley and Sons.
- Noether GE (1987), "Sample Size Determination for Some Common NonParametric Tests," *Journal of the American Statistical Association*, 82, 645-647.
- O'Brien RG, Lohr VI (1984), "Power Analysis for Linear Models: The Time Has Come," *Proceedings of the Ninth Annual SAS Users Group International Conference*, Cary NC: SAS Institute Inc., 840-846.
- O'Brien RG (1986a), "Using the SAS System to Perform Power Analyses for Log-Linear Models," *Proceedings of the Eleventh SAS Users Group International Conference, Cary, NC, SAS Institute*, 778-784.
- O'Brien RG (1986b), "Power Analysis for Linear Models," *Proceedings of the Eleventh SAS Users Group International Conference*, Cary, NC, SAS Institute, 915-922.
- O'Brien RG and Muller KE (1993), "Unified Power Analysis for t Tests through Multivariate Hypotheses," in Edwards, L K (Ed.), *Applied Analysis of Variance in Behavioral Science*, New York: Marcel Dekker, 297-344.
- Lohr VI and O'Brien RG (1984), "Power Analysis for Univariate Linear Models: SAS Makes It Easy," *Proceedings of the Ninth Annual SAS Users Group International Conference*, Cary NC: SAS Institute Inc., 847-852.
- Suissa S, Shuster JJ (1985), "Exact Unconditional Sample Sizes for the 2×2 Comparative Trial," *Journal of the Royal Statistical Society (A)*, 148, 317-327.
- Wright SP and O'Brien RG (1988) "Power Analysis in an Enhanced PROC GLM: What It Might Look Like," *Proceedings of the Thirteenth SAS Users Group International Conference*, Cary, NC, SAS Institute, 1097-1102.

Ralph O'Brien, PhD, directs the Collaborative Biostatistics Center within the Department of Biostatistics and Epidemiology, Cleveland Clinic Foundation, Cleveland, Ohio, 44195. 216-445-9451; robrien@bio.ri.ccf.org.