Constant sex difference across populations in liability of nonmetric traits

Akira Tagaya1*

1Nagano College of Nursing, 1694 Akaho, Komagane, Nagano 399-4117, Japan

Received 9 April 2019; accepted 6 April 2020

Abstract  Studies have revealed the existence of statistically significant sex differences in the frequency of nonmetric traits, but no agreement seems to exist about their variability among populations. This problem was examined using the multifactorial threshold model. Considering the assumption of additive effects of factors on the liability and the nature of effect of sex difference on the development of nonmetric traits, it would be reasonable to assume that the sex difference in the mean of liability is constant across populations. This hypothesis was tested and the magnitude of sex difference was examined using the world-wide dataset collected by Ossenberg and the dual-liability threshold model formulated by the author with a modification to accommodate side difference in the probability of trait occurrence. The data were divided into 16 samples regarded as randomly sampled from regional populations. The data of 31 bilateral traits were analyzed using maximum likelihood estimation procedures. After confirming the homogeneity of the variance of liability between sexes and across populations, the homogeneity and significance of sex difference in the mean of liability were tested. The results indicate the homogeneity of sex difference across populations. The assumed constant sex difference was statistically significant in 17 traits at the 1% level, and its magnitude exceeded half the averaged distance between eight groups of populations in 12 traits. Population comparisons without distinguishing sex are justifiable if they use the traits with enough weak sex difference in comparison with population differences. Since the sex difference has proved to be basically constant across populations, the estimates of the assumed constant sex difference reported in this study would provide references for selecting traits appropriate for each comparison. The Breslow–Day test of homogeneity of sex difference indicated the inapplicability of the genotype model to the data, supporting Ossenberg’s proposal for the use of side counts.

Key words: bilateral nonmetric traits, homogeneity of sex difference, dual-liability threshold model, inapplicability of genotype model, Ossenberg’s proposal for side counts

Introduction

Although Berry and Berry (1967) proposed that no substantial difference exists between sexes in the frequency (rate of positive expression) of skeletal nonmetric traits, statistically significant differences have been reported (Corruccini, 1974; Berry, 1975; Perizonius, 1979; Mouri, 1976, 1988). There seems to have been no agreement, however, about the variability of sex difference across populations. Berry (1975) analyzed the sex difference of cranial nonmetric traits in four populations, and concluded that there was no stable tendency in the direction of sex difference. However, this is rather strange under the multifactorial threshold model, which is recognized to be the most appropriate mathematical model for explaining the expression of a nonmetric trait. This model consists of a latent variable, the ‘liability,’ and a fixed value, the ‘threshold,’ and a trait is positively expressed when its liability exceeds its threshold. The liability is assumed to additively reflect the effects of numerous factors, hence is normally distributed (Falconer, 1960). It would be natural to consider that sex should also additively affect the liability, which means that the sex difference in the liability should be constant across populations.

Mouri (1988) used Cochran’s test for multiple contingency tables to analyze 30 cranial nonmetric traits and detected statistically significant male-dominant occurrence in seven traits, and female-dominant occurrence in another seven traits; these results were consistent across 8–13 populations. This means that the direction of the difference was constant across populations, at least for these traits, although the geographic background of the study populations was limited to the Japanese archipelago and adjacent areas, and no judgment could be made concerning whether the sex difference in these traits was of constant magnitude.

The dual-liability threshold model (DLM) formulated by the present author (Tagaya, 2019) would be useful for exam-
ining this problem as it enables us to test biologically meaningful hypotheses about sex difference and its relationship with populations. The DLM is a combination of two standard multifactorial threshold models, each explaining the expression of a bilateral trait on either side. Each of the two liabilities is a sum of mutually independent two components: (1) the inter-individual component shared by sides explaining the inter-side correlation in the trait expression and (2) the inter-side component whose population variance assumed to be identical between sides and constant across populations is used as the unit of liability. This unit also provides a reference for the evaluation of the magnitude of sex difference. Since the inter-side component is assumed to be normally distributed due to its multifactorial nature, the magnitude of the sex difference must be much smaller than 1 for sex to be regarded as one of the numerous additive factors of liability in the multifactorial threshold model. A hypothesis about the sex difference is expressed as a constraint on the parameters of the DLM, and tested for its goodness of fit to the actual data using the maximum likelihood estimation (MLE) procedure. This method (the DLM–MLE method) will be used in this study to test hypotheses about the homogeneity of sex difference in the mean of liability across populations.

As Mouri (1988) showed, it is also possible to test the significance of sex difference across populations using a non-parametric method, the Cochran–Mantel–Haenszel (CMH) method, developed by Cochran (1954), Mantel and Haenszel (1959), and Breslow and Day (1980). The CMH method defines the sex difference substantially in accordance with the genotype model, and tests the hypotheses in similar fashion to the DLM–MLE method. Therefore, a comparison of the results between the two methods should provide useful information about the nature of the sex difference of nonmetric traits and the appropriateness of the DLM and genotype model for explaining the expression of nonmetric traits.

This study tests the hypothesis of constant sex difference in the mean of liability across populations and, under this hypothesis, evaluates the magnitude and statistical significance of the assumed constant sex difference using the DLM–MLE method and the database published by Ossenberg (2013a, b). The data were compared with the corresponding results obtained with the CMH method.

### Materials and Methods

#### Preparation of data

The data of the bilateral nonmetric traits included in Table 1 were included in the analyses. These data were taken from the database published by Ossenberg (2013a, b). The data were divided into 16 groups, each of which was regarded to be randomly sampled from a population (listed in Table 2). An effort was made to limit and level the within-group heterogeneity, taking into consideration their geographical and historical backgrounds and the reports on nonmetric and craniometric population affinities and regional variabilities (Dodo and Ishida, 1987; Ossenberg, 1994; Shigematsu et al., 2004; Ossenberg et al., 2006; Hanihara, 2008; Hubbe et al., 2009).

| Code | Definition |
|------|------------|
| OMB  | Occipitomastoid ossicle |
| AST  | Asterionic ossicle |
| PNB  | Parietal notch bone |
| POS  | Posterior condylar canal absent |
| HYP  | Hypoglossal canal bridged or double |
| PCP  | Paracodylar process |
| ICC  | Intermediate condylar canal |
| SQS  | Parietal process of temporal squama |
| SPS  | Squamoparietal synostosis |
| MAR  | Marginal foramen of tympanic plate |
| TYM  | Tympanic dehiscence |
| FSP  | Dehiscent wall of foramen spinosum or foramen ovale |
| LPF  | Foramen in lateral pterygoid plate |
| CIV  | Pterygospinous bridge complete (foramen of Civinini) |
| PTB  | Pterygobasal spur or bridge |
| CLN  | Clinoid bridging |
| SOF  | Supraorbital foramen |
| FRG  | Frontal groove(s) |
| TRS  | Trochlear spur |
| OPT  | Accessory optic canal |
| ORB  | Orbital suture variant |
| CON  | Infraorbital suture variant |
| JAP  | Transversozygomatic suture |
| M3U  | Upper third molar congenitally absent |
| MEN  | Accessory mental foramen |
| MHB  | Mylohyoid bridge |
| BUC  | Retromolar foramen |
| M3L  | Lower third molar congenitally absent |
| TRM  | Three-rooted mandibular first molar |
| ATA  | Atlas bridging, condylar process to posterior arch |
| ATB  | Atlas bridging, condylar to transverse process |

Any positive expression was regarded as a trait occurrence. The counts were recorded for four types of trait expression: (1) on neither side, (2) on the left side only, (3) on the right side only, and (4) on both sides. The set of any values (counts, frequencies, probabilities, etc.) corresponding to these four types of trait expression in this order will be called a ‘quartet,’ and the quartet of values {a, b, c, d} will be denoted as {a, b, c, d} henceforth. To ensure the determinability of the DLM parameters (by avoiding zero frequency of trait occurrence), {a + 1/8, b + 1/4, c + 1/4, d + 1/8} was used instead of the count quartet {a, b, c, d} for analyses. This is to assume that there was a chance of observing an additional case at the probabilities {1/8, 1/4, 1/4, 1/8} for respective types although it did not happen. As a result, the side-count trait frequency is calculated as (b + d + 3/8)/(a + b + c + d + 3/4) for the left side and (c + d + 3/8)/(a + b + c + d + 3/4) for the right side, which happens to be equivalent to Anscombe’s (1948) modification in the angular transformation. Samples with a size for a trait of less than 9 were excluded from the analyses of that trait. The side-count trait frequencies are given in Table 3.

### DLM accommodating side difference

The DLM proposed by the present author (Tagaya, 2019).
is used here with a modification to accommodate the side difference because significant side differences are observed in the data used for analyses. This modification introduces a parameter $\delta$ for left-side dominance (i.e., right-side dominance if $\delta$ is negative) in the probability of trait expression, and assumes that the inter-side component of liability is distributed as $N(\delta, 1)$ for the left side and $N(-\delta, 1)$ for the right side, where $N(m, v)$ denotes a normal distribution with mean $m$ and variance $v$. DLM$(V, M, \delta)$ will hereafter denote a DLM with its inter-individual component of liability (IICL) distributed as $N(M, V)$ and the left-side-dominance parameter being $\delta$. All the components of liability are mutually independent, as in the original version of DLM. Figure 1 shows the distribution of the four types of trait expression for given values of IICL under DLM$(2, -1, 0.1)$ calculated using the function DLM_Q_Dist in Appendix 2.

Existence of complete-fit DLM for observed data

The existence of a complete-fit DLM, DLM$(V, M, \delta)$, depends on the estimate of inter-side correlation coefficient of liability, $R$, obtained by the tetrachoric procedure, as follows. When $V$ is given, the values of $M$ and $\delta$ are determined from the right- and left-side rates of trait occurrence. Noting that $R = V/(V + 1)$, $V$ is calculated as $V = R/(1 - R)$. Therefore, a complete fit DLM exists if $0 \leq R < 1$, which is equivalent to $0 \leq \phi < 1$ for the phi correlation coefficient, $\phi$. The data with enough large sample size should satisfy this condition if the model is appropriate.

Hypotheses about sex difference as constraints on DLM parameters

The following hypotheses were tested under DLM$(V_{ij}, M_{ij}, \delta_{ij})$, where $ij$ indicates sex $j$ ($j = 1$ for male and $j = 2$ for female) of population $i$ ($i = 1$ to 16).

H$_1$: $V_{ij} = V_i$ (equality of variance of IICL between sexes in every population)

H$_2$: $V_i = V$ (homogeneity of variance of IICL across populations under H$_1$)

The variance $V_i$ is expected to reflect the variability of factors affecting IICL. The magnitude of such variability is
assumed to be equal between sexes under $H_1$, and to be constant across populations under $H_3$. These seem to be biologically natural assumptions because most of the numerous factors affecting IICL should be common to sexes and populations. Statistically, these are the premise for the inter-sex factor affecting $DLM$ to be common to sexes and populations because most of the numerous factors assumed to be equal between sexes under $H_1$.

The magnitude of sex difference in the mean of liability ($d$) estimated under $H_3$ was evaluated by its ratio to the mean square of Spearman’s $\rho$ coefficient over all combinations of traits based on the pooled data.

### Table 3. Side-count frequencies for each population (data based on eight or fewer individuals were not included in analyses)

| Population | M | F | M | F | M | F | M | F | M | F | M | F | M | F | M | F |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Khoisan    | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| North Africa | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| Sub-Sahara | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| US African | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| Europe     | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| India      | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| Australia  | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |
| Jomon      | 18 | 115 | 40 | 28 | 72 | 64 | 29 | 25 | 237 | 175 | 69 | 54 | 26 | 20 | 35 | 40 |

* Harmonic mean of sample sizes over traits (rounded to integer value)

### Procedures for testing hypotheses

The MLE method was used to estimate the parameter values of $DLM(V_{ij}, M_{ij}, \delta_{ij})$ and obtain the test statistic for each hypothesis (i.e., constraint on $V_{ij}$ and $M_{ij}$). Appendix 1 illustrates the procedure for these analyses. The calculations used the programs coded by the author in Microsoft Visual Basic for Applications (VBA 7.1) and the Solver of Microsoft Excel 2016.

### Comparison of sex difference with group difference

The magnitude of sex difference in the mean of liability ($d$) estimated under $H_3$ was evaluated by its ratio to the group difference in the mean of liability ($D_g$) estimated un-
under $H_3$. This $D_g$ does not depend on sex because the sex difference is assumed to be constant under $H_3$. In this study, $D_g$ was defined as the weighted root mean square (WRMS) of inter-group difference in the mean of liability over all pairs of groups with the harmonic mean of sample sizes being used as the weight of each pair. The mean of liability of a group was obtained as the weighted average of the means of member populations. The group, not the population, was used as the unit in order for the value of $D_g$ to reflect the world-wide variability.

**Non-parametric tests**

The CMH method uses the male-to-female odds ratio (OR) to define the sex difference. Let $R_i$ denote the male-to-female OR of trait occurrence for population $i$. The hypothesis of constant sex difference across populations is expressed as $R_i = R$, which is tested by the Breslow–Day test of homogeneity of OR. The CMH estimate of common OR gives the estimate of this assumed constant $R$. The null hypothesis $R = 1$ for testing significance of the sex difference is examined by the CMH test of the conditional independence.

The Breslow–Day test of the homogeneity of OR can be regarded as a test of applicability of the genotype model to nonmetric data. Under the genotype model, the sex difference in the trait frequency must be explained by the difference in penetrance if the genotype frequency is equal between sexes, as is usually considered. On the other hand, to justify the population comparison based on trait frequencies, the penetrance must be constant across populations. (This applies to both the side-count and individual-count penetrance values because the latter is expressed as $2p - p^2$, using the former value, $p$.) Numerical simulations show that the male-to-female OR of trait occurrence is fairly stable for usual ranges of genotype frequency and penetrance if the penetrance is constant in each sex. Therefore, a substantial homogeneity across populations should be observed in the male-to-female OR if the genotype model is applicable to the trait expression, although there is a possibility that both the models exhibit sufficient goodness of fit because the liability-based homogeneity and the OR-based homogeneity

| Trait | Ainu | Japan | NE Asia | Siberia | Aleutian | Arctic | America | Polynesia |
|-------|------|-------|---------|---------|----------|--------|---------|----------|
| Nh*   | 58   | 124   | 34      | 47      | 123      | 435    | 467     | 559      |
| OMB   | 0.169| 0.172| 0.126   | 0.140   | 0.109    | 0.203  | 0.194   | 0.203    |
| AST   | 0.231| 0.136| 0.140   | 0.127   | 0.179    | 0.194  | 0.184   | 0.184    |
| PNb   | 0.186| 0.198| 0.272   | 0.183   | 0.190    | 0.225  | 0.133   | 0.222    |
| POS   | 0.192| 0.085| 0.302   | 0.245   | 0.275    | 0.331  | 0.338   | 0.213    |
| HYP   | 0.222| 0.233| 0.114   | 0.092   | 0.107    | 0.074  | 0.144   | 0.233    |
| PCh   | 0.274| 0.276| 0.307   | 0.209   | 0.281    | 0.310  | 0.238   | 0.188    |
| ICC   | 0.324| 0.226| 0.229   | 0.197   | 0.279    | 0.305  | 0.231   | 0.221    |
| SQS   | 0.025| 0.008| 0.023   | 0.019   | 0.022    | 0.038  | 0.012   | 0.023    |
| SPS   | 0.006| 0.000| 0.005   | 0.000   | 0.017    | 0.002  | 0.012   | 0.002    |
| MAR   | 0.071| 0.123| 0.126   | 0.153   | 0.151    | 0.126  | 0.136   | 0.300    |
| TYM   | 0.125| 0.275| 0.271   | 0.301   | 0.227    | 0.204  | 0.219   | 0.556    |
| FSP   | 0.165| 0.246| 0.170   | 0.169   | 0.161    | 0.146  | 0.190   | 0.187    |
| LPF   | 0.147| 0.189| 0.088   | 0.048   | 0.119    | 0.063  | 0.045   | 0.085    |
| CIV   | 0.038| 0.009| 0.028   | 0.012   | 0.044    | 0.039  | 0.006   | 0.051    |
| PTB   | 0.260| 0.250| 0.191   | 0.114   | 0.156    | 0.044  | 0.036   | 0.135    |
| CLN   | 0.048| 0.132| 0.052   | 0.030   | 0.064    | 0.093  | 0.057   | 0.261    |
| SOF   | 0.196| 0.208| 0.373   | 0.375   | 0.523    | 0.549  | 0.607   | 0.542    |
| FRG   | 0.158| 0.179| 0.222   | 0.308   | 0.320    | 0.124  | 0.203   | 0.310    |
| TRS   | 0.073| 0.034| 0.109   | 0.070   | 0.063    | 0.066  | 0.055   | 0.037    |
| OPT   | 0.014| 0.038| 0.063   | 0.026   | 0.084    | 0.114  | 0.032   | 0.048    |
| ORB   | 0.122| 0.166| 0.373   | 0.254   | 0.463    | 0.346  | 0.278   | 0.174    |
| CON   | 0.139| 0.207| 0.303   | 0.307   | 0.463    | 0.417  | 0.611   | 0.344    |
| JAP   | 0.640| 0.658| 0.310   | 0.401   | 0.302    | 0.268  | 0.363   | 0.177    |
| M3U   | 0.284| 0.257| 0.377   | 0.517   | 0.266    | 0.319  | 0.438   | 0.089    |
| MEN   | 0.223| 0.143| 0.144   | 0.095   | 0.176    | 0.125  | 0.140   | 0.077    |
| MHB   | 0.134| 0.042| 0.064   | 0.028   | 0.014    | 0.123  | 0.074   | 0.262    |
| BUC   | 0.182| 0.122| 0.057   | 0.052   | 0.049    | 0.000  | 0.042   | 0.126    |
| M3L   | 0.256| 0.204| 0.333   | 0.396   | 0.180    | 0.287  | 0.375   | 0.121    |
| TRM   | 0.068| 0.056| 0.200   | 0.198   | 0.283    | 0.274  | 0.207   | 0.407    |
| ATA   | 0.104| 0.004| 0.014   | 0.038   | 0.109    | 0.109  | 0.127   | 0.110    |
| ATB   | 0.000| 0.000| 0.000   | 0.019   | 0.091    | 0.013  | 0.137   | 0.172    |

* Harmonic mean of sample sizes over traits (rounded to integer value)
of sex difference are numerically close to each other if the inter-population variability of trait frequency is not so large.

The analyses based on the CMH method used individual-count data and IBM SPSS Statistics version 25.

**Results**

**Homogeneity of variance of liability**

The constancy of variance of IICL was tested between sexes first and then across populations (Table 4). No trait exhibited statistically significant sex difference in the variance of IICL. The sum of chi-square values indicated that $H_0$ was acceptable when tested collectively over traits ($P = 0.511$). Statistically significant heterogeneity of variance among populations was exhibited by only four traits (FRG at the 1% level, and BUC, AST, and FSP at the 5% level), none of which satisfied Benjamini and Hochberg’s (1995) criterion of $q < 0.05$ to control the false discovery rate (FDR). The sum of the chi-square values indicated that $H_2$ was acceptable when tested collectively over traits ($P = 0.217$).

**Constant sex difference across populations**

Table 5 shows the results for the test of homogeneity of sex difference ($H_1$) and that of the null hypothesis of the assumed constant sex difference ($H_0$) under the hypothesis of constant variance of IICL across populations ($H_1$ and $H_2$). Three traits (TYM, CLN, and MEN) exhibited a heterogeneity of sex difference that was statistically significant at the 5% level, but no trait satisfied even $q < 0.5$ to control the FDR. The sum of chi-square values indicated that $H_1$ was acceptable when tested collectively over traits ($P = 0.066$).

Statistically significant sex difference was exhibited by 19 traits (TYM, FRG, AST, PNB, ORB, FSP, PCP, CIV, SPS, MEN, JAP, SOF, PNB, LFH, SQS, CLN, M3U, MHB, and ATA) at the 5% level, of which the first 14 cases were significant at the 0.1% level, and the next three cases were significant at the 1% level. All these traits, apart from ATA, satisfied the Benjamini–Hochberg criterion $q < 0.05$ to control the FDR. The absolute magnitude of sex difference was below 1 in all traits, and the 95% CI was within the range from −1 to 1, except for two traits (SPS and ORB). This means that the difference in the mean of liability between sexes was
less than the inter-side variability for all traits, although not necessarily negligible as shown later.

Relative magnitude of sex difference

Table 6 shows the ratios of magnitude of sex difference to the within-population standard deviation of liability ($d/[\sqrt{N} + 1]; S/W$ ratio) and to the WRMS of group difference ($d/D; S/G$ ratio).

The traits can be classified into three types based on the statistical significance and $S/W$ ratio of the assumed constant sex difference. Type 0 includes 12 traits (MAR, CON, HYP, POS, ICC, TRS, OMB, M3L, OPT, BUC, TRM, and ATB) with no statistical significance and very low $S/W$ ratios (1–9%), excepting 15% for ATB). Type 1 includes 9 traits (M3U, MHB, SOF, JAP, CLN, SQS, PNB, LPF, and MEN) with statistical significance at 5% but rather low $S/W$ ratios (9–17%). Type 2 includes 10 traits (SPS, ORB, CIV, TYM, FRG, PCP, AST, PTB, ATA, and FSP) with statistically significant sex difference ($P < 0.001$, excepting $P = 0.044$ for ATA) and high $S/W$ ratios (over 20%). The $S/G$ ratio was 1.6–27.6% for type 0, 16.4–66.1% for type 1, and 47.4–128.4% for type 2.

Non-parametric test of sex difference

Table 7 shows the results of the CMH tests of sex difference. The result of the Breslow–Day test of homogeneity of OR indicates that the deviation from the homogeneity of sex difference ($R_1 - R_2$) was statistically significant in only five traits (TYM at the 1% level, and CLN, OPT, MEN, and M3L at the 5% level), none of which satisfied $q < 0.05$ to control the FDR. When tested collectively over traits, however, the deviation was statistically significant ($P = 0.0013$). The Pearson’s correlation coefficient of log-transformed probability with that of the DLM–MLE method was 0.878.

The null hypothesis for the assumed constant sex difference ($R = 1$) was rejected in 21 traits (TYM, FRG, AST, PTB, FSP, ORB, PCP, CIV, SPS, MEN, JAP, SOF, CLN, LPF PNB, SQS, M3U, MHB, TRM, OMB, and ATA) at the
5% level, of which the first 13 cases were significant at the 0.1% level, and the next four cases were significant at the 1% level. The first 18 of them satisfied $q < 0.05$ to control the FDR. These are the same traits that satisfied the same criterion in the analysis of the DLM–MLE method. The Pearson’s correlation coefficient of log-transformed probability between the methods was very high ($r = 0.993$).

**Discussion**

**Homogeneity of variance of liability**

The hypothesis of equality of variance of liability between sexes proved to be acceptable. This seems natural considering that the variability of genetic and environmental factors affecting the liability should be identical between sexes. The hypothesis of homogeneity of variance across populations was substantially acceptable. This suggests that most of the factors affecting the liability are common to all populations. These results justify the comparison of the means of liability between sexes and populations in the subsequent analyses.

**Homogeneity of sex difference in the mean of liability**

The results based on the world-wide 16 populations indicate that the sex difference in the mean of liability is basically constant across populations for each trait. This is in accordance with the assumption of the multifactorial threshold model that the factors affect the liability additively even though the magnitude of sex difference may be much greater than these factors. Perizonius (1979) raised a question about whether the same traits show sex differences in various populations. This study has given an affirmative answer to this question, and, moreover, showed that both the direction and magnitude of the differences should be assumed to be constant.

The result here disagrees with the conclusion of Berry (1975), who emphasized inconsistency in the direction and magnitude of sex difference among populations. However, the result she presented appears to be generally in accordance with the hypothesis of constant sex difference. In her result, the direction of difference was consistent in five of six traits where multiple samples exhibited statistically significant sex differences. The only exception was the highest nuchal line, where one of three samples exhibited difference...
at the 5% level in the opposite direction to that of other two. Therefore, the conclusion did not faithfully reflect the result of her analysis. This was probably because she preferred the hypothesis of random sex difference, which could explain the failure of Berry and Berry (1967) to detect significant sex differences when they pooled the same samples.

Magnitude of sex difference

In the 12 traits classified as type 0, the constant sex difference was statistically not significant at the 5% level, and the nine traits classified as type 1 exhibited statistically significant but fairly small sex differences. The magnitude of sex difference for these 21 traits seems to be enough low to regard sex as one of the numerous factors affecting the liability, hence these traits may be safely used for population comparisons without distinguishing sex. For the remaining 10 traits classified as type 2 because of their statistically significant higher magnitudes of sex difference, it is not justifiable to regard sex as one of the numerous factors contributing to normal distribution of liability. A limited extent of uneven sex composition of samples may not so much affect the results of population comparisons because the assumed constant sex difference was less than the average group difference except for two traits. However, since the estimates of average group difference were based on the world-wide grouping, the S/G ratios obtained here should be considered as minimum values. In a comparison of closer populations, the effect of uneven sex composition can become negligible.

Mouri (1988), who compared the cranial nonmetric data of Jomon populations in West Japan using the mean character difference as the distance measure, reported the distances of Jomon populations in West Japan using the mean character difference as the distance measure, reported the distances between sexes as comparable to those between populations. Therefore, the conclusion did not faithfully reflect the result of her analysis. This was probably because she preferred the hypothesis of random sex difference, which could explain the failure of Berry and Berry (1967) to detect significant sex differences when they pooled the same samples.

Table 7. Non-parametric test of sex difference

| Trait | Homogeneity of sex difference ($R_i = R$) | Common sex difference (common $R$) | No common sex difference ($R = 1$) |
|-------|-----------------------------------------|-----------------------------------|----------------------------------|
|       | $\chi^2$ | d.f. | $P$ | log(OR) | [95% CI] | $\chi^2$ | $P$ |
| OMB   | 8.96     | 15   | 0.880 | 0.132 | [0.003, 0.262] | 3.88 | 0.049 |
| AST   | 21.23    | 15   | 0.130 | 0.452 | [0.332, 0.573] | 54.33 | 2E-13 |
| PNB   | 12.82    | 15   | 0.616 | 0.182 | [0.071, 0.293] | 10.16 | 0.001 |
| POS   | 17.33    | 15   | 0.299 | 0.007 | [-0.105, 0.119] | 0.01 | 0.927 |
| HYP   | 14.15    | 15   | 0.514 | -0.024 | [-0.137, 0.089] | 0.15 | 0.697 |
| PCP   | 8.35     | 15   | 0.909 | 0.416 | [0.285, 0.546] | 38.79 | 5E-10 |
| ICC   | 24.43    | 15   | 0.058 | 0.069 | [-0.040, 0.177] | 1.47 | 0.225 |
| SQS   | 14.50    | 15   | 0.488 | 0.311 | [0.102, 0.521] | 8.19 | 0.004 |
| SPS   | 17.01    | 13   | 0.199 | 1.265 | [0.809, 1.720] | 32.23 | 1E-08 |
| MAR   | 15.09    | 15   | 0.445 | -0.039 | [-0.162, 0.085] | 0.34 | 0.562 |
| TYM   | 30.98    | 15   | 0.009 | -0.423 | [-0.527, -0.318] | 62.72 | 2E-15 |
| FSP   | 21.26    | 15   | 0.129 | -0.387 | [-0.501, -0.274] | 44.42 | 3E-11 |
| LPF   | 11.32    | 14   | 0.661 | 0.292 | [0.120, 0.465] | 10.77 | 0.001 |
| CIV   | 11.37    | 13   | 0.579 | 0.575 | [0.382, 0.769] | 33.94 | 6E-09 |
| PTB   | 19.15    | 15   | 0.207 | 0.448 | [0.324, 0.571] | 50.32 | 1E-12 |
| CLN   | 27.10    | 15   | 0.028 | 0.238 | [0.098, 0.377] | 11.06 | 9E-04 |
| SOF   | 24.29    | 15   | 0.060 | -0.206 | [-0.311, -0.102] | 14.73 | 1E-04 |
| FRG   | 20.07    | 15   | 0.169 | -0.421 | [-0.531, -0.311] | 55.90 | 8E-14 |
| TRS   | 17.95    | 15   | 0.265 | 0.124 | [-0.063, 0.311] | 1.57 | 0.211 |
| OPT   | 22.65    | 13   | 0.046 | -0.095 | [-0.333, 0.143] | 0.52 | 0.473 |
| ORB   | 12.46    | 13   | 0.491 | 0.544 | [0.378, 0.711] | 41.08 | 1E-10 |
| CON   | 15.07    | 14   | 0.373 | -0.036 | [-0.194, 0.122] | 0.17 | 0.682 |
| JAP   | 10.64    | 15   | 0.778 | -0.301 | [-0.434, -0.168] | 19.53 | 1E-05 |
| M3U   | 16.46    | 14   | 0.286 | -0.209 | [-0.355, -0.063] | 7.73 | 0.005 |
| MEN   | 26.32    | 15   | 0.035 | 0.372 | [0.224, 0.520] | 24.09 | 9E-07 |
| MHB   | 12.61    | 15   | 0.633 | 0.177 | [0.026, 0.327] | 5.11 | 0.024 |
| BUC   | 22.56    | 15   | 0.094 | 0.183 | [-0.015, 0.381] | 3.13 | 0.077 |
| M3L   | 25.97    | 15   | 0.038 | -0.091 | [-0.236, 0.055] | 1.41 | 0.234 |
| TRM   | 5.60     | 12   | 0.935 | 0.191 | [0.008, 0.375] | 4.00 | 0.046 |
| ATA   | 9.14     | 6    | 0.166 | 0.440 | [0.021, 0.860] | 3.86 | 0.049 |
| ATB   | 9.15     | 4    | 0.057 | 0.275 | [-0.186, 0.736] | 1.13 | 0.288 |

Total 525.98 431 0.001 2 546.73 1E-95

1Breslow–Day test.
2Cochran-Mantel-Haenszel estimate of common OR.
3Cochran-Mantel-Haenszel test of conditional independence.

**P < 0.05, ***P < 0.01, ****P < 0.001.**
tude of sex difference in comparison with differences between populations. Since sex difference of liability proved to be basically constant across populations, the statistical significance and relative magnitude of the assumed constant sex difference in Table 6 would provide references for selection of such traits. Since the S/G ratio depends on the populations included in the comparison, the traits most appropriate for a comparison may vary depending on the level of comparison and the combination of populations.

**Comparison of results with non-parametric tests**

The results of the tests of sex difference were similar between the two methods despite the difference in their definitions of sex difference. This is not surprising because the numerical effect caused by the difference in the definition of sex difference tends to be small when the trait frequency varies within a limited range. It is notable, however, that the homogeneity of sex difference in the mean of liability exhibited a higher goodness of fit than the homogeneity of sex difference in the form of male-to-female OR of trait occurrence. This indicates that the multifactorial threshold model is more appropriate than the genotype model for explaining the expression of nonmetric traits. The genotype model seems to be the only logically consistent model that justifies the use of individual counts for recording the expression of bilateral nonmetric traits. The ASUDA system in dental anthropology has adopted this model for justification of the individual counts (Turner et al., 1991).

Ossenberg (1981) adopted the threshold model with a single liability and two thresholds proposed by Falconer (1960) to argue for the use of side counts. Although the strangeness of the two thresholds invited criticisms (McGrath et al., 1984; Hallgrímsson et al., 2005), and the conventional count method has been predominantly used for some reason, the use of side counts she advocated is a mathematical requirement under the DLM (Tagaya, 2019). Thus, the results in this study indicate that the world-wide data of nonmetric traits Ossenberg collected over her lifetime support the use of side counts she proposed early in her academic life.

**Acknowledgments**

My deep gratitude is due to the late Dr. Nancy Suzanne Ossenberg who made her valuable data open to our academic community.

**References**

Anscombe F.J. (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika, 35: 246–254.

Benjamini Y. and Hochberg Y. (1990) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57: 289–300.

Berry A.C. (1975) Factors affecting the incidence of non-metrical skeletal variants. Journal of Anatomy, 120: 519–535.

Berry A.C. and Berry R.J. (1967) Epigenetic variation in the human cranium. Journal of Anatomy, 101: 361–379.

Breslow N.E. and Day N.E. (1880) Statistical Methods in Cancer Research, Vol. 1: The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer.

Cochran W.G. (1954) Some methods for strengthening the common $\chi^2$ tests. Biometrics, 10: 417–451.

Corruccini R.S. (1974) An examination of the meaning of cranial discrete traits for human skeletal biological studies. American Journal of Physical Anthropology, 40: 425–445.

Dodo Y. and Ishida H. (1987) Incidences of nonmetric cranial variants in several population samples from East Asia and North America. Journal of the Anthropological Society of Nippon, 95: 161–177.

Falconer D.S. (1960) Introduction to Quantitative Genetics. Oliver & Boyd, Edinburgh and London.

Hallgrímsson B., Donnabháin B.O., Blom D.E., Lozada M.C., and Willmore K.T. (2005) Why are rare traits unilaterally expressed? Trait frequency and unilateral expression for cranial nonmetric traits in humans. American Journal of Physical Anthropology, 128: 14–25.

Hanihara T. (2008) Morphological variation of major human populations based on nonmetric dental traits. American Journal of Physical Anthropology, 136: 169–182.

Hubbe M., Hanihara T., and Harvati K. (2009) Climate signatures in the morphological differentiation of worldwide modern human populations. Anatomical Record, 292: 1720–1733.

Mantel N. and Haenszel W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22: 719–748.

McGrath J.W., Cheverud J.M., and Buikstra J.E. (1984) Genetic correlations between sides and heritability of asymmetry for nonmetric traits in rhesus macaques on Cayo Santiago. American Journal of Physical Anthropology, 64: 401–411.

Mouri T. (1976) A study of non-metrical cranial variants of the modern Japanese in the Kinki district. Journal of the Anthropological Society of Nippon, 84: 191–203.

Mouri T. (1988) Incidences of cranial nonmetric characters in five Jomon populations from West Japan. Journal of the Anthropological Society of Nippon, 96: 319–337.

Ossenberg N.S. (1981) An argument for the use of total side frequencies of bilateral nonmetric skeletal traits in population distance analysis: the regression of symmetry on incidence. American Journal of Physical Anthropology, 54: 471–479.

Ossenberg N.S. (1994) Origins and affinities of the native peoples of northwestern North America: the evidence of cranial nonmetric traits. In: Bonnichsen R. and Steele D.G. (eds.), Method and Theory for Investigating the Peopling of the Americas. Center for the Study of the First Americans, Oregon State University, pp. 79–115.

Ossenberg N.S. (2013a) Brief communication: Cranial nonmetric trait database on the internet. American Journal of Physical Anthropology, 152: 551–553.

Ossenberg N.S. (2013b) Cranial nonmetric trait database. Kingston, Ontario, Canada: Nancy S. Ossenberg [producer]. Queen’s University Library, Data and Government Information Centre [distributor].

Ossenberg N.S., Dodo Y., Maeda T., and Kawakubo Y. (2006) Ethnogenesis and craniofacial change in Japan from the perspective of nonmetric traits. Anthropological Science, 114: 99–115.

Perizonius W.R.K. (1979) Non-metric cranial traits: sex difference and age dependence. Journal of Human Evolution, 8: 679–684.

Shigematsu M., Ishida H., Goto M., and Hanihara T. (2004) Morphological affinities between Jomon and Ainu: reassessment of discrete traits for human skeletal biological studies. American Journal of Physical Anthropology, 115: 161–172.

Tagaya A. (2019) Tetrachoric correlation of bilateral nonmetric traits: a defect in the conventional procedure and a proposal for two alternative estimation methods. Anthropological Science, 127: 39–45. Erratum, Anthropological Science, 128: 41, 2020.

Turner C.G. II, Nichol C.R., and Scott G.R. (1991) Scoring procedures for key morphological traits of the permanent dentition: the Arizona State University Dental Anthropology System. In: Kelley M.A. and Larsen C.S. (eds.), Advances in Dental Anthropology. Wiley-Liss, New York, pp. 13–31.
Appendix 1. Examples of testing hypotheses by the MLE method

Table S1 illustrates the MLE procedure for testing the hypotheses about the distribution of liability of DLM accommodating side difference. Only two populations are used here, but the number of populations will be expressed as $k$ for extension to a general case. The programs required for calculations are provided in Appendix 2.

| Hypothesis (constraint on DLM parameters) | Popul.  | Sex | MLE estimates | Test of hypothesis |
|------------------------------------------|---------|-----|---------------|--------------------|
|                                          |         |     | Parameters    | Expected counts     | Total d.f.* | Conditional d.f.* | P     |
| (A) DLM with no constraint               | Europe  | M   | $1.807$ 0.009 $0.097$ | $229$ $24$ $17$ $29$ | $299$       |                   |       |
|                                          |         |     | $2.835$ $–3.585$ $0.143$ | $181$ $10$ $6$ $17$ | $214$       |                   |       |
|                                          | Japan   | M   | $2.289$ $–1.257$ $0.039$ | $221$ $31$ $27$ $82$ | $361$       |                   |       |
|                                          |         |     | $2.137$ $–1.911$ $0.099$ | $150$ $17$ $12$ $29$ | $208$       |                   |       |
| (B) $H_1(V_i = V_j)$                     | Europe  | M   | $2.100$ $–2.245$ $0.106$ | $230.5$ $21.8$ $15.0$ | $31.7$ $299$ |                   | $0.00$ |
|                                          |         |     | $2.100$ $–2.813$ $0.117$ | $179.7$ $12.1$ $8.0$ | $14.3$ $214$ |                   |       |
|                                          | Japan   | M   | $2.239$ $–1.235$ $0.039$ | $220.5$ $31.5$ $27.5$ | $81.4$ $361$ |                   |       |
|                                          |         |     | $2.239$ $–1.981$ $0.102$ | $150.4$ $16.5$ $11.5$ | $29.6$ $208$ |                   |       |
| (C) $H_2(V_i = V_j)$ and $H_1$           | Europe  | M   | $2.185$ $–2.313$ $0.109$ | $230.9$ $21.2$ $14.4$ | $32.4$ $299$ |                   | $2.35$ |
|                                          |         |     | $2.185$ $–2.900$ $0.121$ | $179.9$ $11.8$ $7.7$ | $14.6$ $214$ |                   | $0.726$|
|                                          | Japan   | M   | $2.185$ $–1.212$ $0.038$ | $220.0$ $32.2$ $28.1$ | $80.7$ $361$ |                   | $2.48$ |
|                                          |         |     | $2.185$ $–1.944$ $0.100$ | $150.2$ $16.7$ $11.8$ | $29.3$ $208$ |                   | $0.713$|
| (D) $H_3(d_i = d_j)$ and $H_1$ and $H_2$| Europe  | M   | $2.186$ $–2.283$ $0.108$ | $229.7$ $21.5$ $14.6$ | $33.2$ $299$ |                   | $14.49$|
|                                          |         |     | $2.186$ $–2.954$ $0.122$ | $181.0$ $11.5$ $7.4$ | $14.0$ $214$ |                   | $0.001$|
|                                          | Japan   | M   | $2.186$ $–1.233$ $0.038$ | $221.3$ $32.0$ $28.0$ | $79.7$ $361$ |                   | $14.49$|
|                                          |         |     | $2.186$ $–1.904$ $0.100$ | $149.0$ $17.0$ $11.9$ | $30.1$ $208$ |                   | $0.001$|
| (E) $H_4(d = 0)$ and $H_1$ and $H_2$    | Europe  | M   | $2.207$ $–2.558$ $0.114$ | $239.3$ $19.3$ $12.9$ | $27.5$ $299$ |                   | $14.49$|
|                                          |         |     | $2.207$ $–2.558$ $0.112$ | $171.3$ $13.8$ $9.2$ | $19.7$ $214$ |                   | $0.001$|
|                                          | Japan   | M   | $2.207$ $–1.477$ $0.039$ | $235.0$ $30.1$ $26.2$ | $69.7$ $361$ |                   | $14.49$|
|                                          |         |     | $2.207$ $–1.477$ $0.094$ | $135.2$ $19.0$ $13.7$ | $40.1$ $208$ |                   | $0.001$|

*The formula in parenthesis gives the degrees of freedom of the chi-square value when $k$ populations are included in the analysis.

(A) To accomplish the complete fit of the DLM to the observed data, three parameters ($V, M, \delta$) are required for each sex of each population. Therefore, $6k$ (or $12$) parameters (i.e. $V_i, M_i,$ and $\delta_i$) are used in total.

(B) Hypothesis $H_1$ assumes that the variance of liability is equal between sexes (i.e. $V_i = V_j$), hence the number of independent parameters (i.e. $V_i, M_i,$ and $\delta_i$) becomes $5k$, and the decrease of the number of independent parameters, $6k – 5k = k$ ($= 2$) is the degrees of freedom (d.f.) of the chi-square statistic.

(C) Hypothesis $H_2$ assumes that the common-to-sex variance of liability is constant across populations (i.e. $V = V_j$). Under $H_1$ and $H_2$ (i.e. $V_i = V_j$), the number of independent parameters (i.e. $V_i, M_i,$ and $\delta_i$) becomes $4k + 1$, hence d.f. of the chi-square statistic is $6k – (4k + 1) = 2k – 1$ ($= 3$). To test $H_2$ under $H_1$, the increments of the chi-square value and d.f. from those under $H_1$ are used as the chi-square statistic and its d.f.

(D) Hypothesis $H_3$ assumes that the difference in the mean of liability between sexes is constant across populations (i.e. $d_i = d_j$). Under $H_1$ and $H_2$ and $H_3$ (i.e. $V_j = V$ and $d_i = d_j$), the number of independent parameters (i.e. $V, M_i,$ and $\delta_i$) becomes $3k + 2$, hence d.f. of the chi-square statistic is $6k – (3k + 2) = 3k – 2$ ($= 4$). To test $H_3$ under $H_1$ and $H_2$, the increments of the chi-square value and d.f. from those under $H_1$ and $H_2$ are used as the chi-square statistic and its d.f.

(E) Hypothesis $H_4$ assumes that the difference in the mean of liability between sexes assumed to be constant across populations is equal to 0 (i.e. $d = 0$). Under $H_1$ and $H_2$ and $H_1$ and $H_1$ (i.e. $V_i = V$ and $d = 0$), the number of independent parameters (i.e. $V_i, M_i,$ and $\delta_i$) becomes $3k – 1$, hence d.f. of the chi-square statistic is $6k – (3k + 1) = 3k – 1$ ($= 5$). To test $H_4$ under $H_1$ and $H_2$ and $H_3$, the increments of chi-square value and d.f. from those under $H_1$ and $H_2$ and $H_3$ are used as the chi-square statistic and its d.f.
Appendix 2. VBA programs and usage instructions

The VBA functions provided here return the results in an array. To use an array function, select a range of cells, enter any formula including the function, e.g., “= DLM_Q(A3, B3, C3)*(D3 + 0.75)” in the first cell of the range, and press the Enter key while pressing the Shift and Control keys. The example above calculates the expected count quartet for sample size plus 3/4 if the values of V, M, and δ are stored in A3, B3, and C3, respectively, and the sample size is stored in D3. In the following descriptions, the DLM parameters V, M, and δ are represented by V, Mean, and Dev, respectively, in the VBA program codes.

1. Function DLM_Q(V, M, δ)
   This function calculates the probability quartet under DLM (V, M, δ) and returns it in an array. The quartet probability densities for IICL value x calculated by DLM_Q_Dist(x, V, M, δ) are integrated from M − 6√V to M + 6√V using Simpson’s quadrature rule.

2. Function DLM_Q_Dist(x, V, M, δ)
   This function calculates the quartet of probability densities for IICL value x under DLM (V, M, δ). For a given value of x, the conditional probabilities of trait occurrence on the left and right sides are calculated as \( P_L = h(x + δ) \) and \( P_R = h(x - δ) \) using the cumulative standard normal distribution function. Since the left- and right-side liabilities are mutually independent when \( x \) is fixed, the quartet probability densities for x are obtained as \( \{Q_L, Q_R, \Phi_L, \Phi_R\} \), where \( Q_L = 1 - P_L \), \( Q_R = 1 - P_R \), and \( \Phi(x) \) is the probability density of x distributed as N(M, V).

3. Function Chisq_DLM_Q(X, V, M, δ, ID)
   This function calculates the chi-square statistic to test the goodness of fit of DLM(V, M, δ) to observed count data stored in the range X, and returns the chi-square value and the given or estimated values of M and δ in an array. Both or one of the M and δ can be left blank to use the estimates based on the data stored in X and V. The parameter values for complete fit DLM are obtained using the Solver. Find the value of V that minimizes the function value of Chisq_DLM_Q with both M and δ left blank. The value becomes equal to zero when the complete fit is achieved.

4. Iterative application of Solver for MLE solution
   To obtain the MLE solution, it is often necessary to iteratively apply the Solver to each parameter (as the ‘changing variable’) in turn until the chi-square statistic converges to its minimum value. This can be automated using a program created by the VBA Macro Recorder, with some modification of codes and inclusion of the Solver in the references. The use of the initial estimates obtained by the blank-estimation method above enables rapid conversion.

Each border of 95% CI of a parameter is obtained as the value that increases the minimum chi-square value by 3.84146 (the chi-square value with one degree of freedom for \( P = 0.05 \)). This requires iterative application of Solver to find values of the initial estimates obtained by the blank-estimation method above until the chi-square statistic converges to its minimum value. This can be automatized using a program created by the VBA Macro Recorder, with some modification of codes and inclusion of the Solver in the references. The use of the initial estimates obtained by the blank-estimation method above enables rapid conversion.

Programs created by A. Tagaya for analyses based on the DLM

DLM_Q, DLM_Q_Dist, and Chisq_DLM_Q return the results in an array.

Function DLM_Q(V, Mean, Dev, Optional k = 10) As Variant
'Calculates the quartet of probabilities \{P00, P01, P10, P11\} for DLM(V, Mean, Dev)
'Uses DLM_Q_Dist and Simpson
'Dimension of F0 to F3 must be equal to N (=k*Nsd*2) or greater.
Dim F0(960), F1(960), F2(960), F3(960), F(4)
SD = Sqr(V)
Nsd = 6
N = k * Nsd * 2
dx = SD / k
x0 = Mean - Nsd * SD
For i = 0 To N
  X = x0 + i * dx
  FF = DLM_Q_Dist(X, V, Mean, Dev)
  'Calculate prob. densities for x
  F0(i) = FF(0, 0) 'Set i-th prob. density in \{F0(i), F1(i), F2(i), F2(i)\}
  F1(i) = FF(1, 0)
  F2(i) = FF(2, 0)
  F3(i) = FF(3, 0)
Next i
TF = F(0) + F(1) + F(2) + F(3) 'adjustment to make total prob=1

'}
For $i = 0$ To 3
    FF($i$, 0) = $F(i)$ / $TF$ 'Set the results for use as a vertical array
    FF(0, $i$) = $F(i)$ / $TF$ 'Set the results for use as a horizontal array
Next i
DLM_Q = FF
End Function

Function DLM_Q_Dist(X, V, Mean, Dev) As Variant
    'Calculates quartet probab. densities for a given IICL value X under DLM(V, Mean, Dev).
    'Returns the values in an array.
Dim F(4), FF(4, 4) As Variant
Px = Application.WorksheetFunction.Norm_Dist(X, Mean, Sqr(V), 0) 'Prob. density of X
'For given X, the total liability is distributed with var=1 and mean=X+Dev (or X-Dev).
'Hence the probability for the total liability to exceed 0 (=threshold) is given as follows.
PL = Application.WorksheetFunction.Norm_Dist(X+Dev, 0, 1, 1)
PR = Application.WorksheetFunction.Norm_Dist(X-Dev, 0, 1, 1)
QL = 1 - PL           'Conditional probability of negative expression on the left side
QR = 1 - PR           'Conditional probability of negative expression on the right side
F(0) = QL * QR * Px   ' {F(0), F(1), F(2), F(3)} is the quartet of prob. densities
F(1) = PL * QR * Px
F(2) = QL * PR * Px
F(3) = PL * PR * Px
For i = 0 To 3
    FF(i, 0) = F(i)   'Set the results for use as a vertical array
    FF(0, i) = F(i)   'Set the results for use as a horizontal array
Next i
DLM_Q_Dist = FF
End Function

Function Simpson(F, N, dx)
    'Integration by Simpson's quadrature rule.
    'N must be an even number.
    'F is an array containing N+1 function values in F(0) to F(N).
    If Int(N / 2) * 2 < N Then Simpson = "N is not even": Exit Function
    S = F(0) - F(N)
    For i = 1 To N - 1 Step 2
        S = S + 4 * F(i) + 2 * F(i + 1)
    Next i
    Simpson = S * dx / 3
End Function

Function Chisq_DLM_Q(X, V, Optional Mean = "", Optional Dev = "", Optional ID = 1) As Variant
    'Chi-square statistic based on the likelihood of DLM(V, Mean, Dev) for count data X
    'Mean and/or Dev are estimated from X and V if one or both of them are left blank.
    'Set ID=0 to avoid modification of data X by adding dN={1/8, 1/4, 1/4, 1/8}.
Dim XX(4), F(4), Chi2(3, 3)
dN = Array(0.125, 0.25, 0.25, 0.125)
T = 0
For i = 0 To 3
    A = X.Cells(i + 1) + dN(i) * ID 'add {1/8, 1/4, 1/4, 1/8} to avoid zero frequency
    T = T + A
    XX(i) = A
Next i
Next i
PL = (XX(1) + XX(3)) / T 'side-count frequency for left side
PR = (XX(2) + XX(3)) / T 'side-count frequency for right side
If Mean = "" Or dev = "" Then
    MeanL = Application.WorksheetFunction.Norm_S_Inv(PL) * Sqr(V + 1)
    MeanR = Application.WorksheetFunction.Norm_S_Inv(PR) * Sqr(V + 1)
    Mean0 = (MeanL + MeanR) / 2
    Dev0 = MeanL - Mean0
End If
If Mean = "" Then Mean = Mean0 'If Mean is blank, the estimate is used
If Dev = "" Then Dev = Dev0 'If Dev is blank, the estimate is used
FF = DLM_Q(V, Mean, Dev) 'Store the quartet probabilities in FF (two-dimensional)
For i = 0 To 3
    F(i) = T * FF(i, 0) 'F(i) is the count expected for sample size T (=N+0.75*ID)
Next i
LnLR = 0 'LnLR: log likelihood ratio
For i = 0 To 3
    If XX(i) * F(i) > 0 Then LnLR = LnLR + XX(i) * Log(F(i) / XX(i))
Next i
Chi2(0,0) = -2 * LnLR
Chi2(0,1) = Mean: Chi2(1,0) = Mean
Chi2(0,2) = Dev: Chi2(2,0) = Dev
Chisq_DLM_Q = Chi2
End Function