The XXXIVth Annual International Occupational Ergonomics and Safety Conference
Virtual Conference
September 15-16, 2022

Analysis-of-Variance (ANOVA) Assumptions Review: Normality, Variance Equality, and Independence

Alvah C. Bittner, PhD, CPE

BITTNER & ASSOCIATES
Kent, Washington, 98042-3532
USA

Corresponding author's Email: drbittner@comcast.net

Author Note: This report reflects the insights of several generations of authors, mentors, and colleagues including: H. Scheffé, T.S. Donaldson, R.J. Wherry Jr., S.J. Morrissey, G.E.P. Box, and most recently R.C. Carter, Jr.

Abstract: This report offers – *from an ergonomic analyst's standpoint* – means for testing and addressing concerns regarding three assumptions of ANOVA. Misapplications of "normality" tests are first illustrated, especially when before other considerations When applied, skewness (γ_1) and excess-kurtosis (γ_2) evaluations are recommended. Regarding these, skew (γ_1) and excess-kurtosis (γ_2) are explored for a balance one-way ANOVA. Of note the impacts of (γ_2) on ½Ln$\boldsymbol{F}(\nu_1,\nu_2)$ may – as seen in results developed for Row-Column product N – be didactically examined in terms of $[(1/\nu_1 + 1/\nu_2)/(1/\nu_1 + 1/\nu_2 + \gamma_2/N)]^{1/2}$. Importantly, this latter factor points out the amplification of nominal $\boldsymbol{\alpha}$-significance with negative γ_2, but the reverse with positive γ_2, (i.e., if apparently significant with positive γ_2 only real concern, it is!). After normality, variance inequality and nonindependence are individually then jointly considered. Ordinarily, where variance heterogeneity is suspect, a first consideration regards the possible relationship ($\boldsymbol{\sigma}$) between sample standard-deviations ($\hat{s}_i$) and means (i.e., ($\hat{u}_i$): $\hat{s}_i = \boldsymbol{\sigma}(\hat{u}_i)$?). This, if established, points toward a transformation of observations ($\Phi(y_i)$, i=1 to N) to achieve homogeneity (viz., $\Phi(y) = \int dy/\boldsymbol{\sigma}(y)$). Next considered is "nonindependence," the arguably most widely overlooked and unevaluated assumption. Certainly, *systematic* serial correlations between successive measures – which at +.10 impacts nominal $\boldsymbol{\alpha}$ by ~1.5x – may readily be accessed and adjustments made. However, more complicated, and rarely investigated, are "messy" secular shifts, e.g., among respectively university student-participants during semesters or union-hall workers hired on different weekdays. Shown herein, these and other dependencies may be evaluated – and largely controlled – when experimental conditions are collected in time-blocks, and these treated as "repeated-measures." This latter approach is complicated by the assumption of "variance-covariance-symmetry" (variance homogeneity perhaps earlier mitigated by transformation). Box $\boldsymbol{\varepsilon}$-correction methods that readily accommodate variance-covariance asymmetry are well-known and widely available in statistical packages.

It is concluded that: 1) Assessment of relationships between experimental condition means and standard-deviations should be first considered prior to ANOVAs, 2) Observation order blocking employed to address the possible nonindependence of observations (utilizing Box $\boldsymbol{\varepsilon}$-corrections), and 3) Slew (γ_1) and excess-kurtosis (γ_2) evaluated on final condition residuals.

Keywords: ANOVA, Assumptions, Review, Recommendations

1. Introduction

Analysis of Variance (ANOVA) offers an appropriate means for testing the significance of differences between means when it "*effectively*" meets its assumptions – *normality, variance equality, and independence.* Normality – and means for testing -- has been a recurring concern since 1929 (See Srivastava, 1958; Pearson, 1931; Demir, 2022). Likewise, nonindependence – in terms of significant serial correlations between successive measures – was noted as a potential issue by Student (1929) and shown to have potentially "serious effects" (Scheffé, 1999). Other typically ignored, but often equally serious, are systematic serial correlations that can affect worker performances as a function of time of day, e.g., reflecting increasing toxic exposure (e.g., Echeverria, Aposhian, Woods et al. (1998) or variations in students recruited across semesters (e.g., Nicholls, Loveless, Thomas, et al., 2015). Variance equality (homoscedasticity) considerations arguably also have their genesis in 1929 – given statistical relationship $t^2(\nu_2) = F(1, \nu_2)$ – with the first delineation of the two-group Behrens -Fisher Problem (Behrens, 1964/1929). There have been continuing efforts regarding the diverse solutions to the two-group problem over the decades (e.g., Chang & Pal, 2008) as well as numerous approaches toward addressing multiple group (Scheffé, 1999). This continuing focus is not surprising as – despite some robustness with large samples in the balanced two-group case – there can be serious effects if left unattended (Scheffé, 1999).

This report offers, *in its body,* this analyst's approach toward progressively framing and addressing concerns regarding the assumptions of ANOVA. Considered individually in turn – reflecting their order of historical focus – are normality, variance equality, and independence. Didactic examples will be used in consideration of early testing for normality, and results of a 4-group study will be utilized to illustrate considerations of aspects of the variance equality and independence assumptions.

2. Normality

Two aspects of normality addressed in this section: 1) When should it be addressed? and 2) What aspects are important (and when should they be of concern)?

2.1 When Should Normality Be Addressed?

Normality – as independence and homogeneity – is technically an assumption regarding the *errors-of-measurement* (Scheffe, 1999, p. 25). Obviously, these always may be appropriately addressed *after* any requisite transformations and other analytic adjustment procedures are applied. As many analysts, I initiate analyses by examining the data at hand with attention to within groups distributions including skew and kurtosis among basic statistics. This certainly provides an initial impression regarding normality and other assumptions, but its intent is to assure that the input data are *artefact-free* (e.g., no unspecified code-variables, buffer under- or over-flows, or other data quality issues.) After initial screening, it may appear opportune to test normality, but there are pitfalls prior to addressing other assumptions (especially with global tests).

Figure 1 illustrates histograms for two distributions (A&B) each of which is a composite of 200 cases and apparent nonnormality. Indeed, the first (1A) has an obvious negative excess kurtosis and may be found significantly non-normal by any of a variety of tests, including: "ocular" appropriate with the sample size at hand (e.g., Demir, 2022). However, 1A is in-fact the composite of two random samples of 100 from normal distributions with *equal variances, but different means* – demonstrating that early initial global normality assessments can be misleading even with ideally met assumptions when means differ. Figure 1B illustrates, in turn another example of clear non-normality, but involves only the overlaying of random samples (100 cases each) from two normal distributions that are both normally distributed and differ only in their variances [Note nonsymmetrical appearance reflects random variation]. Aggregating impacts (e.g., 1A&B), one may can begin to appreciate how there may appear substantial nonnormality in a global composite sample, but virtually none post disaggregation. Carter et al. (2022) presents a very recent example with profound nonnormality (e.g., $\gamma_1 > 2$, $\gamma_2 > 14$) in aggregate but far less ($\gamma_1 \sim 0.84$, $\gamma_2 \sim 2.83$) for disaggregated components.

The critical point is that normality testing (esp., global) – prior to disaggregation and systematic consideration of other assumptions – may be profoundly misleading.

2.2 What Aspects of Normality are Important and When of Concern?

Scheffé (1999. p. 332) made a summary observation that essentially has continued to hold over the last 60 years. Specifically:

> *[T]he value for [Excess] kurtosis γ_2, and to a lessor degree the value of γ_1 [skew], of the errors…are in the present state of knowledge, the most important indicators of the extent of which nonnormality affects the usual inferences made in the analysis of variance."*

Interestingly, the exploration of γ_1 and γ_2 impacts – much as earlier with Student (1927) then Pearson (1931) – has recently turned toward 1) ranges currently encountered in salient areas of study, and 2) impacts of these on α-significance (e.g., Blanca, Arnau, López-Montiel, et al., 2012; Blanca, Alercón, Arnau, et al., 2017). Remarkably, the recent results for Blanca and colleagues have been taken to suggest that normality is not a significant contemporary concern! This – though not considered by these authors or other contemporary authors we have reviewed – could very often be the case because contemporary researchers utilize metrics and methodologies selected over past generations to be most suitable for ANOVA (e.g., inverse transformation vs. raw reaction-times among others as suggested in Read, 1985). We, in following sections, will sequentially address the impacts of skew (γ_1) and excess-kurtosis (γ_2) as they certainly could be an issue in areas 1) certainly under-represented in past surveys (e.g., ergonomic) and 2) yet to be explored by future researchers for later emergent measures.

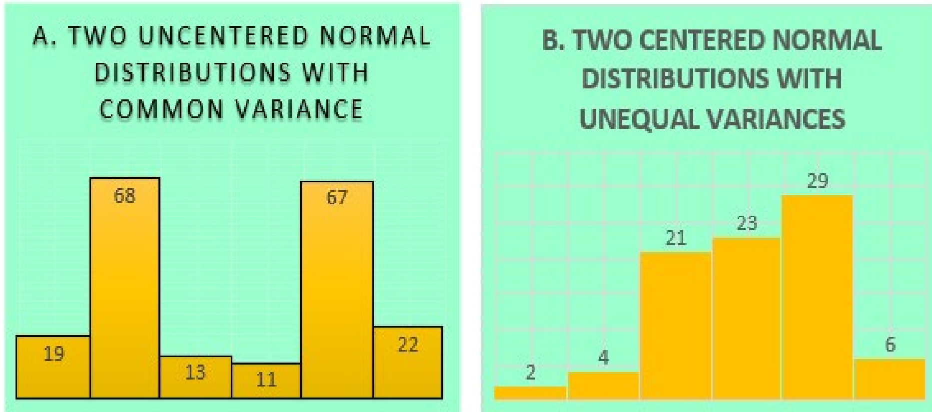

Figure 1. Illustrative Histograms

2.2.1 Skewness (γ_1) Impacts

Skewness ($\boldsymbol{\gamma_1}$) can have substantial impacts on Type-1 α-error, when *directionally testing* a single sample against a criterion-threshold value, but historically has been found negligible for 2-tailed tests (or equivalently ***F***) for groups with common $\boldsymbol{\gamma_1}$ and variances (Scheffé, 1999, p., 346 re: Pearson, 1931 etc.). This latter result is often surprising even to long-experienced statisticians, until their noting that difference in means *always has a zero (γ_1 =0) skew, when drawn from two population with common skew and finite variances (*Scheffé, p., 346). Adding to robustness, the skew for differences between means of independent samples will approach zero with the square-root inverse of their mutual sample size. This suggests – owing to the much larger impacts of variance inequalities (e.g., Scheffé, Table 10.2.3) – a first consideration of transformations that stabilize variances before impacts of skew and kurtosis. This is supported by historic observations that variance stabilizing transformations often enhance normality (e.g., Read, 1985, "Transformations to normality- pp 355-357; Carter et al., 2022).

2.2.2 Excess Kurtosis (γ_2)

Kurtosis ($\boldsymbol{\gamma_2}$) can have substantial impacts on Type-1 α-error, but its influence is directional: *negative ex. kurtoses leading to actual Type-1 errors greater than nominal α; whereas, positive kurtosis conservatively leading to less than nominal*. Interestingly, a hint of this was noted in the analytic results (Box & Anderson, 1953) – for Pearson- and Edgeworth-Distributions shown jointly for skew-squared and ex-kurtosis over respective ranges of 0-1 and -1, 0, +1 for 1-way ANOVAS with 5 groups of 5 observations each (Summarized in Scheffé, 1999, Table 10.3.2). Here, actual α *was* in range of .052-.053 for $\boldsymbol{\gamma_2}$ **= -1,** .049-.051 for $\boldsymbol{\gamma_2}$ = 0, and .048-.049 for γ_2 = +1, virtually constant across skews (negligible skew impacts not surprising in light of our earlier discussion, Sec 2.2.1).

Donaldson (1968) – aware of this and with support from past and his own extensive simulation studies – attributed the result to the general correlation induced between the ***F-ratio's*** numerator and denominator by γ_2. Ted's verbal summary argument being that "Positive Kurtosis (γ_2>0) induces a positive correlation, so that random deviations in the numerator, under the null hypothesis, are offset by a correlated increase in the denominator, and hence conservative; whereas opposite for $\gamma_2 < 0$ (Personal Communication, CSULA, 1967). Donaldson was unable to integrate the analytically established correlation into systematic power curves, which would be useful for his HF/E maintenance and operational research [these he consequently established via simulations in cases of particular interest].

We recently – in the context of a 1-way ANOVA – analytically revisited the effects of $\boldsymbol{\gamma_2}$ as imparting the Fisher-Z Distribution (where Z= ½Ln$\boldsymbol{F}(\nu_1,\nu_2)$ as classically delineated by Fisher (1924) and extensively explored by Aroian (1941). Interestingly, prior to extensive $\boldsymbol{F}(\nu_1,\nu_2)$ significance tables, the normal approximation provided by Z served as a means of testing the significance of $\boldsymbol{F}(\nu_1,\nu_2)$. In any case, noting that Z could be written as ½ the difference between Ln-transformed numerator and denominator, we were utilized a "total differential" error analysis approach (Cameron, 1982, p.449, Deming, 1943, pp 37-48) to derive a first order approximation of $\boldsymbol{Z}$'$(\nu_1,\nu 2,\gamma 2)$'s variance under differing $\boldsymbol{\gamma_2}$. Of note, a key in this derivation was the impact of $\boldsymbol{\gamma_2}$ on mean-square-terms (Scheffé, p. 336) and more specifically the differential impacts on numerator and denominator. We found the approximate impacts of $\boldsymbol{\gamma_2}$ on the relative variance of Z'$(\nu_1,\nu 2,\gamma 2)$ to be of the order of $(1/\nu_1 + 1/\nu_2)/(1/\nu_1 + 1/\nu_2 + \boldsymbol{\gamma_2}/N)]^{1/2}$. This latter factor – as Donaldson, but from a more didactic point of view – again points out the amplification of nominal $\boldsymbol{\alpha}$-significance with negative $\boldsymbol{\gamma_2}$, but the reverse with positive $\boldsymbol{\gamma_2}$. This suggests a rule-of-thumb: *"If apparently significant with group common positive $\boldsymbol{\gamma_2}$ the only real concern, then it is!*

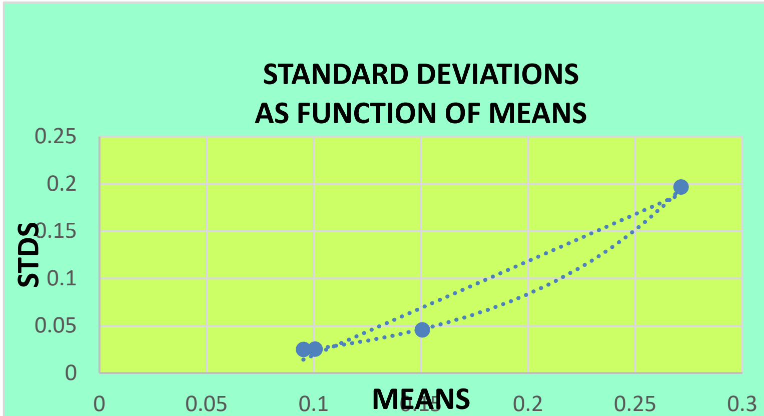

Figure 2. Sample Standard Deviations vs. Means

3. Variance Equality

ANOVAs in 2-group case – and isomorphic 2-tailed *t*-tests – have long been known to be remarkably "robust" to substantial ratios of group variances (≥ 5) when sample sizes are equal, but not so when groups differ in sample sizes (e.g., Scheffé, 1999, Table 10.2.3). Of special note, the paired t-test – where row differences are evaluated against zero – provides some insight into robustness as the variance of differences will always be homogenous. Also, though more conservative vs. most solutions with independent groups (Chang & Pal 2008; re: Behrens–Fisher problem), it also provides for control of row level dependencies (e.g., repeated measures. This joint robustness quality has a parallel with methods for addressing variance-covariance asymmetry that have particular utility when there may be non-independence issues [addressed later in Sec. 4.0].

ANOVAs with >2 Groups and unequal variances, don't share the same degree of robustness as the 2-group case (e.g., Scheffé, 1999, Table 10.4.2). This suggests – as we noted earlier in 2.2.1 – consideration of transformations that stabilize variances (and frequently also serve to markedly enhance normality, e.g., Read, 1985; Broemeling, 1982). Classically where variance heterogeneity is suspect (Scheffé, pp.364-365ff; Read, 1985), a first consideration regards the possible relationship ($\boldsymbol{\emptyset}$) between sample standard-deviations ($\hat{s}_i$) and means (i.e., ($\hat{u}_i$): $\hat{s}_i = \emptyset(\hat{u}_i)$?). This, when established, points toward a transformation of observations ($\Phi(y_i)$, i=1 to N) to achieve homogeneity (viz., $\Phi(y) = \int dy/\emptyset(y)$). In this report, we illustrate this process with data from an investigation of 4 Conditions (*C*s), each with 21 Participances (Ps) for total 84Ps.

Figure 2 shows the graphical exploration of two possible relationships between observed standard deviations and means for the 4 Groups in our illustrative study. As here, we recommend plotting standard-deviations vs. means as curvilinear relationships (cf., Read, 1985), aren't accommodated by Box-Cox or related systems (e.g., Broemeling, 1982) However, this is not the case in the current illustration with a healthy linear fit is healthy, but a Power-transformation ($\hat{s}_i = \hat{u}^k$; $1.99 \leq k \approx 2.01 < 2.03$) provides for an even better fit ($p<.05$ with 99% accounted variance). This latter relationship strongly suggested a reciprocal transformation (re: k = 2) and consequently this was applied in prior to further use of our illustrative data (Sec 3.2).

4. Independence

Non-independence initially appears a daunting consideration, due to the complex ways it may occur *or more importantly their impacts.* Reflecting on this, we will review (Sec 4.1) a cross section of classically studied examples, and then some oft-overlooked alternatives. This will set the stage for an illustration (Sec 4.2) of an "Repeated-measures ANOVA (RANOVA) approach for gaining some control and possible understandings of the sources of non-independence.

4.1 Non-independence Examples

Student (1927) early noted serial correlations between successive measures as a significant independence concern. As it happens, he found only positive 'serials' in his brewery setting, though, negatives can arise in other settings. In either case, serial impacts on *t*-test [and two group *F*s] can potentially be very "serious" (Scheffé, 1999, Table 10.2.2). For example, even a seemingly trivial +.1 "serial" can – for large Ns and nominal .05 α -- result in actual of .074 (and conversely .028 with a serial -.1). Serial correlations, not surprisingly, can have similar serious impacts on multigroup ANOVAS, but with differential factor impacts in 2-way and higher order analyses (Scheffé, Table 10.5.1). Equally serious are *hidden/unrecognized/ignored* factors that also can systematically affect independence. Among these are sometimes messy combinations of: *Time-of-day*, e.g., cumulative environmental exposures (e.g., toxic, Echeverria et al., 1998); *Weekday* (e.g., Borsch-Supan & Weiss, 2016) re: worker performance, or days workers were recruited at union-halls in our experience); *Cross-Month Effects* (e.g., students recruited across semesters, e.g., Nicholls et al., 2015, and/or *Time-of-Year*.

Fortunately, the above range of nonindependences – as indicated earlier – may be addressed given appropriate experimental controls and selective utilization of RANOVA.

4.2 Illustration of Nonindependence Assessment and Control via RANOVA

Table 1 summarizes an extended repeated-measures analysis of reciprocally transformed data (following Sec 3) for our earlier introduced illustrative study: 4 Conditions (*C*s) evaluated in blocks of one randomly assigned subject (*S)* daily for a Period (P) of 21 days. Treating daily blocks as if "repeated-measures" – as in in the table – allows for an assessment of condition effects that are as with 2 group case (Winer, 1962, p.39) "…[relatively] *free of variation due to the unique but systematic effects associated with the elements themselves*." However, there is an implicit variance-covariance matrix assumption that arises with >2Cs as in our illustration (albeit, transformation mediated heterogeneity). In any case, various Box **ε**-correction methods, accommodating asymmetry via degrees-of-freedom (dF) reduction -- are widely available (Bramwell, Bittner, & Morrissey, 1992; Winer, 1962). Applying the most theoretically severe (**ε** = 1/3) to Conditions (with table nominal 6,60 in *dF*s), there *appears* to be both: highly significant effects for Conditions- (F(1,20 = 29.07; $p < 10^{-5}$) and Period (F(6.7,20) = 5.85, $p < 10^{-2}$). {Within Period, Days (D) appears significant ($p < 10^{-2}$) and suggestive of Borsch-Supan and Weiss (/2016).}

Table 1. RANOVA Nonindependence Assessment Summary

	SS	dF	MS	F	P
PERIOD (P)	251.48	20	12.57	5.85	<10^-2
DayofWeek (D)	*223.33*	*6*	*37.22*	*17.31*	*<10^-2*
Week (W)	*3.55*	*2*	*1.77*	*0.82*	*NS*
DxW	*24.61*	*12*	*2.05*	*0.95*	*NS*
CONDITION (C)	534.99	3	178.33	29.07	<10^-5
ERROR (PxC)	129.02	60	2.15		
TOTAL	915.49	83			

Analysts might believe, given forceful capturing as with the illustration, that nonindependence issues are resolved. We recommend, however, an offline check for any residual serial correlations as they can slightly escaped blocking when present.

5. Discussion and Conclusions

This report offered – *from this ergonomic analyst's viewpoint* – an approach for systematically considering three assumptions of ANOVA: *Normality, Variance Equality, and Independence*. For each of these, their impacts were: (1) reviewed in a historic context with particular regard to effects on probabilities of Type1 (α) Errors; and (2) Means for addressing concerns offered that this analyst had found to have particular utility.

Normality was first considered as it often is the initial concern of many researchers. Regarding when to consider, it is argued (Sec 2.2) that it should not be the first consideration and indeed assessments of raw data can be grossly misleading as illustrated with Figure 1. Normality, it is argued, should – in keeping with Scheffé (1999, p. 25). – be addressed in terms of the residual *errors-of-measurement* after considerations of variance-equality (e.g., Sec 4.1) and independence (e.g., 4.2). Then, if the observed (γ_1,γ_2) were found within the set of those evaluated by Blanca et al. (2017), one as they might be comfortable assuming no import. Even, if outside their range but γ_2>0, then one might be also comfortable given Donaldson's (1968) rule-of-thumb: "If apparently significant with group common positive γ_2 the only real concern, then it is!

Variance heterogeneity (Sec 4.1) and nonindependence (4.2) -- both with reviews showing potentially profound impacts on actual α – were sequentially considered using an illustrative data set. Initially, as illustrated by Figure 2, homogeneity was considered via the graphical relationships (***ø***) between sample standard-deviations ($\hat{s}_i$) and means for data as collected (i.e., ($\hat{u}_i$): $\hat{s}_i = ø(\hat{u}_i)$? As it happened, the best fit ($\hat{s}_i \approx \hat{u}_i^{-2}$) informed a reciprocal transformation relationship via $\Phi(y) = \int dy/ø(y)$ as classically delineated (e.g., Read, 1985; Scheffé, 1999). Of particular note, Read (1985) has observed that transformations selected by this means often tend to also "normalize" data. This, in addition to effective variance stabilization, appeared the case for our illustrative data after reciprocal transformations in preparation for nonindependence considerations (Sec 4.2). As delineated, nonindependence may be addressed via treatment of daily blocks of the transformed data as repetitions in RANOVA (see Table 1 Summary). As seen in the tabulated Fs and dFs, there appeared to be a large and highly significant effect across the 21 daily Periods – these essentially reflecting a substantial weekly cycle. Clearly, substantial confounding variance was captured by daily blocking and treatment of blocks of randomly assigned subjects as repeated measures Broadly, the conditions appeared profoundly significant ($p<10^{-5}$), after blocking provided control.

It is concluded – based on the body of this report – that: 1) Assessment of relationships between experimental condition means and standard-deviations should be first considered prior to ANOVAs, 2) Observation order blocking may be employed to address the possible nonindependence of observations (utilizing Box **ε**-corrections), and 3) Slew (γ_1) and excess-kurtosis (γ_2) evaluated on final condition residuals.

6. References

Aroian L.A. (1941). A study of R. A. Fisher's z distribution and related F distribution. *Ann, Math. Stat, 124*, 429-444.

Behrens. W.-U. (1964). The Comparison of Means of Independent Normal Distributions with Different Variances *Biometrics; 20(1)*, 16-27 https://doi.org/10.2307/2527615 {English reprint 1929 paper originally in German]

Blanca M.J., Arnau, J., López-Montiel, D., Bono, R. & Bendayan, R. (2012). Skewness and kurtosis in real data samples. *Euro J. of Res Meth for Behah. Social. Sci.*, *10*, 1614-1641.

Blanca M.J., Alercón R, Arnau, J & Bono, R. (2017). Non-normal data: Is ANOVA still a valid option. *Psicothema; 29*(4), 552-557. *Doi*: 10.7334/psicothema2016.383.

Borsch-Supan A. & Weiss M. (2016). Productivity and the age composition of work teams: Evidence from the assembly line. *J. Econom. Ageing, 7*. 30-42.

Bramwell, A.T., Bittner, A.C. & Morrissey, S.J. (1992). Repeated measures analysis: Issues and options. *International Journal of Industrial Ergonomics*, *10*, 185-197.

Broemeling L.D. (1982). Box and Cox Transformation. In: S. Kotz, N.L. Johnson & C.B. Read (eds). *Encyclopedia of Statistical Sciences, Vol* 1 (pp 306-307). New York, NY: Wiley & Sons

Cameron J.M. (1982). Error analysis. In: In: S. Kotz, N.L. Johnson & C.B. Read (eds). *Encyclopedia of Statistical Sciences, Vol* 2 (pp 545-551). New York, NY: Wiley & Sons

Carter R.C., Garcia P.A., Melgosa M & Brill M.H. (2022). Metrics of color-difference formula improvement. *J. Optical Soc of America; 6* (June, in press)

Chang, C.H., & Pal, N. (2008). A revisit to the Behrens–Fisher problem: comparison of five test methods. *Communications in Statistics—Simulation and Computation®*, *37*(6), 1064-1085.

Deming WE (1943). *Statistical Adjustment of Data.* New York, NY: Dover Publications.

Demir, S. (2022). Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients. *Intern. J. of Assess. To, 39(6ols in Educ; 9(2),* 397-409.

Donaldson, T.S. (1968). Robustness of the F-Test to Errors of Both Kinds and the Correlation between the Numerator and Denominator of the F-Ratio". *J. American Stat Assoc,* 1968*; 63,* 660–676.

Echeverria, D., Aposhian, H.V., Woods, J.S., Heyer, H.J., Aphosian, M.M., Bittner, A.C. & Mahurin, R.K. (1998). Neurobehavioral effects from exposure to dental amalgam Hg^0: New distinctions between recent exposure and Hg^0 body burden. *FASEB Journal*, 12, 971-980.

Fisher RA (1924). On a distribution yielding the error function of several well-known statistics. *Proceedings International Congress of Mathematics*, Toronto. *2*: 813.

Nicholls, M. E., Loveless, K. M., Thomas, N. A., Loetscher, T., & Churches, O. (2015). Some participants may be better than others: Sustained attention and motivation are higher early in semester. *Quart. J. of Exp. Psych*, *68*(1), 10-18.

Pearson, E.S. (1931). The Analysis of Variance in Cases of Nonnormal Variations. *Biometrika*, 1931; *23*, 114–133.

Read. C.B. (1985). Normal Distribution. In: S. Kotz, N.L. Johnson & C.B. Read (eds). *Encyclopedia of Statistical Sciences., Vol 8*, pp. 347-359. New York, NY : John Wiley & Sons

Scheffé, H. (1999). *The Analysis of Variance* [Wiley Classics Library Edition, 1999]. New York, NY: John Wiley & Sons

Srivastava, A.B.L. (1958). Effect of Non-Normality on the Power Function of t-Test. *Biometrika*, 1958;*45*(3/4), 421–430.

Student (1927). Errors of routine analysis. *Biometrika*;*19*, 151-164.

Winer B.J. (1962). *Statistical Principles in Experimental Design*. San Francisco, CA: McGraw-Hill.