Thailand Statistician
2014; 12(2): 113-134
http://statassoc.or.th
Contributed paper

Testing Statistical Agreement Based on Unbalanced Paired Data

Elizabeth Stanwyck [a], **Bimal K. Sinha** [a], **Barry D. Nussbaum** [b]
[a] Department of Mathematics and Statistics,
University of Maryland, Baltimore 21250, USA
[b] Office of Environmental Information,
U.S. Environmental Protection Agency, USA

Received: 23 July 2013
Accepted: 4 December 2013

Abstract

Assessing agreement between two methods of continuous measurement plays a vital role in deciding if one of the methods (newer or cheaper) can be adopted in future experiments. Assuming a bivariate normal distribution for the responses from the two methods, we derive the likelihood ratio test for a combined hypothesis of equality of means, equality of variances and a known high value of pairwise correlation based on *unbalanced* paired data. This situation arises when one observes multiple replications of one response (cheaper) for a specified single value of the other response (costlier) from sampled units. Our results provide a generalization of Yimprayoon, Tiensuwan and Sinha (2006). An example in the context of a USEPA application is highlighted.

Keywords: Agreement, Concordance correlation coefficient, Fisher's test, Likelihood ratio test, Unbalanced data.

1. Introduction

Assessing agreement between two methods of measurement plays a vital role in deciding if one of the methods (newer or cheaper) can be adopted in future experiments. There are numerous examples in clinical, medical and criminal trials that illustrate these situations requiring measurement of agreement between two data-generating sources [1]. Evaluation of agreement starting from the fundamental works of Cohen [2], [3] and Landis and Koch [4] for qualitative responses has recently received considerable attention when the response is continuous. In the latter case, methods based on correlation coefficient, regression analysis, paired t-tests, least squares analysis for intercept and slope, and within-subject coefficient of variation have been traditionally used. It is however the notion of concordance correlation coefficient (CCC) due to Lin [5] which provided a breakthrough for assessing agreement between two distinct methods for continuous data.

Lin [5] defines the degree of concordance or agreement (CCC) between two variables X and Y by the mean of their squared difference (MSD), and CCC as

$$CCC = \rho_C = 1 - \frac{E(X-Y)^2}{E_{indep}(X-Y)^2} = \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \qquad (1)$$

where E_{indep} represents expectation under independence of X and Y, and $\mu_x = E(X)$, $\mu_y = E(Y)$, $\sigma_x^2 = Var(X)$, $\sigma_y^2 = Var(Y)$, and $\sigma_{xy} = Cov(X,Y)$. CCC translates the $MSD = E(X-Y)^2$ into a correlation coefficient that measures the agreement along the identity line, and satisfies: $-1 \leq CCC \leq +1$ with -1 indicating perfect disagreement (or reverse agreement, $Y = -X$), 0 indicating no agreement, and $+1$ indicating perfect agreement ($Y = X$). A sample estimate of CCC based on paired data $[(x_i, y_i), i = 1, \cdots, n]$ is readily obtained by plugging in the standard estimates of the five parameters by their sample analogues.

Several generalizations of Lin [5]'s idea exist in the literature (Chinchilli et al. [6]; Vonesh et al. [7]; Vonesh and Chinchilli [8]; Barnhart [9]; King and Chinchilli [10, 11]; Lin [12]). Lin et al. [1] provided a review and comparison of various measures in this field, including CCC, by evaluating the powers of tests (i) $\mu_x = \mu_y$, (ii) $\sigma_x = \sigma_y$, and (iii) $\rho = \rho_0$ where ρ_0 is a given value. Yimprayoon et al. [13] extended the work of Lin et al. [1] by deriving the likelihood ratio test (LRT) of the combined hypothesis $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho = \rho_0$ where ρ_0 is a given value. In both Lin et al. [1] and Yimprayoon et al. [13], inference is based on usual paired data on X and Y under the assumption of a bivariate normal population.

The objective of this paper is to generalize the results in Yimprayoon et al. [13] in two directions: nature of null hypothesis and nature of data. We test the null hypothesis $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho \geq \rho_0$. We believe that, unlike in Yimprayoon et al. [13], the null hypothesis for ρ, namely, $\rho \geq \rho_0$ makes more sense. Regarding the nature of data, we deal with an *unbalanced* data of the type $[(x_i, y_{i1}, \cdots, y_{im_i}), i = 1, \cdots, n]$, representing multiple replications of Y corresponding to one observation on X. This situation may arise when one variable (Y) is much cheaper than the other variable (X), resulting in multiple observations of Y *all* corresponding to a single value of X. A recent USEPA study (California Environmental Protection Agency Air Resources Board Report [14]) providing an application of this scenario deals with demonstrating equivalence between primary and secondary methods for measuring formaldehyde emissions from composite wood products. The primary method is based on a large chamber test (generating x-values) while the secondary method on a small chamber test (generating y-values), and there are often three y-values corresponding to every x-value. However, the formaldehyde study used a more limited test for agreement based only on the mean and standard deviation of the sample differences $d_i = x_i - \bar{y}_i,\ i = 1, \cdots, n$ as if equality of population means and population variances would guarantee equivalence! That this is not true follows from a simple example of paired data of the type $[(10, 22), (15, 12), (18, 10), (25, 17), (17, 25), (22, 18), (12, 15)]$ for which $\bar{d} = 0$ and also $s_x^2 = s_y^2$ but with obvious severe discrepancies between the two datasets. Moreover, a high correlation by itself also does not guarantee equivalence because the variables may have an almost linear relation without their means and variances being equal. The dataset $[(10, 15), (15, 25), (18, 25), (20, 26), (25, 30), (30, 36)]$ for which $r_{xy} = 0.965, \bar{d} = -6.5, s_x^2 = 50.67$, $s_y^2 = 47.77$ demonstrates this fact. We will return to this EPA example in Section 4. The case when $m_1 = \cdots = m_n = m$ is referred to as *balanced*.

The paper is organized as follows. We derive the LRT of H_0 in Section 2 and provide its critical values by simulation in some cases. We also compute its power for some choices of the alternative hypothesis. In Section 3 we propose new tests based on combinations of p-values of the component tests under

$H_{01}: \mu_x = \mu_y$, $H_{02}: \sigma_x = \sigma_y$, and $H_{03}: \rho \geq \rho_0$. A comparison of the power of such tests with that of the LRT is also done. We return to two EPA datasets, including the one discussed above, in Section 4. An appendix contains some technical results.

2. Main Results

Case 1. We begin with the *restricted* dataset

$$[(x_i, \bar{y}_i), i = 1, \ldots, n] \tag{2}$$

where only the means of the Y variable are provided. The likelihood function can be based on the marginal likelihood of x_i, which is univariate normal $N(\mu_x, \sigma_x^2)$, and conditional likelihood of $\bar{y}_i$, given x_i, which is again univariate normal with conditional mean linear in x_i and conditional variance independent of x_i, namely,

$$\bar{y}_i | x_i \sim N[\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x_i - \mu_x), \frac{\sigma_y^2(1-\rho^2)}{m_i}] \tag{3}$$

resulting into the overall likelihood

$$\begin{aligned} L(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho | data) \quad \propto \quad & (\sigma_x \sigma_y)^{-n}(1-\rho^2)^{-n/2} \\ & \exp\Big[-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - \mu_x)^2}{\sigma_x^2} - \frac{1}{2\sigma_y^2(1-\rho^2)} \\ & \sum_{i=1}^{n} m_i(\bar{y}_i - \mu_y - \rho\frac{\sigma_y}{\sigma_x}(x_i - \mu_x))^2\Big]. \end{aligned} \tag{4}$$

Note that here since only $\bar{y}_i$ is available, the above likelihood represents the joint pdf of $[(x_i, \bar{y}_i), i = 1, \cdots, n]$. Let us define

$$\begin{aligned} A &= \sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad C = \sum_{i=1}^{n} m_i(x_i - \bar{\bar{x}})^2 \\ D &= \sum_{i=1}^{n} m_i(\bar{y}_i - \bar{\bar{y}})^2, \qquad E = \sum_{i=1}^{n} m_i(x_i - \bar{\bar{x}})(\bar{y}_i - \bar{\bar{y}}) \\ \bar{x} &= \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{\bar{x}} = \frac{\sum m_i x_i}{M}, \qquad \bar{\bar{y}} = \frac{\sum m_i \bar{y}_i}{M}, \qquad M = \sum m_i. \end{aligned} \tag{5}$$

Unrestricted maximization. It can be verified that the unrestricted maximization yields the following estimates:

$$\hat{\mu}_x = \bar{x}, \qquad \hat{\mu}_y = \bar{\bar{y}} + \frac{E}{C}(\bar{x} - \bar{\bar{x}}) \tag{6}$$

$$\hat{\sigma}_x^2 = \frac{A}{n}, \quad \hat{\sigma}_y^2 = \frac{1}{n}[D + \frac{MAE^2}{nC^2} - \frac{E^2}{C}], \quad \hat{\rho}^2 = \frac{E^2\hat{\sigma}_x^2}{C^2\hat{\sigma}_y^2}. \tag{7}$$

Plugging in equations (6) and (7) into (4) gives the following unrestricted maximum of the likelihood (apart from some constants):

$$\left[\frac{C}{A(DC - E^2)}\right]^{n/2}. \tag{8}$$

Details appear in the Appendix.

Restricted maximization. To maximize the likelihood under $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \ \ \rho \geq \rho_0$, we first fix $\rho \geq \rho_0$, and maximize it with respect to the common mean μ and common variance σ^2. A direct maximization of the likelihood with respect to μ yields

$$\hat{\mu}(\rho) = \frac{n\bar{x}(1+\rho) + M(\bar{\bar{y}} - \rho\bar{\bar{x}})}{M(1-\rho) + n(1+\rho)}. \tag{9}$$

Also, the MLE of σ^2 is given by

$$\hat{\sigma}^2(\rho) = \frac{A + n(\bar{x} - \hat{\mu})^2 + \frac{\sum_{i=1}^n m_i[\bar{y}_i - \hat{\mu} - \rho(x_i - \hat{\mu})]^2}{1-\rho^2}}{2n} \tag{10}$$

which can be simplified as $2n\hat{\sigma}^2(\rho) = Q_1(\rho)$ with

$$\begin{aligned} Q_1(\rho) \quad &= \quad A + \frac{D + C\rho^2 - 2E\rho}{1-\rho^2} \\ &\quad + \frac{nM[\bar{\bar{y}} - \bar{x} + \rho(\bar{x} - \bar{\bar{x}})]^2}{(1-\rho)[M(1-\rho) + n(1+\rho)]}. \end{aligned} \tag{11}$$

Details appear in the Appendix. It is easy to verify that the likelihood function, maximized with respect to μ and σ^2 for fixed ρ, simplifies to

$$L_1(\rho|data) \propto [(1-\rho^2)^{\frac{1}{2}} \times Q_1(\rho)]^{-n}. \tag{12}$$

To maximize the above expression with respect to $\rho \geq \rho_0$, which is equivalent to minimization of $U_1(\rho) = [(1-\rho^2)^{\frac{1}{2}} \times Q_1(\rho)]$, an analytical approach turns out to be quite difficult. However, one can use numerical computations to evaluate it. It then follows from (8) and (12) that the LRT statistic $\lambda_1(\rho)$ defined by

$$\lambda_1(\rho) = \frac{\sup_{H_0} L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}|\text{ data })}{\sup_{\text{unrestricted}} L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}|\text{data})} \tag{13}$$

is equivalent to rejecting H_0 for large values of T_1 given by

$$T_1 = [min_{\rho \geq \rho_0} U_1(\rho)] \times [\frac{C}{A(DC - E^2)}]^{\frac{1}{2}}. \quad (14)$$

Remark 1. Obviously, when the design is balanced, namely, $m_1 = \cdots = m_n = m$, $C = nA$, and hence the ratio $\frac{C}{A}$ can be dropped from T_1 along with some obvious changes in D, C and E.

It is easy to verify that, under H_0, T_1 is location and scale invariant. In view of the composite nature of the null hypothesis H_0 in terms of ρ, the cut-off point c_1 to be used for rejecting H_0 based on T_1 must satisfy the size condition for all values of $\rho \geq \rho_0$. Our simulation studies reveal that taking $\rho = \rho_0$ and determining the resulting cut-off point $c_1(\rho_0)$ does satisfy the size condition for values of ρ larger than ρ_0. We have simulated (null) values of T_1 under the bivariate normal distribution $N[0, 0, 1, 1, \rho_0]$ for various values of ρ_0, n and $m_1, \cdots, m_n$. Table 1 shows the cut-off points $c_1(\rho_0)$ for $\alpha = 0.05$, and Table 2 shows the estimated Type I error rates when $\rho > \rho_0$; notice that the estimated Type I error rates are always smaller than 0.05. To get an idea about the power of the LRT, we have taken seven kinds of alternatives: (i) means and variances are the same ($\mu = 0$, $\sigma = 1$), but $\rho < \rho_0$; (ii) means are unequal, but variances are the same ($\sigma = 1$) and $\rho = \rho_0$; (iii) means are the same ($\mu = 0$), but variances are unequal and $\rho = \rho_0$; (iv) means are unequal, variances are unequal, $\rho = \rho_0$; (v) means are unequal, variances are equal ($\sigma = 1$), $\rho < \rho_0$; (vi) means are equal ($\mu = 0$), variances are unequal and $\rho < \rho_0$; (vii) means are unequal, variances are unequal and $\rho < \rho_0$. Tables 7-13 in the Appendix provide the power of the LRT under the above alternatives. It turns out that the power is a maximum when the means are different, and a minimum when the correlation is smaller than ρ_0 or when the variances are not the same. Even so, when the sample size is moderate, the LRT performs reasonably.

Case 2. Here we assume that *unrestricted* or *full* data, namely, $[x_i, (y_{i1}, \cdots, y_{im_i})]$, $i = 1, \cdots, n]$ are available. The likelihood function given in (4) is modified as

$$\begin{aligned} L(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho | data) \quad \propto \quad & (\sigma_x)^{-n} [\sigma_y^2 (1 - \rho^2)]^{-M/2} \\ & exp\Big[-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu_x)^2}{\sigma_x^2} - \frac{1}{2\sigma_y^2 (1 - \rho^2)} \\ & \{ \sum_{i=1}^{n} m_i (\bar{y}_i - \mu_y - \rho \frac{\sigma_y}{\sigma_x} (x_i - \mu_x))^2 + W_y \} \Big] \quad (15) \end{aligned}$$

where $W_y = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$ is the within sum of squares of the y-values. The derivation of the above likelihood follows from the simple fact that, conditionally given x_i, $(y_{i1}, \cdots, y_{im_i})$ is a random sample of size m_i from the conditional (normal) distribution of Y with conditional mean linear in x_i and conditional variance independent of x_i, as in Case 1.

Unrestricted maximization. The unrestricted MLEs of the parameters are obtained as

$$\hat{\mu}_x = \bar{x}, \quad \hat{\mu}_y = \bar{\bar{y}} + \frac{E}{C}(\bar{x} - \bar{\bar{x}}), \quad \hat{\sigma}_x^2 = \frac{A}{n} \tag{16}$$

$$\hat{\sigma}_y^2 = \frac{1}{M}[W_y + D + \frac{MAE^2}{nC^2} - \frac{E^2}{C}], \quad \hat{\rho} = \frac{E\hat{\sigma}_x}{C\hat{\sigma}_y}. \tag{17}$$

This results in the unrestricted maximum value of the likelihood as (apart from some constants)

$$\frac{1}{A^{\frac{n}{2}} \times [D - \frac{E^2}{C} + W_y]^{\frac{M}{2}}}. \tag{18}$$

Restricted maximization. As before, we first fix ρ and maximize the likelihood under the hypothesis of a common mean μ and a common variance σ^2. It is easy to see that the restricted MLE of the common mean μ is the same as in the previous case (see (9)), and the restricted MLE of the common variance σ^2 is obtained as $(n + M)\hat{\sigma}^2(\rho) = Q_2(\rho)$ where

$$Q_2(\rho) = A + \frac{D + C\rho^2 - 2E\rho + W_y}{1 - \rho^2} + \frac{nM\{\bar{\bar{y}} - \bar{x} + \rho(\bar{x} - \bar{\bar{x}})\}^2}{(1-\rho)\{M(1-\rho) + n(1+\rho)\}}. \tag{19}$$

Hence, the likelihood when maximized with respect to μ and σ^2 for fixed ρ simplifies to

$$L_2(\rho|data) \propto [(1 - \rho^2)^{\frac{M}{2}} \times Q_2(\rho)^{\frac{n+M}{2}}]^{-1}. \tag{20}$$

Thus in this case to maximize the likelihood under H_0 we need to minimize $U_2(\rho) = (1-\rho^2) \times (Q_2(\rho))^{1+\frac{n}{M}}$ with respect to ρ for values larger than ρ_0. Again, in any specific application, it can be easily done numerically. The LRT in this case is equivalent to rejecting H_0 for large values of T_2 where

$$T_2 = \frac{1}{A} \times \left[\frac{min_{\rho \geq \rho_0} U_2(\rho)}{D - \frac{E^2}{C} + W_y}\right]^{\frac{M}{n}}. \tag{21}$$

Details of the above computations appear in the Appendix.

Remark 2. Again, when the design is balanced, apart from $C = nA$, there will be some obvious simplifications in $Q_2(\rho)$, and in particular, the last term of $Q_2(\rho)$ changes to $\frac{nm(\bar{y}-\bar{x})^2}{(1-\rho)[m(1-\rho)+(1+\rho)]}$. As before, in view of the location and scale invariance of T_2, the critical value $c_2(\rho_0)$ for rejection of H_0 can be determined by setting $\mu = 0$, $\sigma = 1$ and $\rho = \rho_0$, and then verifying that the type I error holds at the stipulated level α for values of $\rho > \rho_0$. We omit the details here.

Table 1: Cutoff points for likelihood method, Case I.

ρ_0	n	m_i	α	cutoff
0.9	5	1	0.05	6.4759
0.9	10	1	0.05	3.1456
0.9	15	1	0.05	2.6378
0.9	5	3	0.05	6.3867
0.9	10	3	0.05	3.1185
0.9	15	3	0.05	2.6402

Table 2: Estimated Type I error rates for $\rho > \rho_0 = 0.9$, Case I.

ρ	n	m_i	α
0.92	5	1	0.0439
0.92	10	1	0.0396
0.92	15	1	0.0371
0.92	5	3	0.0452
0.92	10	3	0.0409
0.92	15	3	0.0335
0.95	5	1	0.033
0.95	10	1	0.0299
0.95	15	1	0.0274
0.95	5	3	0.0374
0.95	10	3	0.0305
0.95	15	3	0.0237
0.99	5	1	0.0299
0.99	10	1	0.0254
0.99	15	1	0.0253
0.99	5	3	0.0309
0.99	10	3	0.0277
0.99	15	3	0.0266

3. Tests based on combinations of p-values

In this section we propose several tests based on combinations of p-values for testing the overall null hypothesis H_0 using tests for the component null hypotheses $H_{01} : \mu_x = \mu_y$ versus both-sided alternatives, $H_{02} : \sigma_x = \sigma_y$ versus both-sided alternatives, and $H_{03} : \rho \geq \rho_0$ versus $H_{13} : \rho < \rho_0$.

1. Assume $m_1 = \cdots = m_n = m$. A standard test for $H_{01} : \mu_x = \mu_y$ is the paired t-test, rejecting H_{01} for large values of $|t_d| = |\frac{\sqrt{n}\bar{d}}{s_d}|$ where $d_i = x_i - \bar{y}_i$, $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$, and $s_d^2 = \frac{\sum_{i=1}^n (d_i-\bar{d})^2}{n-1}$. The p-value here is

defined as

$$p_1 = Pr[|t_{n-1}| > |t_d|]. \tag{22}$$

2. For testing H_{02}, we assume $\rho = \rho_0$, and define the modified Pitman-Morgan test based on $u_i = x_i + \bar{y}_i[\frac{m_i}{1+(m_i-1)\rho_0^2}]^{\frac{1}{2}}$ and $v_i = x_i - \bar{y}_i[\frac{m_i}{1+(m_i-1)\rho_0^2}]^{\frac{1}{2}}$. Then H_{02} is equivalent to $H_{02}^* : \rho_{uv} = 0$, for which a standard test is based on $t_{uv} = \frac{r_{uv}(n-2)^{\frac{1}{2}}}{(1-r_{uv}^2)^{\frac{1}{2}}}$. The p-value for this test is defined as

$$p_2 = Pr[|t_{n-2}| > |t_{uv}|]. \tag{23}$$

3. Assume again $m_1 = \cdots = m_n = m$. Noting that $\rho_{x\bar{y}} = [\frac{m\rho^2}{1+(m-1)\rho^2}] = \rho^*$, a test for H_{03} is readily based on the sample correlation r^* between x and $\bar{y}$. Defining $z^* = \frac{1}{2}\ln\frac{1+r^*}{1-r^*}$ and $\zeta^* = \frac{1}{2}\ln\frac{1+\rho_0^*}{1-\rho_0^*}$, where $\rho_0^* = [\frac{m\rho_0^2}{1+(m-1)\rho_0^2}]$, the p-value here is defined as

$$p_3 = Pr[N(0,1) < (z^* - \zeta^*)(n-3)^{\frac{1}{2}}]. \tag{24}$$

Following Hartung et al. [15], we now propose the following three tests for H_0 based on combinations of the above p-values. Note however that, unlike in standard problems, here the p-values are *not* independent even under H_0! This means that the usual standard null sampling distributions of the following combined tests under independence of the p-values cannot be used in our context because here the p-values are *not* independent.

1. *Tippett's test*. This test rejects H_0 when $p_{(1)} = min(p_1, p_2, p_3) < c_1$.

2. *Fisher's test*. This test rejects H_0 when $-2[\ln p_1 + \ln p_2 + \ln p_3] > c_2$.

3. *Stouffer's test*. This test rejects H_0 when $|\Phi^{-1}(p_1) + \Phi^{-1}(p_2) + \Phi^{-1}(p_3)| < c_3$.

We have evaluated the values of c_1, c_2, c_3 by simulations and also computed simulated powers for the same scenarios as in Section 2. Table 3 shows the cutoff values for each test in the scenario where $\mu_x = \mu_y = 0, \sigma_x^2 = \sigma_y^2 = 1, \rho_0 = 0.9$. Table 4 displays the estimated Type I error rates for $\rho > \rho_0 = 0.9$; notice that for most cases these values are smaller than 0.05, but Tippett's test and Fisher's test have estimated Type I error rates > 0.05 (shown in bold) in the case where $\rho = 0.99$ and $m = 3$. Tables 14-20 in the Appendix provide the power of the tests under various alternatives. As with the LRT approach, the power is a maximum when the means are different and a minimum when the correlation is smaller than the hypothesized value. In general, Fisher's test has the highest power. Stouffer's test has a higher power when the sample size is rather small ($n = 5$). The LRT approach gives comparable power to Fisher's test, and sometimes a higher power.

Table 3: Cutoff points for p-value methods

ρ	n	m_i	Tippett	Fisher	Stouffer
0.9	5	1	0.0223	10.4968	-1.8364
0.9	10	1	0.0198	10.9312	-2.1755
0.9	15	1	0.0185	11.5213	-2.3712
0.9	5	3	0.0226	10.6629	-1.8838
0.9	10	3	0.0186	11.4212	-2.2772
0.9	15	3	0.0185	11.3721	-2.2879

Table 4: Estimated Type I error rates for $\rho > \rho_0 = 0.9$

ρ	n	m_i	Tippett α	Fisher α	Stouffer α
0.92	5	1	0.0498	0.0481	0.0358
0.92	10	1	0.0468	0.0439	0.0327
0.92	15	1	0.0409	0.0343	0.0248
0.92	5	3	0.0484	0.0448	0.0349
0.92	10	3	0.0416	0.0354	0.0271
0.92	15	3	0.0412	0.0402	0.0271
0.95	5	1	0.0457	0.0402	0.0183
0.95	10	1	0.0388	0.0314	0.0092
0.95	15	1	0.039	0.025	0.0053
0.95	5	3	0.0474	0.0442	0.0172
0.95	10	3	0.0473	0.0421	0.0116
0.95	15	3	**0.0551**	0.0427	0.0088
0.99	5	1	0.0399	0.0309	0.0007
0.99	10	1	0.0386	0.0262	0
0.99	15	1	0.0388	0.023	0
0.99	5	3	**0.1112**	**0.1067**	0.0018
0.99	10	3	**0.3148**	**0.2344**	0.0001
0.99	15	3	**0.5378**	**0.4211**	0

4. Applications

In this section we first elaborate on the EPA dataset, indicated in Section

1, dealing with comparing the performances of large chamber (x) and small chamber (y) experiments to quantify formaldehyde emissions from composite wood products. Here large chamber measurements are observed once and the corresponding small chamber measurements are observed three times ($m_1 = m_2 = m_3 = m = 3$). Denoting by $d_i = x_i - \bar{y}_i$, the mean differences between the two methods, the statistical method suggested by the ARB staff of the State of California, borrowing ideas from TOST (Schuirmann [16, 17]), recommends equivalence when

$$T_{EPA} = |\bar{d}| + 0.88 S_d \leq C \tag{25}$$

for a suggested value of C (for low, mid, high ranges). The fundamental assumption behind TOST is that (X, Y) follows a lognormal distribution, and we are interested in testing if the ratio of means is close to 1, without any information as to their variances or correlation! In the log scale, this amounts to testing if the difference of two means is small. That closeness or even equality of means alone does not guarantee equivalence has been elaborated earlier.

The second EPA application deals with comparing conventional purging and low-flow sampling method with *HydraSleeve* sampling method in order to quantify concentrations of VOCs, SVOCs, metals, dissolved gases, and perchlorate in groundwater. Here equivalence is nicely assessed by applying standard regression analysis, and sign and Wilcoxon nonparametric tests (TRC Report: HydraSleeve Comparison Study, Letter report to California Department of Toxic Substances Control [18]). Under the assumption of a bivariate normal distribution of the contamination samples collected from groundwater, we can apply the methods discussed in our paper. Towards this end, we have selected two compounds: TCE (Trichloroethylene) and DCA (Dichloroacetic acid), and carried out the analysis below. It is interesting to note that in both the cases all the proposed tests agree in the final conclusion.

4.1 TCE: a case of agreement

In this case we tested the hypothesis $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho \geq 0.9$. Figure 1 shows a scatterplot of X (the conventional low-flow sampling method measurements) and Y (the average of one or two measurements from the new *HydraSleeve* sampling method). The measurements come from 23 distinct wells.

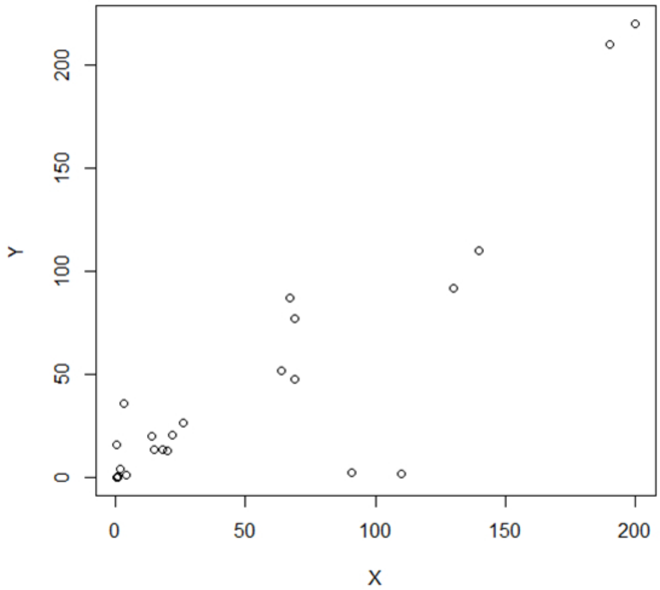

Figure 1: Scatterplot of TCE measurements

The conventional sampling method measurements have a sample mean of 54.6087 and a sample variance of 3,856.09, while the *HydraSleeve* measurements have a sample mean of 46.5435 and a sample variance of 3,888.58. The sample correlation between the measurements is 0.8619. Table 5 gives the results for the tests, leading to acceptance of the agreement hypothesis.

Table 5: Results of TCE analysis.

Test	Cutoff	Test statistic	Conclusion
LRT	2.3755	2.2061	Accept H_0
Tippett	0.01803	0.22175	Accept H_0
Fisher	11.7477	5.8498	Accept H_0
Stouffer	-2.4731	0.4399	Accept H_0

4.2 DCA: a case of non-agreement

Here we tested the hypothesis $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho \geq 0.9$. Figure 2 shows a scatterplot of X (the conventional low-flow sampling method measurements) and Y (the average of one or two measurements from the new *HydraSleeve* sampling method). The measurements come from 19 distinct wells.

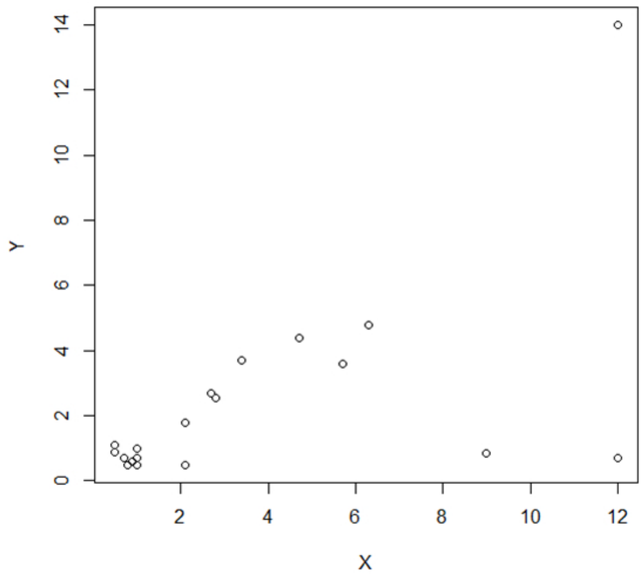

Figure 2: Scatterplot of DCA measurements

The conventional sampling method measurements has a sample mean of 3.6421 and a sample variance of 14.0103, while the *HydraSleeve* measurements has a sample mean of 2.4 and a sample variance of 9.9425. The sample correlation between the measurements is 0.5925 which is rather low. Table 6 gives the results for the tests, leading to rejection of the agreement hypothesis.

Table 6: Results of DCA analysis

Test	Cutoff	Test statistic	Conclusion
LRT	2.4621	3.6415	Reject H_0
Tippett	0.0186	0.00078	Reject H_0
Fisher	11.6593	20.7273	Reject H_0
Stouffer	-2.4187	-4.7037	Reject H_0

Acknowledgement: We thank two anonymous reviewers for their excellent comments and helpful suggestions which improved the presentation.

References

[1] Lin, L.I.K., Hedayat, A.S., Sinha, B., and Yang, M. Statistical methods in assessing agreement: models, issues, and tools. *Journal of the American Statistical Association*, **97**, 2002, 257-270.

[2] Cohen, J. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, **20**, 1960, 37-46.

[3] Cohen, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit, *Psychological Bulletin*, **70**, 1968, 213-220.

[4] Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data." *Biometrics* **33**(1): 1977, 159-174.

[5] Lin, L.I.K. A concordance correlation coefficient to evaluate reproducibility, *Biometrics*, **45**, 1989, 255-268.

[6] Chinchilli, V. M. and Esinhart, J. D. Design and analysis of intra-subject variability in cross over experiments. *Statistical Medicine.*, **15**, 1996, 1619-1634.

[7] Vonesh, E. F., Chinchilli, V.M., and Kewei Pu. Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. *Biometrics.*, **52**, 1996, 572-587.

[8] Vonesh, E. F., and Chinchilli, V. M. **Linear and nonlinear models for the analysis of repeated measurements**. (Vol. 1), 1997, CRC press.

[9] Barnhart, H. X., and Williamson, J. M. Modeling concordance correlation via GEE to evaluate reproducibility. *Biometrics.* **57**(3), 2001, 931-940.

[10] King, T. S., and Chinchilli, V. M. (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Statistics in Medicine.* 20(14):2131-2147.

[11] King, T. S., and Chinchilli, V. M. Robust estimators of the concordance correlation coefficient. *Journal of Biopharmaceutical Statistics.*, **11**(3), 2001, 83-105.

[12] Lin, L.I.K. Total deviation index for measuring individual agreement with application in laboratory performance and bioequivalence, *Statistics in Medicine*, **19**, 2000, 255-270.

[13] Yimprayoon, P., Tiensuwan, M., and Sinha, B. Some statistical aspects of assessing agreement: theory and application. In *Festschrift for Tarmo Pukkila on his 60th birthday, edited by Liski, Isotalo, Niemela, Puntaen, Styan. Department of Mathematics, Statistics, and Philosophy, University of Tampere*, 2006, 327-346.

[14] California Environmental Protection Agency Air Resources Board. Airborne Toxic Control Measure to Reduce Formaldehyde Emissions from Composite Wood Products. Final Regulation Order. April 2008. http://www.arb.ca.gov/regact/2007/compwood07.htm.

[15] Hartung, J., Knapp G. and Sinha, B. **Statistical Meta-Analysis with Applications**, John Wiley and Sons, Inc., 2008, New Jersey.

[16] Schuirmann, D.L. On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics*, **37**, 1981, 137.

[17] Schuirmann, D.L. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, **15**, 1987, 657-680.

[18] TRC Report: HydraSleeve Comparison Study, Letter report to California Department of Toxic Substances Control, November 17, 2011.

Appendix

Here we provide details of the derivations of the MLEs of the model parameters in Cases 1 and 2.

Case 1. Based on the likelihood $L(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|data)$ displayed in (4), taking its logarithm and successtive derivaties with respect to the parameters, we get the following:

1. $\frac{\partial lnL(.|.)}{\partial \mu_x} = 0 \rightarrow \frac{n(\bar{x}-\mu_x)}{\sigma_x} = \frac{M\rho[\bar{\bar{y}}-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(\bar{\bar{x}}-\mu_x)]}{\sigma_y(1-\rho^2)}$.
2. $\frac{\partial lnL(.|.)}{\partial \mu_y} = 0 \rightarrow \mu_y = \bar{\bar{y}} + \rho\frac{\sigma_y}{\sigma_x}(\mu_x - \bar{\bar{x}})$.

Combining the above two equations, we readily get

$$\hat{\mu}_x = \bar{x}, \;\; \hat{\mu}_y = \bar{\bar{y}} + \rho\frac{\sigma_y}{\sigma_x}(\bar{x} - \bar{\bar{x}}). \tag{26}$$

Using the above solutions for μ_x and μ_y, we get the following:

$$\sum_{i=1}^{n} m_i[\bar{y}_i - \hat{\mu}_y - \rho\frac{\sigma_y}{\sigma_x}(x_i - \hat{\mu}_x)]^2 = D + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x} \tag{27}$$

and

$$\sum_{i=1}^{n} m_i(x_i - \bar{x})[\bar{y}_i - \hat{\mu}_y - \rho\frac{\sigma_y}{\sigma_x}(x_i - \hat{\mu}_x)] = E - \rho\frac{\sigma_y}{\sigma_x}C. \tag{28}$$

Using all of the above facts, we observe that

1. $\frac{\partial lnL(.|.)}{\partial \sigma_x^2} = 0 \rightarrow \frac{A}{\sigma_x^2} = n + \frac{\rho(E - \frac{\rho\sigma_y}{\sigma_x}C)}{\sigma_x\sigma_y(1-\rho^2)}$.
2. $\frac{\partial lnL(.|.)}{\partial \sigma_y^2} = 0 \rightarrow n(1-\rho^2)\sigma_y^2 + E\frac{\rho\sigma_y}{\sigma_x} = D \rightarrow n\sigma_y^2 + E\frac{\rho\sigma_y}{\sigma_x} - D = n\rho^2\sigma_y^2$.
3. $\frac{\partial lnL(.|.)}{\partial \rho} = 0 \rightarrow n\rho\sigma_y^2 = C\rho\frac{\sigma_y^2}{\sigma_x^2} - E\frac{\sigma_y}{\sigma_x} + \frac{\rho}{(1-\rho^2)} \times [D + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x}]$.
4. Multiplying both sides of (3) above by ρ and using (2):
$C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - E\rho\frac{\sigma_y}{\sigma_x} + \frac{\rho^2}{1-\rho^2}\left[D + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x}\right] = n\sigma_y^2 + E\rho\frac{\sigma_y}{\sigma_x} - D$.
5. Simplifying (4): $n\sigma_y^2 = \frac{1}{1-\rho^2}\left[C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x} + D\right]$.
6. Using (2) and (5) from above:
$\frac{n\sigma_y^2(1-\rho^2)-D}{C} = \rho^2\frac{\sigma_y^2}{\sigma_x^2} - \rho\frac{2E\sigma_y}{C\sigma_x} = -\frac{E\rho\sigma_y}{C\sigma_x} \rightarrow \rho\frac{\sigma_y}{\sigma_x} = \frac{E}{C}$.

Using the last two results, we readily get

$$\sigma_y^2 = \frac{D - \frac{E^2}{C}}{n[1 - \frac{E^2\sigma_x^2}{C^2\sigma_y^2}]} \tag{29}$$

$$\longleftrightarrow n\sigma_y^2 - \frac{nE^2\sigma_x^2}{C^2} = D - \frac{E^2}{C}$$

$$\rightarrow \sigma_y^2 = \frac{1}{n}[D + n\sigma_x^2\frac{E^2}{C^2} - \frac{E^2}{C}].$$

Using the above, we get $\hat{\sigma}_x^2 = \frac{A}{n}$, and hence all the unrestricted MLEs are evaluated. Finally, since

$$\hat{\sigma}_y^2(1 - \hat{\rho}^2) = \hat{\sigma}_y^2 - \frac{E^2}{C^2}\hat{\sigma}_x^2 = \frac{1}{n}[D - \frac{E^2}{C}] \tag{30}$$

it follows from the expression of the likelihood function that the exponent of the maximum likelihood is a constant, and hence the maximum value of the unrestricted likelihood turns out to be (apart from constants)

$$sup_{unrestricted}L(.|data) \sim \frac{1}{A^{\frac{n}{2}}[D - \frac{E^2}{C} + W_y]^{\frac{n}{2}}}. \tag{31}$$

On the other hand, under H_0, the likelihood function simplifies as

$$L(\mu, \sigma|data) \sim \sigma^{-2n} \tag{32}$$

$$\times exp[-\frac{1}{2\sigma^2}\{\sum_{i=1}^{n}(x_i - \mu_x)^2 + \frac{\sum_{i=1}^{n} m_i(\bar{y}_i - \mu - \rho(x_i - \mu))^2}{1 - \rho^2}\}].$$

This yields the MLE of the common mean μ, for fixed ρ, as

$$\hat{\mu}(\rho) = \frac{n\bar{x}(1 + \rho) + M(\bar{\bar{y}} - \rho\bar{\bar{x}})}{M(1 - \rho) + n(1 + \rho)}. \tag{33}$$

Additionally, the MLE of the common variance σ^2, for fixed ρ, is given by

$$\hat{\sigma}^2 = \frac{1}{2n}[A + n(\bar{x} - \hat{\mu}(\rho))^2 + \frac{\sum_{i=1}^{n} m_i[\bar{y}_i - \hat{\mu}(\rho) - \rho(x_i - \hat{\mu}(\rho))]^2}{1 - \rho^2}]. \tag{34}$$

To simplify the above expression, note that

$$\bar{x} - \hat{\mu}(\rho) = \frac{M[\bar{x}(1 - \rho) - (\bar{\bar{y}} - \rho\bar{\bar{x}})]}{M(1 - \rho) + n(1 + \rho)}. \tag{35}$$

Writing

$$\bar{y}_i - \hat{\mu}(\rho) - \rho(x_i - \hat{\mu}(\rho)) = [(\bar{y}_i - \bar{\bar{y}}) - \rho(x_i - \bar{\bar{x}})] + [\bar{\bar{y}} - \rho\bar{\bar{x}} - \hat{\mu}(\rho)(1 - \rho)] \tag{36}$$

and noting that

$$\bar{\bar{y}} - \rho\bar{\bar{x}} - \hat{\mu}(\rho)(1-\rho) = \frac{n(1+\rho)[\bar{\bar{y}} - \bar{x} + \rho(\bar{x} - \bar{\bar{x}})]}{M(1-\rho) + n(1+\rho)} \quad (37)$$

we get

$$\sum_{i=1}^{n} m_i[\bar{y}_i - \hat{\mu}(\rho) - \rho(x_i - \hat{\mu}(\rho))]^2 = D + C\rho^2 - 2E\rho + \frac{Mn^2(1+\rho)^2[\bar{x}(1-\rho) - (\bar{\bar{y}} - \rho\bar{\bar{x}})]^2}{[M(1-\rho) + n(1+\rho)]^2}. \quad (38)$$

Since

$$\frac{nM^2[\bar{x}(1-\rho) - (\bar{\bar{y}} - \rho\bar{\bar{x}})]^2}{[M(1-\rho) + n(1+\rho)]^2} + \frac{Mn^2(1+\rho)^2[\bar{x}(1-\rho) - (\bar{\bar{y}} - \rho\bar{\bar{x}})]^2}{(1-\rho^2)[M(1-\rho) + n(1+\rho)]^2} = \frac{Mn[\bar{x}(1-\rho) - (\bar{\bar{y}} - \rho\bar{\bar{x}})]^2}{(1-\rho)[M(1-\rho) + n(1+\rho)]} \quad (39)$$

we finally get $2n\hat{\sigma}^2 = Q_1(\rho)$ as

$$\begin{aligned} Q_1(\rho) &= A + \frac{D + C\rho^2 - 2E\rho}{1-\rho^2} \\ &\quad + \frac{nM[\bar{\bar{y}} - \bar{x} + \rho(\bar{x} - \bar{\bar{x}})]^2}{(1-\rho)[M(1-\rho) + n(1+\rho)]}. \end{aligned} \quad (40)$$

Since the null likelihood function $L_1(\rho)$ for fixed ρ, maximized wrt μ and σ^2, simplifies to $L_1(\rho) \sim [(1-\rho^2)^{\frac{1}{2}} \times Q_1(\rho)]^{-n} = [U_1(\rho)]^{-n}$, it follows that the LRT statistic λ defined by

$$\lambda = \frac{\sup_{H_0} L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}|\text{ data })}{\sup_{\text{unrestricted}} L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}|\text{data})} \quad (41)$$

is equivalent to rejecting H_0 for large values of T_1 given by

$$T_1 = [min_{\rho \geq \rho_0} U_1(\rho)] \times [\frac{C}{A(DC - E^2)}]^{\frac{1}{2}}. \quad (42)$$

This is precisely what is stated in (14).

Case 2. A close inspection of the likelihood $L(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|data)$ in this case shows a striking similarity and very minor differences with the likelihood under Case 1. Obviously, the unrestricted MLEs of μ_x and μ_y are the same as in Case 1, namely,

$$\hat{\mu}_x = \bar{x},\ \hat{\mu}_y = \bar{\bar{y}} + \rho\frac{\sigma_y}{\sigma_x}(\bar{x} - \bar{\bar{x}}). \quad (43)$$

As before, we then have

$$\sum_{i=1}^{n} m_i[\bar{y}_i - \hat{\mu}_y - \rho\frac{\sigma_y}{\sigma_x}(x_i - \hat{\mu}_x)]^2 = D + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x} \quad (44)$$

and

$$\sum_{i=1}^{n} m_i(x_i - \bar{x})[\bar{y}_i - \hat{\mu}_y - \rho\frac{\sigma_y}{\sigma_x}(x_i - \hat{\mu}_x)] = E - \rho\frac{\sigma_y}{\sigma_x}C. \tag{45}$$

Using all of the above facts, we observe that

1. $\frac{\partial lnL(.|.)}{\partial\sigma_x^2} = 0 \rightarrow \frac{A}{\sigma_x^2} = n + \frac{\rho(E - \frac{\rho\sigma_y}{\sigma_x}C)}{\sigma_x\sigma_y(1-\rho^2)}$.
2. $\frac{\partial lnL(.|.)}{\partial\sigma_y^2} = 0 \rightarrow M(1-\rho^2)\sigma_y^2 + E\frac{\rho\sigma_y}{\sigma_x} = D + W_y \rightarrow M\sigma_y^2 + E\frac{\rho\sigma_y}{\sigma_x} - (D + W_y) = M\rho^2\sigma_y^2$.
3. $\frac{\partial lnL(.|.)}{\partial\rho} = 0 \rightarrow M\rho\sigma_y^2 = C\rho\frac{\sigma_y^2}{\sigma_x^2} - E\frac{\sigma_y}{\sigma_x} + \frac{\rho}{(1-\rho^2)} \times [D + W_y + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x}]$.
4. Multiplying both sides of (3) above by ρ and using (2): $C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - E\rho\frac{\sigma_y}{\sigma_x} + \frac{\rho^2}{(1-\rho^2)} \times [D + W_y + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x}] = M\sigma_y^2 + E\frac{\rho\sigma_y}{\sigma_x} - (D + W_y)$.
5. $M\sigma_y^2 = \frac{1}{1-\rho^2}[D + W_y + C\rho^2\frac{\sigma_y^2}{\sigma_x^2} - 2E\rho\frac{\sigma_y}{\sigma_x}]$.
6. Using (2) and (5) from above: $\frac{M\sigma_y^2(1-\rho^2) - D - W_y}{C} = \rho^2\frac{\sigma_y^2}{\sigma_x^2} - \frac{2E\rho\sigma_y}{C\sigma_x} = -\frac{E\rho\sigma_y}{C\sigma_x} \rightarrow \rho\frac{\sigma_y}{\sigma_x} = \frac{E}{C}$.

Using the last two results, we readily get

$$\begin{aligned} \sigma_y^2 &= \frac{D + W_y - \frac{E^2}{C}}{M[1 - \frac{E^2\sigma_x^2}{C^2\sigma_y^2}]} \\ &\longleftrightarrow M\sigma_y^2 - \frac{ME^2\sigma_x^2}{C^2} = D + W_y - \frac{E^2}{C} \\ &\rightarrow \sigma_y^2 = \frac{1}{M}[D + W_y + \frac{M\sigma_x^2E^2}{C^2} - \frac{E^2}{C}]. \end{aligned} \tag{46}$$

This yields $\hat{\sigma}_x^2 = \frac{A}{n}$, and hence all the unrestricted MLEs are evaluated. Finally, since

$$\hat{\sigma}_y^2(1 - \hat{\rho}^2) = \hat{\sigma}_y^2 - \frac{E^2}{C^2}\hat{\sigma}_x^2 = \frac{1}{M}[D + W_y - \frac{E^2}{C}] \tag{47}$$

it follows from the expression of the likelihood function that the exponent of the maximum likelihood is a constant, and hence the maximum value of the unrestricted likelihood turns out to be (apart from constants)

$$sup_{unrestricted}L(.|data) \sim \frac{1}{A^{\frac{n}{2}}[D - \frac{E^2}{C} + W_y]^{\frac{M}{2}}}. \tag{48}$$

On the other hand, under H_0 and for fixed ρ, the MLE of the common mean μ is the same as in Case 1, namely,

$$\hat{\mu}(\rho) = \frac{n\bar{x}(1+\rho) + M(\bar{\bar{y}} - \rho\bar{\bar{x}})}{M(1-\rho) + n(1+\rho)} \tag{49}$$

and, following the computations under H_0 in Case 1, the MLE of the common variance σ^2 is given by $\hat{\sigma}^2(\rho) = \frac{Q_2(\rho)}{n+M}$, where

$$Q_2(\rho) = A + \frac{D + C\rho^2 - 2E\rho + W_y}{1-\rho^2} + \frac{nM[\bar{\bar{y}} - \bar{x} + \rho(\bar{x} - \bar{\bar{x}})]^2}{(1-\rho)[M(1-\rho) + n(1+\rho)]}. \tag{50}$$

Combining the above results and arguing as in Case 1 results in the expression of the LRT given in (21).

Table 7: $\rho = 0.5 < \rho_0 = 0.9$

n	m_i	power
5	1	0.2458
10	1	0.642
15	1	0.8527
5	3	0.42657
10	3	0.8875
15	3	0.9723

Table 8: $\mu_y = 1 \neq \mu_x = 0$

n	m_i	power
5	1	0.8815
10	1	0.9999
15	1	1
5	3	0.9996
10	3	1
15	3	1

Table 9: $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	power
5	1	0.5481
10	1	0.961
15	1	0.9984
5	3	0.9096
10	3	0.9996
15	3	1

Table 10: $\mu_y = 1 \neq \mu_x = 0$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	power
5	1	0.8197
10	1	0.9976
15	1	1
5	3	0.9885
10	3	1
15	3	1

Table 11: $\rho = 0.5 < \rho_0 = 0.9$ and $\mu_y = 1 \neq \mu_x = 0$

n	m_i	power
5	1	0.6795
10	1	0.9836
15	1	0.9988
5	3	0.9515
10	3	1
15	3	1

Table 12: $\rho = 0.5 < \rho_0 = 0.9$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	power
5	1	0.5043
10	1	0.9442
15	1	0.9955
5	3	0.5077
10	3	0.9486
15	3	0.9888

Table 13: $\rho = 0.5 < \rho_0 = 0.9$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$ and $\mu_y = 1 \neq \mu_x = 0$

n	m_i	power
5	1	0.6653
10	1	0.9862
15	1	0.9995
5	3	0.8536
10	3	0.9978
15	3	0.9998

Table 14: $\rho = 0.5 < \rho_0 = 0.9$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.2151	0.2762	0.3224
10	1	0.6453	0.6981	0.5593
15	1	0.8661	0.8714	0.6836
5	3	0.2984	0.3835	0.4372
10	3	0.8323	0.8956	0.7832
15	3	0.9764	0.9898	0.9391

Table 15: $\mu_y = 1 \neq \mu_x = 0$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.8507	0.8843	0.6941
10	1	0.9998	0.9998	0.9243
15	1	1	1	0.9796
5	3	0.9981	0.9984	0.8461
10	3	1	1	0.9781
15	3	1	1	0.9987

Table 16: $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.403	0.4249	0.3596
10	1	0.9154	0.9615	0.754
15	1	0.994	0.9984	0.9189
5	3	0.6942	0.7543	0.5457
10	3	0.9971	0.9994	0.916
15	3	1	1	0.9925

Table 17: $\mu_y = 1 \neq \mu_x = 0$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.5252	0.7668	0.7903
10	1	0.9759	0.9979	0.9904
15	1	0.9991	0.9999	0.9993
5	3	0.823	0.9734	0.9505
10	3	1	1	0.9994
15	3	1	1	1

Table 18: $\rho = 0.5 < \rho_0 = 0.9$ and $\mu_y = 1 \neq \mu_x = 0$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.4099	0.6822	0.7232
10	1	0.9163	0.9835	0.9622
15	1	0.9957	0.9997	0.9963
5	3	0.6486	0.9415	0.9381
10	3	0.995	0.9999	0.9993
15	3	1	1	1

Table 19: $\rho = 0.5 < \rho_0 = 0.9$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.3051	0.5042	0.5782
10	1	0.8448	0.9602	0.9209
15	1	0.9982	0.9969	0.9846
5	3	0.3223	0.458	0.5203
10	3	0.8789	0.9489	0.8779
15	3	0.9886	0.9962	0.9783

Table 20: $\rho = 0.5 < \rho_0 = 0.9$ and $\sigma_y^2 = 4 \neq \sigma_x^2 = 1$ and $\mu_y = 1 \neq \mu_x = 0$

n	m_i	Tippett power	Fisher power	Stouffer power
5	1	0.3575	0.6788	0.7638
10	1	0.8987	0.9887	0.979
15	1	0.9929	0.9995	0.9984
5	3	0.5109	0.852	0.8831
10	3	0.9769	0.9987	0.9964
15	3	0.9999	1	0.9998