0.1 Measurement Error Models

DunnSEME proposes a measurement error model for use in method comparison studies. Consider n pairs of measurements X_i and Y_i for $i = 1, 2, ...n$.

$$X_i = \tau_i + \delta_i \tag{1}$$

$$Y_i = \alpha + \beta\tau_i + \epsilon_i$$

In the above formulation is in the form of a linear structural relationship, with τ_i and $\beta\tau_i$ as the true values , and δ_i and ϵ_i as the corresponding measurement errors. In the case where the units of measurement are the same, then $\beta = 1$.

$$\begin{aligned} E(X_i) &= \tau_i \\ E(Y_i) &= \alpha + \beta\tau_i \\ E(\delta_i) &= E(\epsilon_i) = 0 \end{aligned} \tag{2}$$

The value α is the inter-method bias between the two methods.

$$z_0 = d = 0 \tag{3}$$

$$z_{n+1} = z_n^2 + c \tag{4}$$

0.2 Model Formulation and Formal Testing

? formulates a model for un-replicated observations for a method comparison study as a mixed model.

$$\begin{aligned} Y_{ij} = \quad & \mu_j + S_i + \epsilon_{ij} \quad i = 1, 2...n \quad j = 1, 2 \\ & S \sim N(0, \sigma_s^2) \qquad \epsilon_{ij} \sim N(0, \sigma_j^2) \end{aligned} \tag{5}$$

As with all mixed models, the variance of each observation is the sum of all the associated variance components.

$$\begin{aligned} var(Y_{ij}) = \quad & \sigma_s^2 + \sigma_j^2 \\ cov(Y_{i1}, Y_{i2}) = \quad & \sigma_s^2 \end{aligned} \tag{6}$$

? offers maximum likelihood estimators, commonly known as Grubbs estimators, for the various variance components:

$$\begin{aligned} \hat{\sigma_s^2} &= \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} = Sxy \\ \hat{\sigma_1^2} &= \sum \frac{(x_i - \bar{x})^2}{n-1} = S^2x - Sxy \\ \hat{\sigma_2^2} &= \sum \frac{(y_i - \bar{y})^2}{n-1} = S^2y - Sxy \end{aligned} \tag{7}$$

The standard error of these variance estimates are:

$$\begin{aligned} var(\sigma_1^2) = & \quad \frac{2\sigma_1^4}{n-1} + \quad \frac{\sigma_S^2\sigma_1^2 + \sigma_S^2\sigma_2^2 + \sigma_1^2\sigma_2^2}{n-1} \\ var(\sigma_2^2) = & \quad \frac{2\sigma_2^4}{n-1} + \quad \frac{\sigma_S^2\sigma_1^2 + \sigma_S^2\sigma_2^2 + \sigma_1^2\sigma_2^2}{n-1} \end{aligned} \tag{8}$$

?presents confidence intervals for the relative precisions of the measurement methods, $\Delta_j = \sigma_S^2/\sigma_j^2$ (where $j = 1, 2$), as well as the variances σ_S^2, σ_1^2 and σ_2^2.

$$\Delta_1 > \quad \frac{C_{xy} - t(|A|/n-2))^{\frac{1}{2}}}{C_x - C_{xy} + t(|A|/n-2))^{\frac{1}{2}}} \tag{9}$$

where

$$\begin{aligned} C_x = & \quad (n-1)S_x^2 \\ C_{xy} = & \quad (n-1)S_{xy} \\ C_y = & \quad (n-1)S_y^2 \\ A = & \quad C_x \times C_y - (C_{xy})^2 \end{aligned}$$

t is the $100(1-\alpha/2)\%$ quantile of Student's t distribution with $n-2$ degrees of freedom. Δ_2 can be found by changing C_y for C_x. A lower confidence limit can be found by calculating the square root. This inequality may also be used for hypothesis testing.

For the interval estimates for the variance components, **?** presents three relations that hold simultaneously with probability $1 - 2\alpha$ where $2\alpha = 0.01$ or 0.05.

$$\begin{aligned} |\sigma^2 - C_{xy}K| &\leqslant M(C_xC_y)^{\frac{1}{2}} \\ |\sigma_1^2 - (C_x - C_{xy})K| &\leqslant M(C_x(C_x + C_y - 2C_{xy}))^{\frac{1}{2}} \\ |\sigma_2^2 - (C_y - C_{xy})K| &\leqslant M(C_y(C_x + C_y - 2C_{xy}))^{\frac{1}{2}} \end{aligned} \tag{10}$$

The case-wise differences and means are $D_i = Y_{i1} - Y_{i2}$ and $A_i = (Y_{i1} + Y_{i2})/2$ respectively. Both D_i and A_i follow a bivariate normal distribution with $E(D_i) = \mu_D = \mu_1 - \mu_2$ and $E(A_i) = \mu_A = (\mu_1 + \mu_2)/2$. The variance matrix Σ is

$$\Sigma = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2) \\ \frac{1}{2}(\sigma_1^2 - \sigma_2^2) & \sigma_S^2 + \frac{1}{4}(\sigma_1^2 + \sigma_2^2) \end{bmatrix} \tag{11}$$

? demonstrates how the Grubbs estimators for the error variances can be calculated using the difference values, providing a worked example on a data set.

$$\begin{aligned} \hat{\sigma_1^2} \quad &= \sum (y_{i1} - \bar{y1})(D_i - \bar{D}) \\ \hat{\sigma_2^2} \quad &= \sum (y_{i2} - \bar{y_2})(D_i - \bar{D}) \end{aligned} \tag{12}$$

0.2.1 Morgan Pitman

The test of the hypothesis that the variance of both methods are equal is based on the correlation value $\rho_{D,A}$ which is evaluated as follows;

$$\rho(D,A) = \frac{\sigma_1^2 - \sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)(4\sigma_S^2 + \sigma_1^2 + \sigma_2^2)}} \tag{13}$$

The correlation constant takes the value zero if, and only if, the two variances are equal. Therefore a test of the hypothesis $H : \sigma_1^2 = \sigma_2^2$ is equivalent to a test of the hypothesis $H : \rho(D, A) = 0$. The corresponds to the well-known t test for a correlation coefficient with $n - 2$ degrees of freedom.

? describes the Morgan-Pitman test as identical to the test of the slope equal to zero in the regression of Y_{i1} on Y_{12}, adding that this result can be shown using straightforward algebra.

0.2.2 Bartko's Bradley-Blackwood Test

This is a regression based approach that performs a simulataneous test for the equivalence of means and variances of the respective methods.

$$D = (X_1 - X_2) \tag{14}$$

$$M = (X_1 + X_2)/2 \tag{15}$$

The Bradley Blackwood Procedure fits D on M as follows:

$$D = \beta_0 + \beta_1 M \tag{16}$$

Both beta values, the intercept and slope, are derived from the respective means and standard deviations of their respective data sets.
We determine if the respective means and variances are equal if both beta values are simultaneously equal to zero. The Test is conducted using an F test, calculated from the results of a regression of D on M.
We have identified this approach to be examined to see if it can be used as a foundation for a test perform a test on means and variances individually.
Russell et al have suggested this method be used in conjunction with a paired t-test , with estimates of slope and intercept.

subsectiont-test

0.2.3 Blackwood Bradley Model

This is a regression based approach that performs a simultaneous test for the equivalence of means and variances of the respective methods.
We have identified this approach to be examined to see if it can be used as a foundation for a test perform a test on means and variances individually.

$$D = (X_1 - X_2) \tag{17}$$

$$M = (X_1 + X_2)/2 \tag{18}$$

The Bradley Blackwood Procedure fits D on M as follows:

$$D = \beta_0 + \beta_1 M \tag{19}$$

Both beta values, the intercept and slope, are derived from the respective means and standard deviations of their respective data sets.
We determine if the respective means and variances are equal if both beta values are simultaneously equal to zero. The Test is conducted using an F test, calculated from the results of a regression of D on M.
Russell et al have suggested this method be used in conjunction with a paired t-test , with estimates of slope and intercept. Bradley and Blackwood have developed a regression based approach assessing the agreement.
The Bradley Blackwood test is a simultaneous test for bias and precision. They propose a regression approach which fits D on M, where D is the difference and average of a pair of results.

0.2.4 Pitman & Morgan Test

This test assess the equality of population variances. Pitman's test tests for zero correlation between the sums and products.
Correlation between differences and means is a test statistics for the null hypothesis of equal variances given bivariate normality.

0.3 Thompson 1963

? defines Δ_j to be a measure of the relative precision of the measurement methods, with $\Delta_j = \sigma^2_S/\sigma^2_j$(where $j = 1, 2$). Confidence intervals for Δ_j are also presented.

$$\Delta_1 > \frac{C_{xy} - t(\frac{|A|}{n-1}))^{\frac{1}{2}}}{C_x - C_{xy} + t(\frac{|A|}{n-1}))^{\frac{1}{2}}}, \tag{20}$$

where

$$\begin{aligned} C_x &= (n-1)S^2_x, \\ C_{xy} &= (n-1)S_{xy}, \\ C_y &= (n-1)S^2_y, \\ A &= C_x \times C_y - (C_{xy})^2. \end{aligned}$$

The value t is the $100(1 - \alpha/2)\%$ quantile of Student's t distribution with $n - 2$ degrees of freedom. The ratio Δ_2 can be found by interchanging C_y and

C_x. A lower confidence limit can be found by calculating the square root. The inequality in equation 1.10 may also be used for hypothesis testing.

For the interval estimates for the variance components, **?** presents three relations that hold simultaneously with probability $1 - 2\alpha$ where $2\alpha = 0.01$ or 0.05.

$$
\begin{aligned}
|\sigma^2 - C_{xy}K| &\leqslant M(C_x C_y)^{\frac{1}{2}} \\
|\sigma_1^2 - (C_x - C_{xy})K| &\leqslant M(C_x(C_x + C_y - 2C_{xy}))^{\frac{1}{2}} \\
|\sigma_2^2 - (C_y - C_{xy})K| &\leqslant M(C_y(C_x + C_y - 2C_{xy}))^{\frac{1}{2}}
\end{aligned}
$$

? contains tables for K and M.

? offers a formal simultaneous hypothesis test for the mean and variance of two paired data sets. Using simple linear regression of the differences of each pair against the sums, a line is fitted to the model, with estimates for intercept and slope ($\hat{\beta}_0$ and $\hat{\beta}_1$). The null hypothesis of this test is that the mean (μ) and variance (σ^2) of both data sets are equal if the slope and intercept estimates are equal to zero(i.e $\sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ if and only if $\beta_0 = \beta_1 = 0$)

A test statistic is then calculated from the regression analysis of variance values (**?**) and is distributed as 'F' random variable. The degrees of freedom thereof are $\nu_1 = 2$ and $\nu_1 = n - 2$ (where n is the number of pairs). The critical value is chosen for $\alpha\%$ significance with those same degrees of freedom. **?** amends this methodology for use in method comparison studies, using the averages of the pairs, as opposed to the sums, and their differences. This approach can facilitate simultaneous usage of test with the Bland-Altman methodology. Bartko's test statistic take the form:

$$F.test = \frac{(\Sigma d^2) - SSReg}{2MSReg} \tag{21}$$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Averages	1	0.04	0.04	0.74	0.4097
Residuals	10	0.60	0.06		

Table 1: Regression ANOVA of case-wise differences and averages for Grubbs Data

For the Grubbs data, $\Sigma d^2 = 5.09$, $SSReg = 0.60$ and $MSreg = 0.06$ Therefore the test statistic is 37.42, with a critical value of 4.10. Hence the means and variance of the Fotobalk and Counter chronometers are assumed to be simultaneously equal.

Importantly, this methodology determines whether there is both inter-method bias and precision present, or alternatively if there is neither present. It has previously been demonstrated that there is a inter-method bias present, but as this procedure does not allow for separate testing, no conclusion can be drawn on the comparative precision of both methods.

0.3.1 Formal Testing

The Bland Altman plot is a simple tool for inspection of the data, but in itself it offers no formal testing procedure in this regard. To this end, the approach proposed by **?** is a formal test on the Pearson correlation coefficient of casewise differences and means (ρ_{AD}). According to the authors, this test is equivalent to a well established tests for equality of variances, known as the 'Pitman Morgan Test' (**??**).

For the Grubbs data, the correlation coefficient estimate (r_{AD}) is 0.2625, with a 95% confidence interval of (-0.366, 0.726) estimated by Fishers 'r to z' transformation (**?**). The null hypothesis (ρ_{AD} =0) would fail to be rejected. Consequently the null hypothesis of equal variances of each method would also fail to be rejected.

There has no been no further mention of this particular test in the subsequent article published by Bland and Altman, although **?** refers to Spearmans' rank correlation coefficient.

0.4 Bartko's Regression and Ellipse

? offers a formal simultaneous hypothesis test for the mean and variance of two paired data sets. Using simple linear regression of the differences of each pair against the sums, a line is fitted to the model, with estimates for intercept and slope (β_0 and β_1). The null hypothesis of this test is that the mean (μ) and variance (σ^2) of both data sets are equal if the slope and intercept estimates are equal to zero(i.e $\sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ if and only if $\beta_0 = \beta_1 = 0$)

A test statistic is then calculated from the regression analysis of variance values (**?**) and is distributed as 'F' random variable. The degrees of freedom thereof are $\nu_1 = 2$ and $\nu_1 = n - 2$ (where n is the number of pairs). The critical value is chosen for $\alpha\%$ significance with those same degrees of freedom. **?** amends this metholodogy for calculation using the from the averages of the pairs, as opposed to the sums, and their differences. This would facilitate simultaneous usage of test with the Bland Altman methodology. Bartko's test statistic take the form:

$$F.test = \frac{(\Sigma D^2) - SSReg}{2MSReg} \tag{22}$$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Averages	1	0.04	0.04	0.74	0.4097
Residuals	10	0.60	0.06		

Table 2: Regression ANOVA of case-wise differences and averages for Grubbs Data

For the Grubbs data, $\Sigma D^2 = 5.09$, $SSReg = 0.60$ and $MSreg = 0.06$ Therefore the test statistic is 37.42, with a critical value of 4.102821 (calculate using r code $qf(0.95, 2, 10)$). Hence the means and variance of the Fotobalk and Counter chronometers are assumed to be simultaneously equal.

Importantly, this methodology determines whether there is both inter-method bias and precision present, or alternatively if there is neither present. It has previously been demonstrated that there is a inter-method bias present, but as this procedure does not allow for seperate testing, no conclusion can be drawn on the comparative precision of both methods.

0.5 Formal Models and Tests

The Bland-Altman plot is a simple tool for inspection of data, and **?** comments on the lack of formal testing offered by that methodology. **?** formulates a model for single measurement observations for a method comparison study as a linear mixed effects model, i.e. model that additively combine fixed effects and random effects.

$$Y_{ij} = \quad \mu + \beta_j + u_i + \epsilon_{ij} \qquad i = 1, \dots, n \qquad j = 1, 2$$

The true value of the measurement is represented by μ while the fixed effect due to method j is β_j. For simplicity these terms can be combined into single terms; $\mu_1 = \mu + \beta_1$ and $\mu_2 = \mu + \beta_2$. The inter-method bias is the difference of the two fixed effect terms, $\beta_1 - \beta_2$. Each of the i individuals are assumed to give rise to random error, represented by u_i. This random effects terms is assumed to have mean zero and be normally distributed with variance σ^2. There is assumed to be an attendant error for each measurement on each individual, denoted ϵ_{ij}. This is also assumed to have mean zero. The variance of measurement error for both methods are not assumed to be identical for both methods variance, hence it is denoted σ^2_j. The set of observations (x_i, y_i) by methods X and Y are assumed to follow the bivariate normal distribution with expected values $E(x_i) = \mu_i$ and $E(x_i) = \mu_i$ respectively. The variance covariance of the observations $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \left[\begin{array}{cc} \sigma^2 + \sigma^2_1 & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma^2_2 \end{array} \right]$$

The inter-method bias is the difference of the two fixed effect terms, $\beta_1 - \beta_2$.

? demonstrates the estimation of the variance terms and relative precisions relevant to a method comparison study, with attendant confidence intervals for both. The measurement model introduced by **??** provides a formal procedure for estimate the variances σ^2,σ^2_1 and σ^2_2 devices. **?** offers estimates, commonly known as Grubbs estimators, for the various variance components. These estimates are maximum likelihood estimates, a statistical concept that shall be revisited in due course.

$$\hat{\sigma^2} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} = Sxy$$
$$\hat{\sigma^2_1} = \sum \frac{(x_i - \bar{x})^2}{n-1} = S^2x - Sxy$$
$$\hat{\sigma^2_2} = \sum \frac{(y_i - \bar{y})^2}{n-1} = S^2y - Sxy$$

? defines Δ_j to be a measure of the relative precision of the measurement methods, with $\Delta_j = \sigma^2/\sigma^2_j$. Thompson also demonstrates how to make statistical inferences about Δ_j. Based on the following identities,

$$\begin{aligned} C_x &= (n-1)S_x^2, \\ C_{xy} &= (n-1)S_{xy}, \\ C_y &= (n-1)S_y^2, \\ |A| &= C_x \times C_y - (C_{xy})^2, \end{aligned}$$

the confidence interval limits of Δ_1 are

$$\Delta_1 > \frac{C_{xy} - t(\frac{|A|}{n-2}))^{\frac{1}{2}}}{C_x - C_{xy} + t(\frac{|A|}{n-2}))^{\frac{1}{2}}} \qquad (23)$$

$$\Delta_1 > \frac{C_{xy} + t(\frac{|A|}{n-2}))^{\frac{1}{2}}}{C_x - C_{xy} - t(\frac{|A|}{n-1}))^{\frac{1}{2}}}$$

The value t is the $100(1-\alpha/2)\%$ upper quantile of Student's t distribution with $n-2$ degrees of freedom (**?**). The confidence limits for Δ_2 are found by substituting C_y for C_x in (1.3). Negative lower limits are replaced by the value 0.

The case-wise differences and means are calculated as $d_i = x_i - y_i$ and $a_i = (x_i + y_i)/2$ respectively. Both d_i and a_i are assumed to follow a bivariate normal distribution with $E(d_i) = \mu_d = \mu_1 - \mu_2$ and $E(a_i) = \mu_a = (\mu_1 + \mu_2)/2$. The variance matrix $\Sigma_{(a,d)}$ is

$$\Sigma_{(a,d)} = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2) \\ \frac{1}{2}(\sigma_1^2 - \sigma_2^2) & \sigma^2 + \frac{1}{4}(\sigma_1^2 + \sigma_2^2) \end{bmatrix}. \qquad (24)$$

0.5.1 Morgan-Pitman Testing

An early contribution to formal testing in method comparison was made by both **?** and **?**, in separate contributions. The basis of this approach is that if the distribution of the original measurements is bivariate normal. Morgan and Pitman noted that the correlation coefficient depends upon the difference $\sigma_1^2 - \sigma_2^2$, being zero if and only if $\sigma_1^2 = \sigma_2^2$.

The classical Pitman-Morgan test is a hypothesis test for equality of the variance of two data sets; $\sigma_1^2 = \sigma_2^2$, based on the correlation value $\rho_{a,d}$,and is evaluated as follows;

$$\rho(a,d) = \frac{\sigma_1^2 - \sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)(4\sigma_S^2 + \sigma_1^2 + \sigma_2^2)}} \qquad (25)$$

The correlation constant takes the value zero if, and only if, the two variances are equal. Therefore a test of the hypothesis $H : \sigma_1^2 = \sigma_2^2$ is equivalent to a test of the hypothesis $H : \rho(D,A) = 0$. The corresponds to the well-known

t test for a correlation coefficient with $n-2$ degrees of freedom. **?** describes the Morgan-Pitman test as identical to the test of the slope equal to zero in the regression of Y_{i1} on Y_{12}, a result that can be derived using straightforward algebra.