

Contrast Variable Potentially Providing a Consistent Interpretation to Effect Sizes

Xiaohua Douglas Zhang

Biometrics Research, WP53B-120, Merck Research Laboratories, West Point, PA

Abstract

Various effect sizes have been proposed. However, different effect size measures are suitable for different types of data, and the interpretations of effect sizes are generally arbitrary and remain problematic. In this article, the concepts of contrast variable, its standardized mean (SMCV) and c^+ -probability are explored to link together the commonly used effect sizes including the probabilistic index and ratios of mean difference to variability. A contrast variable can provide both a probabilistic meaning and an index of signal-to-noise ratio to interpret the strength of a comparison, which offers us a strong base to classify the strength of a comparison. Contrast variable, SMCV and c^+ -probability not only give interpretations to both Cohen's and McLean's criteria but also work effectively and consistently for either relationship or group comparison in either independent or correlated situations and in either two or more than 2 groups. Treatment effect, main effect, interaction effect, linear relationship, quadratic relationship and any other contrasts in factorial experiments can all be addressed consistently using contrast variables and the SMCV-based classifying rule, as demonstrated using examples in this article. Therefore, contrast variable, SMCV, c^+ -probability and associated classifying rules may have the potential to offer a consistent interpretation to effect sizes.

Keywords: Contrast variable; Standardized mean of contrast variable; C^+ -Probability; Effect size; ANOVA

Introduction

Traditional contrast analysis tackles the question of whether a linear combination of group means is exactly zero using significance testing. However, in reality, the key question of interest is usually how far a linear combination of values in multiple groups involved in a comparison is away from zero in a distribution level. Therefore, to effectively compare groups, we need additional analysis to incorporate information in a distribution level. In addition, the value of a traditional contrast has issues in capturing data variability. The p -value from classical t -test of testing a traditional contrast can capture data variability; however, it is affected by both sample size and the strength of a comparison. This issues with significance testing of mean difference are reflected in Tukey's comments [33] "It is foolish to ask 'Are the effects of A and B different?' They are always different --- for some decimal place." All the problems in traditional contrast analysis attribute to the calls for a critical re examination of the common use of p -value and significance testing in medical research [9,15,16,23,30,36].

To address the issues of traditional contrast analysis, various effect sizes have been proposed [8,13,20,34]. Many of them may fall into two categories: probabilistic indices for comparing groups in a distribution level [1,4,6,12,14,19,21,25,28,31] and metrics for capturing both mean and variability, which includes various effect sizes similar to standardized mean differences [4,7,10,14,21,25,26,28,29,31,32,35,36]. However, different effect size measures are suitable for different types of data, and the interpretations of effect sizes are generally arbitrary and remain problematic [13,24,26]. Therefore, clear and consistent interpretations for the strength of group comparison are important.

In this article, I explore how to use recently proposed concept of contrast variable [39,40] to integrate the two categories of effect sizes. Meanwhile, based on the concepts of contrast variable and standardized mean of contrast variable (SMCV), I explore a rule for classifying the strength of comparison. The thresholds in this rule have meanings from both strength and probability perspectives;

hence this rule may provide a clear and consistent interpretation to the strength of group comparisons. The concept of contrast variable can be applied not only to measure the strength of any contrast including difference between two groups in general but also to assess the size of main effects, interaction effects and any other contrasts (such as linear or quadratic effects) in Analysis of Variance (ANOVA). Hence, the use of contrast variable and associated classifying rule may have the potential to establish a consistent standard for assessing the strength of comparisons and the size of various effects in ANOVA.

Contrast variable, SMCV and c^+ -probability

Contrast analysis is a widely used statistical method to compare the values in multiple groups or treatment levels. Traditionally, a contrast is defined as a linear combination of means in the groups or treatment levels where the coefficients sum to zero. Traditional contrast analysis focuses on testing whether an average effect is exactly zero, above zero, or below zero. However, in many cases, scientists are interested in knowing the magnitude of effect beyond the question of average effect being exactly zero and the direction of a contrast. As a potential solution to these issues, contrast variable, SMCV and c^+ -probability have been proposed for addressing the strength of a comparison [39,40].

Briefly, a contrast variable is defined as a linear combination of random variables representing the values in the groups (or treatment levels) themselves, not just a linear combination of their means, where the coefficients sum to zero. The mean of a contrast variable

***Corresponding author:** Xiaohua Douglas Zhang, Biometrics Research, WP53B-120, Merck Research Laboratories, West Point, PA 19486, E-mail: xiaohua_zhang@merck.com

Received November 10, 2010; **Accepted** December 22, 2010; **Published** December 24, 2010

Citation: Zhang XD (2010) Contrast Variable Potentially Providing a Consistent Interpretation to Effect Sizes. J Biomet Biostat 1:108. doi:10.4172/2155-6180.1000108

Copyright: © 2010 Zhang XD. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



equals a traditional contrast. SMCV is the ratio of mean and standard deviation of a contrast variable representing a comparison or relationship of interest. Consequently, an effect size (i.e., SMCV) and a traditional contrast (i.e., contrast mean) are now two characteristics of the same random variable; thus a contrast variable integrates both an effect size and a traditional contrast. c^+ -probability is the probability that a contrast variable obtains a positive value. In other words, when one draws one value from each group involved in a comparison and calculate a linear combination of these drawn values based on the contrast's coefficients, c^+ -probability is the proportion that the calculated value is greater than zero when the drawing process is repeated infinite times. c^+ -probability is a probabilistic index accounting for distributions of compared groups whereas SMCV is an extended variant of standardized mean difference (such as Cohen's d , Glass's Δ and Hedge's g) incorporating both mean and variance of groups. There is a link between SMCV and c^+ -probability [39]. Thus, standardized mean difference and probabilistic index are now integrated to effectively assess the strength of a comparison. In addition, the concepts of SMCV and c^+ -probability are applicable to not only the comparison of two groups but also the comparison of more than two groups.

The concepts in contrast analysis can be formalized mathematically as follows. Suppose the random values in t groups represented by random variables G_1, G_2, \dots, G_t have means $\mu_1, \mu_2, \dots, \mu_t$ and variances $\sigma_1^2, \sigma_2^2, \sigma_t^2$ respectively. Then a traditional contrast L is $L = \sum_{i=1}^t c_i \mu_i$ where $\sum_{i=1}^t c_i = 0$. Traditional contrast analysis focuses on testing $H_0: L = 0, H_0: L \leq 0$ or $H_0: L \geq 0$. A contrast variable V is a linear

combination of the random variables, i.e., $V = \sum_{i=1}^t c_i G_i$ where c_i 's are a set of coefficients representing a comparison of interest and satisfy $\sum_{i=1}^t c_i = 0$. The SMCV of contrast variable V , denoted by λ is

$$\lambda = \frac{\mu_V}{\sigma_V} = \frac{\sum_{i=1}^t c_i \mu_i}{\sqrt{\text{var}\left(\sum_{i=1}^t c_i \mu_i\right)}} = \frac{\sum_{i=1}^t c_i \mu_i}{\sqrt{\sum_{i=1}^t c_i^2 \sigma_i^2 + 2 \sum_{i=1}^t \sum_{j=1}^t c_i c_j \sigma_{ij}}}$$

where σ_{ij} is the covariance of G_i and G_j . When G_1, G_2, \dots, G_t are independent,

$$\lambda = \frac{\sum_{i=1}^t c_i \mu_i}{\sqrt{\sum_{i=1}^t c_i^2 \sigma_i^2}}.$$

The c^+ -probability for a contrast variable V is the probability that the contrast variable obtains a positive value, i.e., $\Pr(V > 0)$. There is a strong relationship between c^+ -probability and SMCV of a contrast variable V as follows [39]:

1). for any distribution of V , c^+ -probability = $1 - F(-\lambda)$, where $F(\cdot)$ is the cumulative distribution function of the standardized contrast variable

$$W = \frac{V - \mu_V}{\sigma_V};$$

2). if V has a normal distribution, c^+ -probability = $\Phi(\lambda)$, where

A: SMCV-based Criterion			
Effect Type	SMCV (or SSMD)	c^+ -probability (or d^+ -probability)	Cohen's d
Extra large positive effect	[1.645, +∞]	[0.95, 1]	[2.324, +∞]
Large positive effect	[1, 1.645)	[0.84, 0.95)	[1.414, 2.326)
Medium large positive effect	[0.5, 1)	[0.7, 0.84)	[0.707, 1.414]
Medium positive effect	(0.25, 0.5)	(0.60, 0.7)	(0.354, 0.707)
Small positive effect	(0, 0.25]	(0.50, 0.60]	(0, 0.354]
No effect	0	0.50	0
Small negative effect	[-0.25, 0)	[0.40, 0.50)	[-0.354, 0)
Medium negative effect	(-0.5, -0.25)	(0.3, 0.40)	(-0.707, -0.354)
Medium large negative effect	(-1, -0.5]	(0.16, 0.3)	(-1.414, -0.707]
Large negative effect	(-1.645, -1]	(0.05, 0.16]	(-2.326, -1.414]
Extra large negative effect	(-∞, -1.645]	[0, 0.05]	(-∞, -2.326]
B: Cohen's Criterion			
Effect Type	SSMD	d^+ -probability or CL	Cohen's d
Large positive effect	0.5657	0.7142	0.8
Medium positive effect	0.3536	0.6382	0.5
Small positive effect	0.1414	0.5562	0.2
Small negative effect	-0.1414	0.4438	-0.2
Medium negative effect	-0.3536	0.3618	-0.5
Large negative effect	-0.5657	0.2858	-0.8
C: Mclean's Criterion			
Effect Type	SSMD	d^+ -probability or CL	Cohen's d
Large positive effect	(0.707, +∞]	(0.76, 1]	(1, +∞]
Moderate positive effect	(0.354, 0.707]	(0.64, 0.76]	(0.5, 1]
Small positive effect	(0, 0.354]	(0.5, 0.64]	(0, 0.5]
No effect	0	0.5	0
Small negative effect	[-0.3536, 0)	[0.36, 0.5)	[-0.5, 0)
Moderate negative effect	[-0.707, -0.354)	[0.24, 0.36)	[-1, -0.5)
Large negative effect	(-∞, -0.707)	[0, 0.24)	(-∞, -1)

Note: CL denotes the common language effect size indicator in the 2-group case proposed by [21].

Table 1: Effect size criteria for classifying the strength of a contrast.

$\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution;

3). if V has a unimodal distribution with finite variance $\sigma_v^2 > 0$,

$$\begin{cases} c^+ - \text{probability} \geq 1 - \frac{4}{9\lambda^2}, & \text{for } \lambda \geq \sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \geq \frac{4}{3} - \frac{4}{3\lambda^2}, & \text{for } 1 \leq \lambda \leq \sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \leq \frac{4}{9\lambda^2}, & \text{for } \lambda \leq -\sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \geq \frac{4}{3\lambda^2} - \frac{1}{3}, & \text{for } -1 \geq \lambda \geq -\sqrt{\frac{8}{3}} \end{cases}$$

4). if V has a symmetric unimodal distribution with finite variance $\sigma_v^2 > 0$,

$$\begin{cases} c^+ - \text{probability} \geq 1 - \frac{2}{9\lambda^2}, & \text{for } \lambda \geq \sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \geq \frac{7}{6} - \frac{4}{3\lambda^2}, & \text{for } 1 \leq \lambda \leq \sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \leq \frac{2}{9\lambda^2}, & \text{for } \lambda \leq -\sqrt{\frac{8}{3}} \\ c^+ - \text{probability} \geq \frac{2}{3\lambda^2} - \frac{1}{6}, & \text{for } -1 \geq \lambda \geq -\sqrt{\frac{8}{3}} \end{cases}$$

A classifying rule for interpreting the strength of comparisons

Clear and consistent interpretations for the strength of group comparison are important and in urgent needs, which is reflected in Rosenthal, Rosnow and Rubin [26]'s comment, "Despite the growing awareness of the importance of estimating sizes of effects along with obtaining levels of significance, problems of interpretation remain." as well as in [13]'s comments, "The interpretation of the index value magnitude is, perhaps, the biggest limitation of the use of effect size." Based on contrast variable, SMCV and c^+ -probability, we can provide a clear and consistent interpretation to the strength of a comparison.

With contrast variable, we can assess the strength of a contrast from two aspects: one is based on the integration of both mean and variability of a contrast, represented by SMCV and the other is based on distributions in multiple conditions, represented by c^+ -probability. Moreover, because of the relationship between SMCV and c^+ -probability, we can classify the strength of contrast based on SMCV which simultaneously contains information from c^+ -probability. Based on SMCV, some key values of interest are 0.25, 0.5 and 1, which means that the average value of a contrast is one quarter, one half and one time of the standard deviation of the contrast variable. If one treats the mean of a variable as signal and standard deviation of a variable as noise, then the SMCV values of 0.25, 0.5 and 1 for a contrast variable means that the signal-to-noise ratio for the contrast variable representing a comparison is 25 percent, 50 percent and 100 percent respectively. Under normality, SMCV = 0.25, 0.5 and 1 indicates c^+ -probability = 0.60, 0.7 and 0.84 respectively. If based on c^+ -probability alone, one key value of interest for SMCV is 1.645. This is because SMCV = 1.645 corresponds to c^+ -probability = 0.95 under normality. In addition, 1.645 equals approximately $\sqrt{\frac{8}{3}}$, an important value in the relationship between SMCV and c^+ -probability for non-normal distributions. SMCV = 1.645 indicate that

the corresponding c^+ -probability is at least 0.918 when the contrast variable has a symmetric unimodal distribution, and is at least 0.836 when the contrast variable has a unimodal distribution. Therefore, based on both SMCV and c^+ -probability, some key values of interest for SMCV are 0, 0.25, 0.5, 1 and 1.645.

Based on the key values of SMCV and their corresponding values of c^+ -probabilities, it is reasonable to construct SMCV-based criteria for assessing the strength of a comparison. For example, when the mean of a contrast variable is zero (i.e., SMCV=0), c^+ -probability is 0.5. When SMCV is between 0 and 0.25, the signal-to-noise ratio is between 0 and 25 percent and the c^+ -probability is between 0.50 and 0.60, which is slightly above 0.50; thus the strength of the comparison (or effect size) is small. When SMCV is between 0.25 and 1, the signal-to-noise ratio is between 25 percent and 100 percent and the c^+ -probability is between 0.60 and 0.84; thus the strength of the comparison is medium. When SMCV is between 1 and 1.645, the signal-to-noise ratio is between 100 percent and 164.5 percent and the c^+ -probability is between 0.84 and 0.95; thus the strength of the comparison is large. Finally, when SMCV is greater than or equal to 1.645, the signal-to-noise ratio is at least 164.5 percent and the c^+ -probability is at least 0.95; thus the strength of comparison is extra large. This is for the positive direction. Similarly, we can classify the strength of comparison in the negative direction. The resulting SMCV-based classifying rule is shown in Panel A of Table 1.

The criteria in Panel A of Table 1 can be applied to any combination of random variables in which the interest is how far the combination is away from zero although here we focus on the linear combination of random variables. The concepts of SMCV and c^+ -probability accommodate various situations with or without independence and with or without variance homogeneity. Thus, if one uses SMCV only, one can use the criterion in the second column of Panel A in Table 1. This criterion should be applicable in a group comparison context with or without independence and with or without variance homogeneity and is consistent in various experiments in terms of signal-to-noise ratio. On the other hand, if one prefers to use c^+ -probability, one can use the criterion in the third column of Panel A in Table 1. Regardless of data distributions, this c^+ -probability-based criterion is consistent in various experiments in terms of the probability of a contrast variable being greater than zero. However, The correspondence between SMCV-based and c^+ -probability-based criteria listed in Table 1 relies on the normality assumption. Consequently, if one uses the c^+ -probability-based criterion in Table 1, corresponding values of SMCV may be adjusted according to the underlying nonnormal distributions. Similarly, if one uses the SMCV-based criterion in Table 1, corresponding values of c^+ -probability may be adjusted according to the underlying nonnormal distributions.

Two-group comparisons

For two-group comparisons, the difference of values between two groups is a contrast variable with coefficients (1, -1). Thus, SMCV becomes strictly standardized mean difference (SSMD), i.e., the ratio of mean to standard deviation of a difference, and c^+ -probability becomes d^+ -probability, i.e., the probability that a difference is positive [35-37,41]. The d^+ -probability is equivalent to the well-known common language effect size indicator (CL) in the 2-group case [21] or $P(X > Y)$ [1, 28, 42]. By replacing SMCV and c^+ -probability with SSMD and d^+ -probability, we can obtain the relationships between SSMD and d^+ -probability described in [38].

Huberty [13] pointed out, it appears that the only cutoffs of effect size to which applied researchers have paid attention are those

standards initiated by Cohen [5]: Cohen's d being 0.20 for small effects, 0.50 for medium effects, and 0.80 for large effects. Later, McLean [22] suggested the following criteria: Cohen's d being 0.50 for small effects, between 0.50 and .00 for moderate effects, and above 1.00 for large effects. The original and probability meanings of SSMD can give clear interpretations to Cohen's criterion. The relationship between Cohen's d and SSMD is Cohen's $d = \sqrt{2}$ SSMD [41]. Based on this relationship and d^+ -probability= Φ (SSMD) under normality, Cohen's $d = 0.20, 0.50, 0.80$ correspond to SSMD = 0.1414, 0.3536, 0.5656 and d^+ -probability=0.556, 0.638, 0.714 (Panel B of Table 1). Hence Cohen's small, medium and large effects have a signal-to-noise ratio of 14.14, 35.36 and 56.56 percent respectively and a d^+ -probability of 0.556, 0.638 and 0.714 respectively under normality. In other words, if one randomly draws one value from each of two groups, Cohen's small, medium and large effects means that the chance of the value from the first group is greater than the second group is 0.556, 0.638 and 0.714 respectively in the positive direction. Cohen's d being [0, 0.50), [0.50, 1.00) and [1.00, ∞) correspond to SSMD being [0, 0.3536), [0.3536, 0.7071), [0.7071, ∞) respectively and d^+ -probability being [0.50, 0.64), [0.64, 0.76), [0.76, 1] respectively (Panel C of Table 1). That is, Mclean's small, moderate and large effects have a signal-to-noise ratio between 0 and 35.36 percent, between 35.36 and 70.71 percent, and greater than 70.71 percent respectively. Based on the d^+ -probability, if one randomly draws one value from each of two groups, Mclean's small, moderate and large effects means that the chance of the value from the first group is greater than the second group is between 0.5 and 0.64, between 0.64 and 0.76, and greater than 0.714 respectively in the positive direction. Therefore, SSMD criterion gives interpretations to Cohen's and Mclean's criteria from both strength and probability perspectives. From Table 1, we can see that the small, medium and large effects based on Cohen's criterion are respectively small, medium and medium large effects based on SSMD. Mclean's small and medium effects are roughly SSMD's small and medium effects respectively. Mclean's large effects contain SSMD's medium large, large and extra large effects.

Contrast analysis in one-way ANOVA

Concepts: The contrast analysis in one-factor experiments aims at the comparison of multiple levels in one factor. Consider one-way ANOVA, in which t groups (or treatment levels) are used in contrast analysis, and the random variable representing the i^{th} group G_i has population mean μ_i and variance σ_e^2 . Then a traditional contrast is defined as $L = \sum_{i=1}^t c_i \mu_i$ where $\sum_{i=1}^t c_i = 0$; correspondingly a contrast variable is

$$V = \sum_{i=1}^t c_i G_i \quad (1)$$

Clearly, the mean of V equals its corresponding traditional contrast. The SMCV (λ) of V is the ratio of mean and standard deviation of contrast variable V .

In the situation where n_i observations Y_{ij} 's are obtained from the i^{th} group, the sample mean of Y_{ij} 's is \bar{Y}_i namely $\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ and the mean square of errors (MSE) is $MSE = \frac{1}{v_e} \sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ where $v_e = N - t$ is the degree of freedom, namely and N is the total number of observations. Then the estimate of traditional contrast L is $\hat{L} = \sum_{i=1}^t c_i \bar{Y}_i$. A simple estimate of SMCV λ is the method-of-moment (MM) estimate

$$\text{which is } \hat{\lambda}_{MM} = \frac{\sum_{i=1}^t c_i \bar{Y}_i}{\sqrt{MSE \sum_{i=1}^t c_i^2}} \quad (2)$$

A better estimate is the uniformly minimal variance unbiased estimate (UMVUE) [39], which is

$$\hat{\lambda}_{UMVUE} = \frac{\sqrt{K}}{\sqrt{v_e}} \frac{\sum_{i=1}^t c_i \bar{Y}_i}{\sqrt{MSE \sum_{i=1}^t c_i^2}} \quad \text{Where } K = 2 \cdot \left(\frac{\Gamma\left(\frac{v_e}{2}\right)}{\Gamma\left(\frac{v_e-1}{2}\right)} \right) \approx v_e - 1.48 \quad (3)$$

It can be shown that

$$T = \frac{\sum_{i=1}^t c_i \bar{Y}_i}{\sqrt{MSE \sum_{i=1}^t c_i^2 / n_i}} \sim \text{noncentral } t(v_e, b\lambda) \quad \text{where } b = \frac{\sum_{i=1}^t c_i^2}{\sum_{i=1}^t c_i^2 / n_i} \quad [39, 40].$$

This non-central t-distribution can be used to construct the CI of SMCV [39, 40].

The above method of contrast variable and SMCV can be applied to any contrast in one-way ANOVA although different contrasts may have different coefficients. After constructing contrast variable, we can apply the same classifying rule in Table 1 to assess the strength of any contrast. For example, the effect in the i^{th} group is commonly

Treatment	SMCV and 95% CI	C+-probability	SES and 95%CI	$\hat{\alpha}$ and 95% CI	p-value for α_i
Age 11	-0.42 (-1.05, 0.21)	0.34	-0.38 (-0.94, 0.19)	-15 (-37.6, 7.6)	0.188
Age 12	-0.28 (-0.90, 0.34)	0.39	-0.25 (-0.81, 0.31)	-10 (-32.6, 12.6)	0.378
Age 13	0 (-0.62, 0.62)	0.5	0.00 (-0.56, 0.56)	0 (-22.6, 22.6)	1
Age 14	0.28 (-0.34, 0.90)	0.61	0.25 (-0.31, 0.81)	10 (-12.6, 32.6)	0.378
Age 15	0.42 (-0.21, 1.05)	0.66	0.38 (-0.19, 0.94)	15 (-7.6, 37.6)	0.188

Table 2: The estimates and 95% confidence intervals (CI) of standardized mean of contrast variable (SMCV), standardized effect size (SES) and effect (i.e., α_i) at each age level in the Rosenthal, Rosnow and Rubin's example about psychomotor skills.

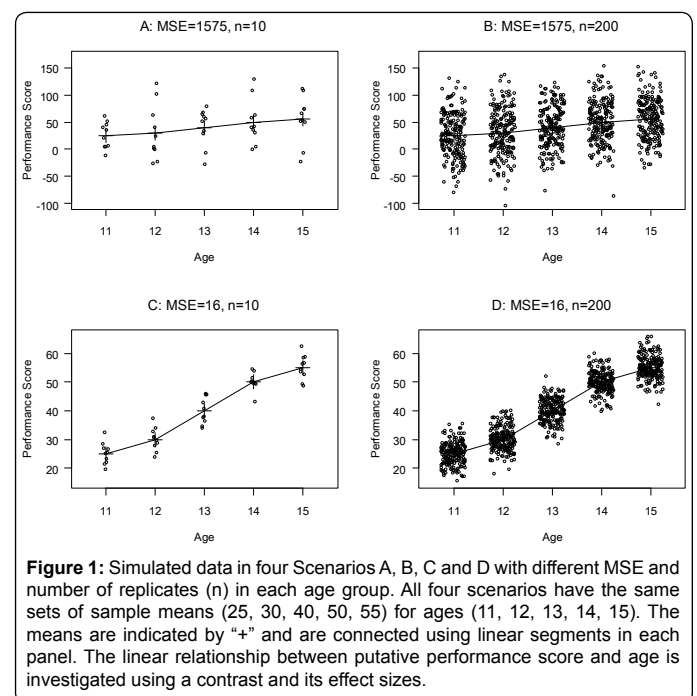


Figure 1: Simulated data in four Scenarios A, B, C and D with different MSE and number of replicates (n) in each age group. All four scenarios have the same sets of sample means (25, 30, 40, 50, 55) for ages (11, 12, 13, 14, 15). The means are indicated by "+" and are connected using linear segments in each panel. The linear relationship between putative performance score and age is investigated using a contrast and its effect sizes.

defined as $\alpha_i = \mu_i - \mu$ where μ is the mean of μ_i $S \alpha_i$ is the mean of contrast variable $V_e = \sum_{i=1}^t c_i G_i$ where $c_k = \begin{cases} 1 - \frac{1}{t}, & \text{When } k = i \\ -\frac{1}{t}, & \text{When } k \neq i \end{cases}$. Thus the size of α_i effect for the i^{th} group can be estimated using formula 2 or 3. The average magnitude of effects in all t levels is estimated to be the average value of $|\hat{\lambda}_i|$ namely $\frac{1}{t} \sum_{i=1}^t |\hat{\lambda}_i|$. Correspondingly, the average absolute value of effects in all t levels is estimated as $\frac{1}{t} \sum_{i=1}^t |\bar{Y}_i|$. The confidence interval of SMCV can be obtained using a non-central t-distribution [39].

Currently, standardized effect size (SES) is used to assess the size of treatment effects. SES at the i^{th} level is defined as $\psi_i = \frac{\alpha_i}{\sigma_e}$ and overall SES Ψ is $\psi = \sqrt{\frac{1}{t} \sum_{k=1}^t \left(\frac{\alpha_k}{\sigma_e} \right)^2}$ [29]. Their estimates are $\hat{\psi}_i = \frac{\hat{\alpha}_i}{\sqrt{MSE}}$ and $\hat{\psi} = \sqrt{\frac{1}{t} \sum_{k=1}^t \left(\frac{\hat{\alpha}_k}{\sqrt{MSE}} \right)^2}$ respectively. Clearly, $\psi_i = \lambda_i \sqrt{\sum_{k=1}^t c_k^2}$ and $\hat{\psi}_i = \hat{\lambda}_i \sqrt{\sum_{k=1}^t c_k^2}$ where c_k 's are the coefficients of the contrast representing the treatment effect at the i^{th} level. Recently, Bonett [3] has proposed a more general term: standardized linear combination of means $\delta = \frac{\sum_{i=1}^t c_i \mu_i}{\sqrt{\frac{1}{t} \sum_{i=1}^t \sigma_i^2}}$ which is an extension of SES from homoscedasticity to heteroscedasticity.

Rosenthal, Rosnow and Rubin [24] proposed four effect size correlations r_{alerting} , r_{contrast} , $r_{\text{effectsize}}$ and r_{BESD} to measure the magnitude of experimental effects. The definition of these four effect size correlations are as follows.

$$r_{\text{alerting}} = \sqrt{\frac{F_{\text{contrast}}}{F_{\text{contrast}} + F_{\text{noncontrast}} \cdot df_{\text{noncontrast}}}}, \quad r_{\text{contrast}} = \sqrt{\frac{F_{\text{contrast}}}{F_{\text{contrast}} + df_{\text{within}}}},$$

		Scenario A: MSE=1575, n=10	Scenario B: MSE=1575, n=200	Scenario C: MSE=16, n=10	Scenario D: MSE=16, n=200
Contrast Mean L	p-value	0.0498	0	0	0
	Estimate	80	80	80	80
	95% CI	(0.067, 159.93)	(62.59, 97.41)	(71.94, 88.06)	(78.24, 81.76)
r_{alerting}		0.9923	0.9923	0.9923	0.9923
r_{contrast}		0.2878	0.2748	0.9481	0.9431
$r_{\text{effectsize}}$		0.2876	0.2746	0.9415	0.9366
r_{BESD}		0.2794	0.2744	0.8944	0.4017
SES	Estimate	2.0158	2.0158	20	20
	95% CI	(0.0011, 4.009)	(1.568, 2.463)	(15.42, 24.54)	(19.01, 20.98)
SMCV	Estimate	0.6375	0.6375	6.325	6.325
	95% CI	(0.0004, 1.268)	(0.496, 0.779)	(4.875, 7.760)	(6.012, 6.634)
	c^+ -probability	0.738	0.738	~1	~1
	Effect type	Medium large	Medium large	Extra large	Extra large

Table 3: Contrast mean (i.e., traditional contrast L), SMCV, SES and their confidence intervals (CI), correlational effect sizes for the contrast variable representing a linear relationship between performance score and age as well as average magnitude of effects at treatment levels in a simulation.

$$r_{\text{effect size}} = \sqrt{\frac{F_{\text{contrast}}}{F_{\text{contrast}} + F_{\text{noncontrast}} \cdot df_{\text{noncontrast}} + df_{\text{within}}}} \quad \text{and}$$

$$r_{\text{BEST}} = \sqrt{\frac{F_{\text{contrast}}}{F_{\text{contrast}} + F_{\text{noncontrast}} \cdot (df_{\text{noncontrast}} + df_{\text{within}})}}$$

where $df_{\text{within}} = v_e = N - t$, $df_{\text{noncontrast}} = t - 2$, $MS_{\text{contrast}} = \frac{\left(\sum_{i=1}^t c_i \bar{Y}_i \right)^2}{\sum_{i=1}^t (c_i^2 / n_i)}$,

$$F_{\text{contrast}} = \frac{MS_{\text{contrast}}}{MSE} \quad \text{and} \quad F_{\text{noncontrast}} = \frac{\sum_{i=1}^t \left(n_i \left(\bar{Y}_i - \frac{1}{t} \sum_{i=1}^t \bar{Y}_i \right)^2 \right) - MS_{\text{contrast}}}{MSE \cdot df_{\text{noncontrast}}}. \quad \text{See}$$

Rosenthal, Rosnow and Rubin [24] for more details about these four effect size correlations.

A case study and simulation

An example: A study was conducted to investigate psychomotor skills against ages [26]. In this study, the sample means of performance scores of ten children at each of five age levels (11, 12, 13, 14, 15) were 25, 30, 40, 50 and 55 respectively, and the MSE was 1575. The

contrast for level effect at age 11 has coefficients $\left(\frac{4}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right)$

. The value of SMCV for this contrast is estimated to be -0.42 and corresponding c^+ -probability is 0.34, which suggests that the effect at the level of age 11 is medium and negative. The c^+ -probability of 11 0.34 indicates that, if one value is drawn from each age group, the chance that the value at age 11 is larger than the average value of the 5 draws at the remaining ages is 34%. Using the noncentral t-distribution described above, the 95% confidence interval (CI) of SMCV of this contrast is (-1.049, 0.209). Similarly we can investigate the effect at other age levels. We also calculated SES and its 95% CI. The results are displayed in Table 2. In this example, the values of SES are all very close to their corresponding SMCV values. This is because the square root of the sum of squared coefficients for each treatment level is close to 1 in this example.

A simulation study: to see how strong the linear relationship between performance score and age is, four scenarios A, B, C and D were simulated using normal distributions and transformation based on the psychomotor example (Figure 1). The age groups have the same sample means (25, 30, 40, 50, 55) in each scenario whereas the MSE and sample size vary (the first row of Table 3). Although all four scenarios have the same set of group means, the simulated data displayed in Figure 1 clearly show that the linear relationship between the putative performance score and age is not large in Scenarios A and B but extra large in Scenarios C and D. This is because the simulated studies have large data variability (i.e., data variance being 1575) in Scenarios A and B but much smaller data variability (i.e., data variance being 16) in Scenarios C and D.

The contrast variable representing a linear relationship between performance score and age has coefficients (-2, -1, 0, 1, 2). For this linear relationship, in each scenario, we calculated contrast mean (i.e., tradition contrast L), SES, SMCV and their 95% confidence intervals as well as γ_{alerting} , γ_{contrast} , $\gamma_{\text{effectsize}}$ and γ_{BESD} that are proposed by Rosenthal, Rosnow and Rubin [26]. The SMCV $\hat{\lambda}_{\text{linear}}$ of this contrast variable is estimated to be 0.6375 in scenarios A and B, which indicates that the contrast for the linear relationship has a medium large positive effect; in other words, the study shows that the performance score

has a medium large positive linear relationship with age. The 95% confidence interval of SMCV of this contrast is (0.0004, 1.268). The estimate of SES and its 95% CI for the linear relationship are 2.0158 and (0.0011, 4.009) respectively (Panel A of Table 3).

The traditional contrast L has the same value of 80 in all four scenarios; thus it cannot indicate the strength of the linear relationship. The 95% confidence interval of L becomes narrower each step from Scenarios A to D, but can hardly represent the strength of the linear relationship. If using p -value of testing $L=0$, we would conclude marginally significance in Scenario A and strong significance in the remaining three scenarios, which hardly offers useful information about the strength of the linear relationship in the data. Therefore, traditional contrast cannot provide useful information about the strength of the linear relationship. The results in Table 3 also show that (1) using γ_{altering} , the linear relationship is the same for all four scenarios; (2) using γ_{BESD} , the linear relationship is weak in Scenarios A and B, strong in Scenario C, and in between in Scenario D; (3) using γ_{contrast} or $\gamma_{\text{effectsize}}$, the linear relationship is weak in Scenarios A and B but very strong in Scenarios C and D; (4) using SMCV, the linear relationship is medium large in Scenarios A and B but extra large in Scenarios C and D. The results obtained using γ_{altering} and γ_{BESD} do not match with the observation and intuition about the strength of linear relationship displayed in Figure 1. SMCV, γ_{contrast} and $\gamma_{\text{effectsize}}$ give similar sensible judgment about the strength of the linear relationship.

The SES values in Scenarios A and B (both being 2.0158) is much smaller than those in Scenarios C and D (both being 20). Thus SES represents the relative strength of the linear relationship effectively. However, the values of SES are affected by the coefficients of contrasts; thus they are not comparable for contrasts with different coefficients. Consequently, there are not any consistent SES-based criteria for classifying the strength of contrasts with different coefficients. The SMCV values in Scenarios A and B (both being 0.6375) is smaller than those in Scenarios C and D (both being 6.325). The confidence intervals of SMCV in the scenarios with 10 per group are much wider than those in the corresponding scenarios with 200 per group. Thus, like SES, the sample size has the right impact on SMCV estimation: it affects the precision but not mean of SMCV estimation. Unlike SES, the value of SMCV is comparable to various contrast variables with different coefficients and has the same probabilistic meanings. Moreover, the values of corresponding c^+ probability provide additional information about the linear relationship: if one draws five values (y_1, y_2, y_3, y_4, y_5) each from one of the five age groups, the probability that a contrast variable representing the linear relationship (i.e., $-2y_1 - y_2 + y_4 + 2y_5$) is greater than 0 (i.e., no positive linear relationship) is 0.738 in Scenarios A and B and nearly 1 in Scenarios C and D.

Contrast analysis in multi-factor experiments

Concepts: In multifactor experiments with multiple levels in each factor, we use a random variable P_k to represent the random values in the l^{th} ($l = 1, \dots, n$) level of the k^{th} ($k = 1, \dots, K$) factor. Suppose a set of coefficients represent a comparison for the levels in the k^{th} factor. A contrast variable based on this set of coefficients is defined as $V = \sum_{l=1}^n c_l G_l$ where G_l is a random variable whose mean (and variance) equals the weighted mixture of means (and variances) of m combinations of factor levels containing the l^{th} level of the k^{th} factor.

Consider a two-factor experiment in which the two factors have I and J levels respectively as shown in Table 4. Let us use a random P_{ij}

variable to represent the random values in the combination of the i^{th} level of factor 1 and the j^{th} level of factor 2. Assume P_{ij} has mean μ_{ij} and variance σ_{ij}^2 . A contrast variable for a comparison in the levels of Factor 1 is

$$V = \sum_{i=1}^I c_i P_{i\cdot} \quad \text{Where } \sum_{i=1}^I c_i = 0 \quad (4)$$

$P_{i\cdot}$ is a random variable defined as $P_{i\cdot} = \sum_{j=1}^J \sqrt{w_{ij}} (P_{ij} - \mu_{ij}) + \sum_{j=1}^J w_{ij} \mu_{ij}$ where w_{ij} 's are the weights of means with the constraint of $\sum_{j=1}^J w_{ij} = 1$ for each i . Thus, its mean (and variance) equals the weighted mixture of means (and variances) of m combinations of levels in factors containing the i^{th} level of factor 1, i.e., $\mu_{i\cdot} = \sum_{j=1}^J w_{ij} \mu_{ij}$, $\sigma_{i\cdot}^2 = \sum_{j=1}^J w_{ij} \sigma_{ij}^2$ and $\text{cov}(P_{i\cdot}, P_{i'\cdot}) = \sum_{j=1}^J (w_{ij} \text{cov}(P_{ij}, P_{i'j}))$. Let the mean and variance of V are μ_V and σ_V^2 respectively. Then the SMCV of V is

$$\lambda_V = \frac{\mu_V}{\sigma_V} = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_i w_{ij} \mu_{ij})}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (c_i^2 w_{ij} \sigma_{ij}^2) + 2 \sum_{i=1}^I \sum_{i'=1}^I \sum_{j=1}^J (c_i c_{i'} w_{ij} \text{cov}(P_{ij}, P_{i'j}))}} \quad (5)$$

The corresponding c^+ probability is the probability that the contrast variable V based on $P_{i\cdot}$'s is greater than 0.

In situations where we are interested in mixture of m levels ($m \leq J$) of factor 2 with equal weights (i.e., for each $i, w = \frac{1}{m}$ if the j^{th} level is of interest and $w_{ij} = 0$ otherwise), if $P_{i\cdot}$'s are independent,

$$\lambda_V = \frac{\mu_V}{\sigma_V} = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_i w_{ij} \mu_{ij})}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (c_i^2 w_{ij} \sigma_{ij}^2)}} = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_i w_{ij} \mu_{ij})}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (c_i w_{ij}) \sigma_{ij}^2}} \cdot \frac{1}{\sqrt{m}} \quad (6)$$

As shown in Formula 4, the definition of a contrast variable for a comparison in levels of a factor is based on multiple random variables $P_{i\cdot}$'s which contain unknown parameter μ_{ij} . Thus, it is not easy to directly use Formula 4 to derive SMCV. In reality, we can use the standardized mean of a linear combination of $P_{i\cdot}$'s to calculate SMCV. Corresponding to the SMCV in Formula 6, we can construct a linear combination of $P_{i\cdot}$'s as follows:

$$U = \sum_{i=1}^I \sum_{j=1}^J c_{ij} P_{ij} = \sum_{i=1}^I \sum_{j=1}^J c_{ij} P_{i\cdot} \quad \text{where } c_{ij} = c_i w_{ij} \quad (7)$$

The standardized mean of the linear combination U is

$$\lambda_U = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_{ij} w_{ij} \mu_{ij})}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (c_{ij} w_{ij}) \sigma_{ij}^2}} \quad (8)$$

$$\text{Therefore, } \lambda_V = \lambda_U \cdot \frac{1}{\sqrt{m}} \quad (9)$$

Because U directly consists of factor levels and is easier to handle than contrast variable V , we can work on U to get statistical inference on parameters from V , based on formulas 7–9. For convenience, we use a term “contrast core” to refer to the comparison (represented by c_{ij} 's) embraced in a linear combination U . Meanwhile, we use another term “core number” to refer to the number of combinations with equal weights (i.e., m) [40].

In a two-way ANOVA, we usually assume $P_{i\cdot}$'s that are independent with equal variance σ_e^2 . In situations where we are interested in mixture of m levels ($m \leq J$) of factor 2 with equal weights (i.e., w_{ij}

$= \frac{1}{m}$ if the j^{th} level is of interest and $w_{ij} = 0$ otherwise), we have

$$\lambda_v = \frac{\mu_v}{\sigma_v} = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_{ij} w_{ij} \mu_{ij})}{\sigma_v \sqrt{\sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} = \frac{\sum_{i=1}^I \sum_{j=1}^J (c_{ij} w_{ij} \mu_{ij})}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (c_{ij} w_{ij})^2 \sigma_v^2}} \cdot \frac{1}{\sqrt{m}} \quad (10)$$

Based on the method of contrast variable and SMCV presented above, we can apply the same classifying rule in Table 1 to assess the strength of any contrasts in two-way ANOVA including main effect and interaction effect although different contrasts may have different coefficients. The treatment effect, main effect and interaction effect can be explored using traditional contrasts [2]. Now, each of them can be expressed as the mean of a contrast variable or a linear combination of random variables $\sum_{i=1}^I \sum_{j=1}^J c_{kl} P_{kl}$ where the coefficients c_{kl} 's are different for different effects. For treatment effect $\tau_{ij} = \mu_{ij} - \mu_{..}$ in the i^{th} level of factor 1 and the j^{th} level of factor 2, the coefficients are

$$c_{kl} = \begin{cases} 1 - \frac{1}{I \times J}, & \text{When } k = i \text{ and } l = j \\ -\frac{1}{I \times J}, & \text{When } k \neq i \text{ and } l \neq j \end{cases} \quad (11)$$

For main effect $\tau_{i.} = \mu_{i.} - \mu_{..}$ in the i^{th} level of factor 1, the core number is J and the coefficients are

$$c_{kl} = \begin{cases} \frac{1}{J} - \frac{1}{I \times J}, & \text{When } k = i \\ -\frac{1}{I \times J}, & \text{When } k \neq i \end{cases} \quad (12)$$

For main effect $\tau_{.j} = \mu_{.j} - \mu_{..}$ in the j^{th} level of factor 2, the core number is I and the coefficients are

$$c_{kl} = \begin{cases} \frac{1}{I} - \frac{1}{I \times J}, & \text{When } l = j \\ -\frac{1}{I \times J}, & \text{When } l \neq j \end{cases} \quad (13)$$

For interaction effect $\tau_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$ between the i^{th} level of factor 1 and the j^{th} level of factor 2, the coefficients are

$$c_{kl} = \begin{cases} 1 - \frac{1}{I} - \frac{1}{J} - \frac{1}{I \times J}, & \text{When } k = i, l = j \\ -\frac{1}{J} + \frac{1}{I \times J}, & \text{When } k = i, l \neq j \\ -\frac{1}{I} + \frac{1}{I \times J}, & \text{When } k \neq i, l = j \\ \frac{1}{I \times J}, & \text{When } k \neq i, l \neq j \end{cases} \quad (14)$$

Using contrast variable, SMCV and the coefficients in (11)–(14), we can use the classifying rule in Table 1 to provide a consistent interpretation to any effects that are commonly used in two-way ANOVA. And the same method can readily be extended to ANOVA with more than two factors.

Estimation and Inference of SMCV

A traditional contrast is defined as $L = \sum_{i=1}^I \sum_{j=1}^J c_{ij} \mu_{ij}$ where

$L = \sum_{i=1}^I \sum_{j=1}^J c_{ij} = 0$ and the corresponding linear combination of factor levels can be constructed as $U = \sum_{i=1}^I \sum_{j=1}^J c_{ij} P_{ij}$. Suppose this linear combination contains a contrast core (represented by c_{kl} 's) and core number m for a contrast variable of interest.

Consider a sample of size n_{ij} , $Y_{ij} = (Y_{ij1}, Y_{ij2}, \dots, Y_{ijn_{ij}})$ from the ij^{th} ($i = 1, 2, \dots, I, j = 1, 2, \dots, J$) treatment combination P_{ij} . Y_{ij} 's are independent.

Let

$$N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}, n_{i.} = \sum_{j=1}^J n_{ij}, v_e = \sum_{i=1}^I \sum_{j=1}^J (n_{ij} - 1), K = 2, \left(\frac{\Gamma\left(\frac{v_e}{2}\right)}{\Gamma\left(\frac{v_e - 1}{2}\right)} \right)^2 \approx v_e - 1.5, \bar{Y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}.$$

$$S_{ij}^2 = \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2, \text{ and } \hat{\sigma}_e^2 = MSE = \frac{1}{v_e} \sum_{i=1}^I \sum_{j=1}^J (n_{ij} - 1) S_{ij}^2 = \frac{1}{v_e} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2.$$

Then the estimate of traditional contrast L is $\hat{L} = \sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}$. It is well-known that the following properties hold in the situation where the $I \times J$ groups independently have normal distributions and equal variance $\hat{\sigma}_e^2$: (1) \bar{Y}_{ij} 's and S_{ij}^2 's are all independent with each other, (2) $v_e MSE / \sigma_e^2 \sim \chi^2(v_e)$, and (3) $(\bar{Y}_{11}, \bar{Y}_{12}, \dots, \bar{Y}_{IJ}, S_{11}^2, S_{12}^2, \dots, S_{IJ}^2)$ is a complete sufficient statistic of $(\mu_{11}, \mu_{12}, \dots, \mu_{IJ}, \sigma_e^2)$.

$Y_{ijk} \sim N(\mu_{ij}, \sigma_e^2)$, so

$$\bar{Y}_{ij} \sim N\left(\mu_{ij}, \frac{\sigma_e^2}{n_{ij}}\right), \sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij} \sim N\left(\sum_{i=1}^I \sum_{j=1}^J c_{ij} \mu_{ij}, \sigma_e^2 \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}\right) \text{ and}$$

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij} / \sqrt{\sigma_e^2 \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}} \sim N\left(\sum_{i=1}^I \sum_{j=1}^J c_{ij} \mu_{ij} / \sqrt{\sigma_e^2 \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}}, 1\right).$$

Considering $v_e MSE / \sigma_e^2 \sim \chi^2(v_e)$, and \bar{Y}_{ij} 's and S_{ij}^2 's are all independent, we have

$$T = \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}}} \sim \text{noncentral } t(v_e, b\lambda), \quad (15)$$

$$\text{where } b = \sqrt{m \cdot \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}{\sum_{i=1}^I \sum_{j=1}^J c_{ij}^2 / n_{ij}}}$$

Let $F_{\tau(V, b\lambda)}(\cdot)$ be the cumulative distribution function of noncentral $\tau(V, b\lambda)$ and T_{obs} be the observed value of T . Then we can find λ_L and λ_U such that $F_{\tau(V, b\lambda_L)}(T_{\text{obs}}) = 1 - \frac{\alpha}{2}$ and $F_{\tau(V, b\lambda_U)}(T_{\text{obs}}) = \frac{\alpha}{2}$ subsequently (λ_L, λ_U) is a $1 - \alpha$ confidence interval of SMCV λ .

The estimates of SMCV λ and its distribution can be derived as follows. Based on the above three properties for the independent, homoscedastic and normal model in two-way ANOVA,

$$E\left(\frac{\sqrt{K}}{\sqrt{v_e}} \cdot \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}}} \cdot \frac{1}{\sqrt{m}}\right) = \sqrt{K} \cdot E\left(\frac{1}{\sqrt{v_e MSE}}\right) \cdot \frac{E\left(\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}\right)}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \cdot \frac{1}{\sqrt{m}} = \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \mu_{ij}}{\sqrt{\sigma_e^2 \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}}} \cdot \frac{1}{\sqrt{m}} = \lambda,$$

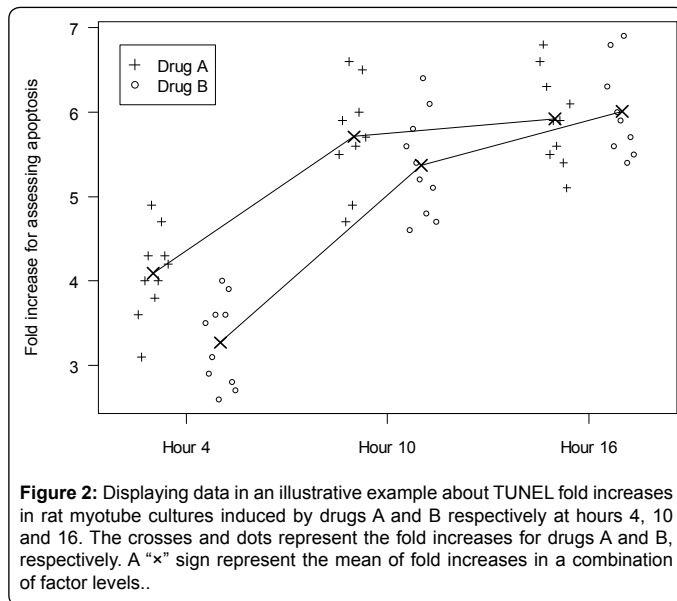


Figure 2: Displaying data in an illustrative example about TUNEL fold increases in rat myotube cultures induced by drugs A and B respectively at hours 4, 10 and 16. The crosses and dots represent the fold increases for drugs A and B, respectively. A "*" sign represent the mean of fold increases in a combination of factor levels..

where $n_{11}, n_{12}, \dots, n_{ij} \geq 2$. In deriving the above equations, I used

$$v_e MSE / \sigma_e^2 \sim \chi^2(v_e), \text{ and subsequently } E\left(\frac{1}{\sqrt{v_e MSE}}\right) = \frac{1}{\sigma_e^2 \sqrt{K}}.$$

$$\text{Set } \hat{\lambda} = \frac{\sqrt{K}}{\sqrt{v_e}} \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \frac{1}{\sqrt{m}}.$$

Then $\hat{\lambda}$ is an unbiased estimate of λ and is a function of complete sufficient statistic $(\bar{Y}_{11}, \bar{Y}_{12}, \dots, \bar{Y}_{IJ}, S_{11}^2, S_{12}^2, \dots, S_{IJ}^2)$ thus $\hat{\lambda}$ is a UMVUE of λ , namely

$$\hat{\lambda}_{UMVUE} = \frac{\sqrt{K}}{\sqrt{v_e}} \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \frac{1}{\sqrt{m}} \quad (16)$$

The MM and MLE estimates of SMCV are

$$\hat{\lambda}_{MM} = \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \frac{1}{\sqrt{m}} \quad (17)$$

$$\hat{\lambda}_{MLE} = \frac{\sqrt{N}}{\sqrt{v_e}} \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{Y}_{ij}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \frac{1}{\sqrt{m}} \quad (18)$$

$$\text{Considering } \hat{\lambda}_{UMVUE} = \frac{\sqrt{K}}{\sqrt{v_e}} bT, \hat{\lambda}_{MM} = bT, \hat{\lambda}_{MLE} = \frac{\sqrt{N}}{\sqrt{v_e}} bT$$

and $T \sim \text{noncentral } t(v_e, b)$, we have

$$\hat{\lambda}_{UMVUE} \sim \text{pnc } t(v_e, b\lambda, a_{UMVUE}), \hat{\lambda}_{MM} \sim \text{pnc } t(v_e, b\lambda, a_{MM}) \text{ and}$$

$$\hat{\lambda}_{MLE} \sim \text{pnct}(v_e, b\lambda, a_{MLE}),$$

where

$$a_{UMVUE} = \frac{\sqrt{K}}{\sqrt{v_e}} b \approx \frac{\sqrt{v_e - 1.5}}{\sqrt{v_e}} b, a_{MM} = b, a_{MLE} = \frac{\sqrt{N}}{\sqrt{v_e}} b \text{ and } \text{pnc } t(v, c, a)$$

is a proportional noncentral t-distribution with v degrees of

freedom, non-central parameter c and proportion a . The proportional noncentral t-distribution is defined as follows: if $T \sim \text{noncentral } t(v, c)$ then $aT \sim \text{pnc } t(v, c, a)$ [37].

In ANOVA, we usually assume that the values in each combination of factor levels independently have a normal distribution with equal variance. Consequently, any contrast variable also has a normal distribution. Therefore, we can plug the estimate of SMCV into the relationship $c^* - \text{probability} = \Phi(\lambda)$ to obtain an estimate of $c^* \text{probability}$ for contrast variable V .

An illustrative example

Here I construct an illustrative example based on Johnson et al's safety assessment study [18] to demonstrate how the method works. In this example, TUNEL staining fold increases were measured to assess apoptosis in rat myotube cultures treated with two drugs A and B respectively at 4, 10 and 16 hours. The data is listed in Table 5 and displayed in Figure 2.

This is an example of ANOVA with two factors, drug and time. The factor drug has two levels, drugs A and B, and the factor time has three levels, hours 4, 10 and 16. That is, $I=2$ and $J=3$. Let random variables $P_{11}, P_{12}, P_{13}, P_{21}, P_{22}, P_{23}$ represent the fold increases by drug A at hours 4, 10, 16 and by drug B at hours 4, 10, 16, respectively. Then the main effect for the drug can be assessed using a contrast variable

$$V_{1*} = \frac{1}{2} P_{1*} - \frac{1}{2} P_{2*}, \text{ where } P_{1*} \text{ and } P_{2*} \text{ are random variables constructed}$$

as P_{1*} in Table 4. Thus, P_{1*} is a random variable for the drug A whose mean (and variance) equals the equally weighted mixture of means (and variances) of 3 combinations of levels, i.e., drug A at hours 4, 10, 16, and P_{2*} is a similar random variable for drug B. The mean

and SMCV of V_{1*} can be obtained using the linear combination

$$U_{1*} = \frac{1}{6} P_{11} + \frac{1}{6} P_{12} + \frac{1}{6} P_{13} - \frac{1}{6} P_{21} - \frac{1}{6} P_{22} - \frac{1}{6} P_{23}. \text{ For } U_{1*}, \text{ the coefficient}$$

Populations		Factor 2				
		Level 1	Level 2	...	Level	Pooled
Factor 1	Level 1	P_{11}	P_{12}	...	P_{1J}	P_{1*}
	Level 2	P_{21}	P_{22}	...	P_{2J}	P_{2*}
	⋮	⋮	⋮	⋮	⋮	⋮
	Level I	P_{I1}	P_{I2}	...	P_{IJ}	P_{I*}
	Pooled	P_{*1}	P_{*2}	...	P_{*J}	

Note: P_{ij} has mean μ_{ij} and variance $\sigma_{ij}^2 - P_{i*} = \sum_{j=1}^J \sqrt{w_{ij}} (P_{ij} - \mu_{ij}) + \sum_{j=1}^J w_{ij} \mu_{ij}$

where $\sum_{j=1}^J w_{ij} = 1$ for each i . $P_{*j} = \sum_{i=1}^I \sqrt{w'_{ij}} (P_{ij} - \mu_{ij}) + \sum_{i=1}^I w'_{ij} \mu_{ij}$ where $\sum_{i=1}^I w'_{ij} = 1$ for each j .

Table 4: Random variables to represent values in each combination of levels in a two-factor experiment.

Fold increase	Hour 4	Hour 10	Hour 16
Drug A	3.6, 3.1, 4.0, 4.3, 4.9, 3.8, 4.0, 4.7, 4.3, 4.2	5.5, 5.9, 4.7, 6.6, 4.9, 5.6, 6.0, 6.5, 5.7	6.6, 6.8, 6.3, 5.5, 5.9, 5.6, 5.9, 5.4, 5.1, 6.1
Drug B	3.5, 2.9, 3.1, 3.6, 2.6, 4.0, 3.6, 3.9, 2.8, 2.7	5.6, 4.6, 5.8, 5.4, 5.2, 6.4, 4.8, 6.1, 5.1, 4.7	6.3, 6.8, 5.6, 6.0, 5.9, 6.9, 5.4, 5.7, 5.5
MSE: 0.316			
Means: A.Hour4	4.09	5.71	5.92
A.Hour10		3.27	5.37
B.Hour16			6.01

Table 5: Data in an illustrative safety assessment study for two drugs at three time points.

set is $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, -\frac{1}{6}, -\frac{1}{6}, -\frac{1}{6})$, the contrast core is $(\frac{1}{2}, -\frac{1}{2})$, and the core number is 3. Thus, the SMCV of V is estimated to be 0.442 and the confidence interval of $\lambda_{1\bullet}$ can be obtained using

$$T = \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \bar{y}_{ij\bullet}}{\sqrt{MSE \sum_{i=1}^I \sum_{j=1}^J \frac{c_{ij}^2}{n_{ij}}}} \sim \text{noncentral } t(v_e, b\lambda_{1\bullet})$$

where $v_e = 52$ and $b = 5.3785$. Similarly, we applied the method of contrast variable to investigate other main effects at levels of either factor and to explore linear and quadratic relationships. The results are listed in Table 6.

The results in Table 6 show that the SMCV of the contrast variable for the main effect of Drug A is estimated to be 0.442 and its corresponding c^+ -probability is 0.671. Thus the main effect of Drug A is medium; so is the main effect of Drug B. The average magnitude of drug effects is 0.442 which is a medium effect. In other words, the drug factor has a medium effect or the overall difference between Drugs A and B is medium. For each time level, the difference between Drugs A and B is large (i.e., $SMCV=1.017$) at Hour 4, medium (i.e., $SMCV=0.423$) at Hour 10 and small (i.e., $SMCV=-0.113$) at Hour 16. The SMCV's of contrasts for main effects at Hours 4, 10 and 16 are -2.968, 1.028 and 1.940 respectively. The corresponding c^+ -probabilities are 0.001, 0.848 and 0.974. These results indicate that the strength of main effects at Hours 4, 10 and 16 are extra large, large and extra large respectively (Table 6). The average magnitude of time effects is 1.979. Thus, the time factor has an extra large effect.

The SMCV is estimated to 2.834 for the contrast variable representing the linear relationship between time and fold increase, which indicates that this linear relationship is extra large. From the

coefficients and core numbers, the contrast variable representing the linear relationship also denotes the difference of fold increases between Hours 16 and 4. The SMCV is estimated to 1.028 for the contrast variable representing the quadratic relationship between time and fold increase, which indicates that this quadratic relationship is large. From the coefficients, the contrast variable representing the quadratic relationship also denotes the main effect at Hour 10, namely the difference between Hour 10 and the average of Hours 4 and 16. Judged by observing the data displayed in Figure 2, all the above conclusions about the strength of main effects and other contrasts obtained using SMCV are reasonable.

If we use the values of contrast means (namely traditional contrasts) and associated pvalues, we would conclude that all the contrast means listed in Table 7 are significant and we can hardly obtain useful information about the strength of comparison. One more case is that we know the contrasts with coefficients (-1, 2, -1, -1, 2, -1) and $(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, -\frac{1}{6}, \frac{1}{3}, -\frac{1}{6})$ respectively can both represent the quadratic relationship between fold increase and time. However, the estimated values of traditional contrast for these two contrasts (i.e., Contrasts D and G) are very different, i.e., 0.287 and 0.048, respectively, and the SES are also very different, i.e., 5.11 and 0.85 respectively. By contrast, the values of SMCV are the same (i.e., 1.028) for both contrasts (Table 6), which also indicates that the results reached using SMCV are reasonable, better than those reached using a traditional contrast and SES.

The correlational effect sizes including γ_{alerting} , γ_{contrast} , $\gamma_{\text{effect size}}$ and γ_{BESD} [26] were also calculated for each contrast. Like SMCV, multiplying a constant to the coefficients will not change the values of the correlational effect sizes. For example, for each of γ_{alerting} , γ_{contrast} , $\gamma_{\text{effect size}}$ and γ_{BESD} , the value is the same

Effect or Relationship	Coefficients of linear combination	Core Number	SMCV	95% confidence interval of SMCV	c^+ - probability	Strength Type
A: Drug A $\tau_{1\bullet}$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, -\frac{1}{6}, -\frac{1}{6}, -\frac{1}{6})$	3	0.442	(0.072, 0.821)	0.671	Medium
B: Drug B $\tau_{2\bullet}$	$(-\frac{1}{6}, \frac{1}{6}, -\frac{1}{6}, \frac{1}{6}, -\frac{1}{6}, \frac{1}{6})$	3	-0.442	(-0.821, -0.072)	0.329	Medium
Average drug effect magnitude			0.442			Medium
C: Hour 4 $\tau_{\bullet 1}$	$(\frac{1}{3}, -\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, \frac{1}{6})$	2	-2.968	(-3.733, -2.279)	0.001	Extra large
D: Hour 10 $\tau_{\bullet 2}$	$(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, -\frac{1}{6}, \frac{1}{3}, -\frac{1}{6})$	2	1.028	(0.548, 1.529)	0.848	Large
E: Hour 16 $\tau_{\bullet 3}$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, -\frac{1}{6}, -\frac{1}{6}, -\frac{1}{6})$	2	1.940	(1.377, 2.549)	0.974	Extra large
Average time effect magnitude			1.979			Extra large
F: Linear	(-1, 0, 1, -1, 0, 1)	2	2.834	(2.162, 3.578)	0.998	Extra large
G: Quadratic	(-1, 2, -1, -1, 2, -1)	2	1.028	(0.548, 1.529)	0.848	Large

Table 6: Contrast analysis using contrast variable for an illustrative example about safety assessment of Drugs A and B at hours 4, 10 and 16.

Effect or Relationship	L	p-value of testing $L=0$	SES	USES	r_{alerting}	r_{contrast}	$r_{\text{effectsize}}$	r_{BESD}
A: Drug A $\tau_{1\bullet}$	0.18	0.019	0.32	0.78	0.173	0.320	0.155	0.047
B: Drug B $\tau_{2\bullet}$	-0.18	0.019	-0.32	-0.78	-0.173	0.320	0.155	0.047
Average gender effect magnitude			0.32	0.78				
C: Hour 4 $\tau_{\bullet 1}$	-1.38	0	-2.46	-4.26	-0.953	0.880	0.847	0.643
D: Hour 10 $\tau_{\bullet 2}$	0.48	3×10^{-5}	0.85	1.47	0.330	0.539	0.293	0.093
E: Hour 16 $\tau_{\bullet 3}$	0.90	1×10^{-11}	1.61	2.78	0.623	0.770	0.554	0.208
Average time effect magnitude			1.77	2.84				
F: Linear	4.57	0	8.13	4.07	0.910	0.870	0.809	0.506
G: Quadratic	2.87	3×10^{-5}	5.11	1.47	0.330	0.539	0.293	0.093

Table 7: Contrast analysis using traditional contrast L and effect sizes for an illustrative example on safety assessment of two drugs at three time points.

		Time			
		Day 2	Day 4	Day 6	Combined
Drug	C	14.4	14.4	14.4	14.4
	D	10.6	10.6	10.6	10.6

Table 8: Population means for combinations of factor levels in a hypothetical experiment.

for contrasts D and G (Table 7). This is one advantage of these correlational effect sizes. As shown in the simulation in one-way ANOVA shown in Table 3, γ_{alerting} cannot take into account data variability. In that simulation study, γ_{contrast} approximately equals to $\gamma_{\text{effect size}}$ in each situation. However, the results of contrasts in Table 7 shows that, γ_{contrast} approximately equals to $\gamma_{\text{effect size}}$ only when $\gamma_{\text{effect size}}$ is large (above 0.8); the difference between γ_{contrast} and $\gamma_{\text{effect size}}$ is large when $\gamma_{\text{effect size}}$ is small. γ_{BESD} has much different value from γ_{contrast} and $\gamma_{\text{effect size}}$. It seems that $\gamma_{\text{effect size}}$ is the best among these four correlational effect sizes. However, it seems that the consistent meaning of $\gamma_{\text{effect size}}$ has yet to be discovered whereas SMCV has consistent meanings.

As demonstrated by the comparison of contrasts D and G in Tables 6 and 7, one issue of SES is that multiplying a constant to the coefficients of a contrast will change the SES value. For example, To address this issue, one argument for adopting SES is that before calculating SES, one should convert the coefficients c_{ij} 's into c_{ij}^* 's

$$\text{by } c_{ij}^* = \frac{c_{ij}}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J c_{ij}^2}} \text{ so that } \sum_{i=1}^I \sum_{j=1}^J c_{ij}^2 = 1. \text{ For convenience, let us use}$$

USES to denote SES based on the converted coefficients. The USES's of contrasts D and G are the same (i.e., 1.47) as one wishes (Table 7). However, the use of USES incurs another issue as demonstrated in the following example.

In this example, assume population means are known as shown in Table 8; MSE is 100 (or equivalently $\sigma = 10$) and sample size is 10 in each cell. It is trivial to derive that the main effect of Drug C is 1.9, corresponding USES is 0.4654. Now, if we combine the 3 days, we would have cell means shown in the last column of Table 8 and we would still have $\sigma = 10$ because there is no column effect. Then USES becomes 0.2687. Thus, USES produces anomalous behavior. By contrast, SMCV is 0.2687 (SES is 0.19) for the effect of Drug C either in the factorial experiment or in the combined days. Therefore, USES is problematic and cannot help the adoption of SES.

Conclusions and Discussions

For 2-group comparisons, the most commonly used probabilistic index for effect sizes may be McGraw and Wong's CL and the most commonly used ratios of mean difference to variability for effect sizes are Cohen's d , Glass's $\hat{\Delta}$ and Hedge's g . With the recently proposed concept of contrast variable and associated terms of SMCV and c^+ -probability, we can link all of them together. When the concept of contrast variable is applied to 2-group comparison, the contrast variable represents the difference between two groups. Subsequently, c^+ -probability becomes d^+ -probability and SMCV becomes SSMD. d^+ -probability is equivalent to CL [21] or $P(X > Y)$ [1, 17, 28, 42]. Meanwhile, Cohen's $d = \sqrt{2}$ SSMD in situations where there is no correlation between two groups. SMCV and c^+ -probability are two characteristics of a contrast variable and there is a strong link between SMCV and c^+ -probability. Therefore, a contrast variable and its two characteristics can link the commonly used effect sizes together and give interpretations to the strength of comparison from

two perspectives: a probabilistic index and a signal-to-noise ratio.

Based on SMCV and c^+ -probability, we have a rule for classifying the strength of a comparison in a contrast framework. The SMCV-based classifying rule can be consistently applied to any comparison in which the interest is in how far the comparison is away from zero. The common comparison is a contrast where coefficients sum to zero. It works effectively for either relationship or group comparison in either independent or correlated situations and in either two or more than 2 groups. As explored in this article, treatment effect, main effect, interaction effect, linear relationship, quadratic relationship and many other effects and relationships can all be addressed consistently using contrast variables and contrast-based classifying rule. The examples analyzed in the article also show that the results reached using contrast variable and SMCV-based classifying rule are sensible and matched with observations and intuitions from the data. The SMCV-based classifying rule also gives an interpretation to both Cohen's criterion and McLean's criterion. Based on these interpretations, if one prefers the use of Cohen's criterion or McLean's criterion, one can readily extend these criteria to any contrasts involving more than 2 groups by transforming Cohen's d to d^+ -probability and then to SMCV under normality.

As described above, SMCV and c^+ -probability have strong connections with commonly used effect sizes. However, SMCV and c^+ -probability have their own features that other effect sizes do not have. For example, although d^+ -probability is equivalent to CL in 2-group comparison, it is easier for one to guess its meaning from the word " d^+ -probability": the probability of a difference being positive. Moreover, it is much easier to extend d^+ -probability to n -group comparison with a consistent content than CL. When d^+ -probability is extended to n -group comparison, it becomes c^+ -probability which is still the probability of a linear combination of random variables (with coefficients summing to zero) being greater than zero. Now we only need to extend the number of groups from 2 to n . By contrast, when CL is extended to n -group comparison, it has a different meaning: the probability of a value from one group being greater than a value from the other groups [21].

In contrast analysis for factorial experiments, currently the most commonly used effect sizes may be SES's explored by Steiger [29] and correlational effect sizes proposed by Rosenthal et al [26,27] in addition to those described by Huberty [13] and Vacha-Haase and Thompson [34]. As shown in this article, SES effectively takes into account data variability and the accuracy of its estimation is robust to sample size. The value of SES for the contrast with the same coefficients is comparable across different experiments. The drawback is that the value of SES is not comparable for contrasts with different coefficients and timing a constant to a contrast may change the value of SES. The SES after transforming the coefficients may solve this drawback but generates new problems with the merging of factor levels. As to the correlational effect sizes, like SMCV, both γ_{contrast} and $\gamma_{\text{effect size}}$ have good properties that they capture data variability effectively and they are robust to sample size and coefficients. γ_{alerting} does not represent the size of relationship effectively due to lack of ability in capturing data variability and γ_{BESD} does not work effectively because it is not robust to sample size (Table 3). γ_{alerting} , γ_{contrast} , $\gamma_{\text{effect size}}$ and γ_{BESD} have a drawback that all variance-accounted-for effect sizes have. As Hedges and Olkin [11] pointed out, variance-accounted-for effect sizes are intuitively appealing but are not well suited for combination across studies because these indices are inherently nondirectional, and can have the same value even though the research studies exhibit substantively different results.

In summary, the concepts of contrast variable and associated terms of SMCV and c^+ -probability can link together the most commonly used probabilistic index for effect sizes such as McGraw and Wong's CL [21] and the most commonly used ratios of mean difference to variability such as Cohen's d [5]. A contrast variable can provide both a probabilistic meaning and an index of signal-to-noise ratio to interpret the strength of a comparison, which offers us a strong base to classify the strength of a comparison as shown in Panel A of Table 1. SMCV and c^+ -probability also give interpretations to both Cohen's and McLean's criteria [5,22]. The contrast variable, SMCV and c^+ -probability works effectively and consistently for either relationship or group comparison in either independent or correlated situations and in either two or more than 2 groups. Treatment effect, main effect, interaction effect, linear relationship, quadratic relationship and any other contrasts can all be addressed consistently using contrast variables. The examples in the article show that the results reached using contrast variables and the classifying rule are sensible and matched with observations and intuitions from the data. Therefore, contrast variable, SMCV, c^+ -probability and associated classifying rules may have the potential to offer a consistent interpretation to effect sizes.

Acknowledgement

The author would like to thank Drs. Daniel Holder, Keith Soper and Joseph Heyse for their support in this research.

References

- Senn S (2006) Probabilistic index: an intuitive non-parametric approach to measuring the size of the treatment effects by L. Acion, J. J. Peterson, S. Temple and S. Arndt. *Stat Med* 25: 591-602.
- Bird KD (2004) *Analysis of Variance via Confidence Intervals*. SAGE, Thousand Oaks, California.
- Bonett DG (2008) Confidence intervals for standardized linear contrasts of means. *Psychol Methods* 13: 99-109.
- Cohen J (1962) The statistical power of abnormal-social psychological research: A review. *J Abnorm Soc Psychol* 65: 145-153.
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*. (2nd edn), Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Davidson R, Duclos JY (2000) Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* 68: 1435-1464.
- Glass GV (1976) Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3-8.
- Glass GV, McGraw B, Smith ML (1981) *Meta-Analysis in Social Research*. SAGE.
- Goodman S (2008) A dirty dozen: Twelve P-value misconceptions. *Semin Hematol* 45: 135-140.
- Hedges LV (1981) Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Behav Stat* 6: 107-128.
- Hedges LV, Olkin I (1985) *Statistical Methods for Meta-Analysis*. Academic Press, San Diego.
- Hollander M, Wolfe DA (1999) *Nonparametric Statistical Methods*. (2nd edn), John Wiley & Sons, New York.
- Huberty CJ (2002) A history of effect size indices. *Educ Psychol Meas* 62: 227-240.
- Huberty CJ, Lowman LL (2000) Group overlap as a basis for effect size. *Educ Psychol Meas* 60: 543-563.
- Ioannidis JP (2005) Microarrays and molecular research: noise discovery?. *Lancet* 365: 454-455.
- Ioannidis JPA (2005) Why most published research findings are false. *Plos Medicine* 2: 696-701.
- Jing BY, Yuan JQ, Zhou W (2009) Jackknife Empirical Likelihood. *J Am Stat Assoc* 104: 1224-1232.
- Johnson TE, Zhang X, Bleicher KB, Dysart G, Loughlin AF et al. (2004) Statins induce apoptosis in rat and human myotube cultures by inhibiting protein geranylgeranylation but not ubiquitination. *Toxicol Appl Pharmacol* 200: 237-250.
- Kelley TL (1920) Measurement of overlapping. *Journal of Educational Psychology* 11: 458-461.
- Kirk RE (1996) Practical significance: A concept whose time has come. *Educ Psychol Meas* 56: 746-759.
- McGraw KO, Wong SP (1992) A Common Language Effect Size Statistic. *Psychol Bull* 111: 361-365.
- McLean JE (1995) *Improving education through action research: A guide for administrators and teachers*. Corwin Press, Thousand Oaks, CA.
- Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365: 488-492.
- Onwuegbuzie AJ, Levin JR (2003) Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect. *J Mod Appl Stat Methods* 2: 133-151.
- Rom DM, Hwang E (1996) Testing for individual and population equivalence based on the proportion of similar responses. *Stat Med* 15: 1489-1505.
- Rosenthal R, Rosnow RL, Rubin DB (2000) *Contrasts and Effect Sizes in Behavioral Research*. Cambridge University Press, Cambridge, United Kingdom.
- Rosenthal R, Rubin DB (2003) r (equivalent): A simple effect size indicator. *Psychol Methods* 8: 492-496.
- Senn S (1997) Testing for individual and population equivalence based on the proportion of similar responses. *Stat Med* 16: 1303-1306.
- Steiger JH (2004) Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychol Methods* 9: 164-182.
- Sterne JA, Davey Smith G (2001) Sifting the evidence - what's wrong with significance tests? *BMJ* 322: 226-231.
- Stine RA, Heyse JF (2001) Non-parametric estimates of overlap. *Stat Med* 20(2): 215-236.
- Timm NH (2004) Estimating effect sizes in exploratory experimental studies when using a linear model. *Am Stat* 58: 213-217.
- Tukey JW (1991) The philosophy of multiple comparisons. *Stat Sci* 6: 100-116.
- Haase TV, Thompson B (2004) How to estimate and interpret various effect sizes. *J Couns Psychol* 51: 473-481.
- Zhang XD (2007a) A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J Biomol Screen* 12: 645-655.
- Zhang XD (2007b) A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* 89: 552-561.
- Zhang XD (2008a) Genome-wide screens for effective siRNAs through assessing the size of siRNA effects. *BMC Res Notes* 1: 33.
- Zhang XD (2008b) Novel analytic criteria and effective plate designs for quality control in genome-scale RNAi screens. *J Biomol Screen* 13: 363-377.
- Zhang XD (2009) A method for effectively comparing gene effects in multiple conditions in RNAi and expression-profiling research. *Pharmacogenomics* 10: 345-358.
- Zhang XD (2010a) Assessing the size of gene or RNAi effects in multifactor high-throughput experiments. *Pharmacogenomics* 11: 199-213.
- Zhang XD (2010b) Strictly standardized mean difference, standardized mean difference and classical t-test for the comparison of two groups. *Stat Biopharm Res* 2: 292-299.
- Zhou W (2008) Statistical inference for $P(X < Y)$. *Stat Med* 27: 257- 279.