

SHORT REPORT

Open Access

W_d^* -test: robust distance-based multivariate analysis of variance



Bashir Hamidi^{1,2}, Kristin Wallace³, Chenthamarakshan Vasu⁴ and Alexander V. Alekseyenko^{1,2,3,5*} 

Abstract

Background: Community-wide analyses provide an essential means for evaluation of the effect of interventions or design variables on the composition of the microbiome. Applications of these analyses are omnipresent in microbiome literature, yet some of their statistical properties have not been tested for robustness towards common features of microbiome data. Recently, it has been reported that PERMANOVA can yield wrong results in the presence of heteroscedasticity and unbalanced sample sizes.

Findings: We develop a method for multivariate analysis of variance, W_d^* , based on Welch MANOVA that is robust to heteroscedasticity in the data. We do so by extending a previously reported method that does the same for two-level independent factor variables. Our approach can accommodate multi-level factors, stratification, and multiple post hoc testing scenarios. An R language implementation of the method is available at <https://github.com/alekseyenko/WdStar>.

Conclusion: Our method resolves potential for confounding of location and dispersion effects in multivariate analyses by explicitly accounting for the differences in multivariate dispersion in the data tested. The methods based on W_d^* have general applicability in microbiome and other 'omics data analyses.

Keywords: Welch MANOVA, Distance MANOVA, Heteroscedastic test

Introduction

Beta diversity analyses or community-wide ecological analyses are important tools for understanding the differentiation of the entire microbiome between experimental conditions, environments, and treatments. For these analyses, specialized distance metrics are used to capture the multivariate relationships between each pair of samples in the dataset. Analysis of variance-like techniques, such as PERMANOVA [1], may then be used to determine if an overall difference exists between conditions. The distances use all of the measured taxa information simultaneously without the need to explicitly estimate individual covariances. The utility of these methods is hard to underestimate as virtually every recent major microbiome report has used some form of a community-

wide association analysis. On many occasions, the comparison reveals major differences between the groups. However, one is not guaranteed to find one. For example, in Redel et al. [2], the authors have found that there are significant differences in cutaneous microbiota in diabetic vs. non-diabetic subject feet, but not on their hands (see fig. 5). This lack of difference is an important indicator about the potential pathobiological processes that lead to diabetic foot ulcers. Therefore, getting the correct result in such comparisons is important. The Redel et al. analysis can ultimately be achieved by pairwise comparisons only (diabetic vs. non-diabetic); however, many study designs have more than two groups that need to be considered simultaneously. Dietary intervention studies among others often include several experimental groups. For example, Cox et al. [3] analysis of the impact of diet on the murine gut microbiome included animal groups receiving low fat, high fat, and high fat with fiber supplement diets. Although it is possible to treat such design using multi-way comparisons of dietary fat and dietary fiber, a simultaneous analysis of all three groups can be

*Correspondence: alekseye@muscc.edu

¹Program for Human Microbiome Research, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA

²Biomedical Informatics Center, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA

Full list of author information is available at the end of the article



more intuitive. Hence, there is a need for methods that can compare more than two experimental groups at the same time. PERMANOVA among other methods allows for such analyses.

From the statistical stand point, community-wide analyses test the hypothesis that the data from two or more conditions share the location parameter (centroid or multivariate mean). Caution, however, needs to be taken to ensure that potential violations of assumptions do not lead to adverse statistical behavior of PERMANOVA. Two such assumptions that are commonly violated are the multivariate uniformity of variability (homoscedasticity) and sample size balance. We have previously shown that simultaneous violation of both assumptions leads to PERMANOVA analysis with indiscriminate rejection and type I error inflation or to significant loss of power up to inability to make any rejections at all [4]. Unfortunately, heteroscedasticity across conditions is a very common feature of microbiome data. Thus, new robust methods are needed to ensure correct data analysis.

We have previously described a T_w^2 test, which presents a robust solution for comparing two groups of microbiome samples [4]. The two-group scenario is common, but not universally satisfying as many study designs often include many different sample types, e.g., from affected and unaffected sites of a study subject and from a matched healthy control [5] and interventions as in the Cox et al. [3] study mentioned above. Here we describe a further extension of T_w^2 to allow for arbitrary number of groups with possibly different within group variability to be compared using an omnibus test for equality of means. Our method presents an advance to the state-of-the-art by introducing a way to compare data from multiple conditions where heteroscedasticity is a nuisance and only the differences between location of the data are important.

Univariate Welch MANOVA

Univariate solutions for a heteroscedastic test to compare k -means deal with finding asymptotic distributions for $\sum w_j(\bar{x}_j - \hat{\mu})^2$, as defined later in Eqs. (2) and (3). Welch’s solution [6] is perhaps the most known and well adopted in statistical literature. Next we briefly describe it, as we will build on extending this statistic to multivariate data.

Suppose we observe data from k populations $\mathbf{x}_j = (x_j^{(1)}, \dots, x_j^{(n_j)})$ with potentially unequal number of observations, n_j for $j = 1, \dots, k$, in each. Let \bar{x}_j and s_j^2 denote the means and variances for each sample. The Welch ANOVA statistic is:

$$W^* = \frac{\sum w_j(\bar{x}_j - \hat{\mu})^2 / (k - 1)}{1 + [2(k - 2) / (k^2 - 1)] \sum h_j}, \tag{1}$$

where

$$w_j = n_j / s_j^2, \tag{2}$$

$$\hat{\mu} = \sum w_j \bar{x}_j / W, \tag{3}$$

$$W = \sum w_j, \text{ and} \tag{4}$$

$$h_j = (1 - w_j / W)^2 / (n_j - 1). \tag{5}$$

The Welch test uses $F(k - 1, f)$, for $f = (k^2 - 1) / (3 / \sum h_j)$ distribution to draw inference with W^* [6].

Calculation of multivariate Welch W -statistic on distances

To derive a Welch W^* statistic suitable for analysis of microbiome data, W_d^* , we follow the same approach as we did in our derivation of T_w^2 . We first demonstrate that in the univariate case, W_d^* can be expressed in terms of sums of pairwise square differences. Next we observe that these sums represent the squares of the univariate Euclidean distances, which allows for a direct extension of the W_d^* statistic computation for multivariate Euclidean distances and in fact any arbitrary distance or dissimilarity metric. The derivation of the statistic in terms of dissimilarities makes it suitable for analysis of microbiome data via a permutation test.

We have previously shown [4] that the sample variances can be written as:

$$s_j^2 = \frac{1}{n_j(n_j - 1)} \sum_{\substack{p < q \\ p, q = 1}}^{n_j} (x_j^{(p)} - x_j^{(q)})^2 \\ = \frac{1}{n_j(n_j - 1)} \sum_{\substack{p < q \\ p, q = 1}}^{n_j} d_{pq}^{(j)2}, \tag{6}$$

where $x_j^{(p)}$ and $x_j^{(q)}$ denote p -th and q -th observations in the j -th level, $d_{pq}^{(j)}$ is the distance between them. Hence:

$$w_j = n_j / s_j^2 = (n_j - 1) n_j^2 \left(\sum_{p < q} d_{pq}^{(j)2} \right)^{-1}. \tag{7}$$

Now consider:

$$\sum_{j=1}^k w_j(\bar{x}_j - \hat{\mu})^2 = \sum_{j=1}^k w_j(\bar{x}_j - \sum_{i=1}^k w_i \bar{x}_i / W)^2 \tag{8}$$

$$= \sum_{j=1}^k \frac{w_j}{W^2} \left(W \bar{x}_j - \sum_{i=1}^k w_i \bar{x}_i \right)^2 \tag{9}$$

$$= \sum_{j=1}^k \frac{w_j}{W^2} \left(W^2 \bar{x}_j^2 - 2W \bar{x}_j \sum_{i=1}^k w_i \bar{x}_i + \left[\sum_{i=1}^k w_i \bar{x}_i \right]^2 \right) \tag{10}$$

$$= \sum_{j=1}^k w_j \bar{x}_j^2 - \frac{2}{W} \sum_{i,j=1}^k w_i w_j \bar{x}_i \bar{x}_j + \sum_{j=1}^k \frac{w_j}{W^2} \left[\sum_{i=1}^k w_i \bar{x}_i \right]^2 \quad (11)$$

$$= \sum_{j=1}^k w_j \bar{x}_j^2 - \frac{2}{W} \sum_{i,j=1}^k w_i w_j \bar{x}_i \bar{x}_j + \sum_{j=1}^k \frac{1}{W} \sum_{i,j=1}^k w_i w_j \bar{x}_i \bar{x}_j \quad (12)$$

$$= \frac{1}{2W} \left(2W \sum_{j=1}^k w_j \bar{x}_j^2 - 2 \sum_{i,j=1}^k w_i w_j \bar{x}_i \bar{x}_j \right) \quad (13)$$

$$= \frac{1}{2W} \left(\sum_{i,j=1}^k w_i w_j \bar{x}_j^2 - 2 \sum_{i,j=1}^k w_i w_j \bar{x}_i \bar{x}_j + \sum_{i,j=1}^k w_i w_j \bar{x}_i^2 \right) \quad (14)$$

$$= \frac{1}{2W} \sum_{i,j=1}^k w_i w_j (\bar{x}_i - \bar{x}_j)^2 \quad (15)$$

$$= \frac{1}{W} \sum_{i < j} w_i w_j (\bar{x}_i - \bar{x}_j)^2. \quad (16)$$

Equation (16) means that $\sum_j w_j (\bar{x}_j - \hat{\mu})^2$ can be expressed as weighted sum of squares of pairwise inter-group mean differences, which makes for a convenient expression to compute. Finally, we have previously shown that squares of mean differences can be expressed in terms of squares of pairwise sample differences [4], i.e:

$$(\bar{x}_i - \bar{x}_j)^2 = \frac{n_i + n_j}{n_i n_j} \left[\frac{1}{n_i + n_j} \sum_{\substack{i < j \\ i,j=1}}^{n_i+n_j} (z_i^{(i,j)} - z_j^{(i,j)})^2 - \left(\frac{1}{n_i} \sum_{\substack{p < q \\ p,q=1}}^{n_i} (x_i^{(p)} - x_i^{(q)})^2 + \frac{1}{n_j} \sum_{\substack{p < q \\ p,q=1}}^{n_j} (x_j^{(p)} - x_j^{(q)})^2 \right) \right], \quad (17)$$

where $\mathbf{z}^{(i,j)} = (z_1^{(i,j)}, \dots, z_{n_i+n_j}^{(i,j)}) = (x_i^{(1)}, \dots, x_i^{(n_i)}, x_j^{(1)}, \dots, x_j^{(n_j)})$. The squares of the pairwise differences under the summations in Eq. (17) can be thought of as the squares of the pairwise Euclidean distances in one dimension. This allows us to generalize the univariate Euclidean Welch ANOVA to MANOVA with arbitrary distances, where the distances can be suitably defined for the data at hand, including all of common distances used with microbiome data.

Note that in contrast to the PERMANOVA statistic, the distance-based T_w^2 and W_d^* explicitly account for potentially unbalanced number of observations and differences in multivariate spread in the two samples. Finally, observe that W_d^* reduces to T_w^2 when $k = 2$, as W^* reduces to Welch t -statistic.

As with T_w^2 , the exact distribution of the multivariate distance-based W_d^* statistic is dependent on many factors,

such as the dimensionality of underlying data, distributions of the random variables comprising the data, the exact distance metric used, and the number of groups compared k . To make a practical general test, we use permutation testing to establish the significance. To do so, we compute $W_d^*(i)$ on m permutations of the original data, for $i = 1, \dots, m$, and estimate the significance as the fraction of times the permuted statistic is greater than or equal to W_d , i.e., $p = \frac{1}{m} \sum_i^m \mathbb{1}(W_d^* \leq W_d^*(i))$. Here, $\mathbb{1}(\cdot)$ designates the indicator function. Larger p values are more easily estimated with permutations as the number of more extreme permuted statistics will be quite large. For smaller, p values often, the precise p value is not necessary, but only an indication if it is smaller than a particular threshold (e.g., 0.01). As a rule of thumb, to conclude that a p value is less than a threshold α , we recommend performing at least $5/\alpha$ permutations.

Confounder modeling and repeated measures are often key elements of microbiome study design. These can be accounted for in permutation testing procedures using restricted permutation. For example, the effect of a discrete valued confounder can be removed from the p value calculation by restricting permutations to only within the levels of the confounding variable. This amounts to an application of stratified analysis of variance. Similarly, restricting permutations to within individual subjects only results in a repeated measures analysis. Notice that the test statistic under restricted permutations remains the same, but the null distribution is changed to reflect the desired comparison. Methods for W_d^* and these restricted permutation methods are available in our reference implementation at <https://github.com/alekseyenko/WdStar>.

When multiple means are compared with W_d^* , a statistically significant result may prompt the question about attribution of the differences to a specific group or groups. Post hoc testing procedures are used to perform that kind of analysis. There are many possible ways to design the post hoc testing procedures, but the guiding principle due to potential for loss of power to multiple testing should be to minimize the number of tests performed. For this reason, in addition to all possible pairwise (one versus one) tests, it may be interesting and relevant to test one group versus all others. In this scenario, samples from one experimental group are compared to pooled samples from the remaining groups. The statistical test for this comparison can equivalently be either T_w^2 or W_d^* on two level factors. We illustrate the use of one versus all post hoc procedure in our application example in “Application example: colorectal cancer disparity and microbiome” section and provide corresponding computation routines in our reference implementation.

Empirical evaluation of W_d^* type I error

The principal evaluation that is required to assure statistical properties of W_d^* is demonstration of appropriate type I error control. For this purpose, we consider the univariate heteroscedastic case with three groups, $\{x_1^{(k_1)}\}$, $\{x_2^{(k_2)}\}$, $\{x_3^{(k_3)}\}$, $k_1 = 1 \dots, n_1$, $k_2 = 1 \dots, n_2$, and $k_3 = 1 \dots, n_3$, of samples to compare, where n_1, n_2, n_3 are the numbers of observations in each group. We let $x_1^{(k_1)} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ be the reference group and $x_2^{(k_2)} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, s^2)$ and $x_3^{(k_3)} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, s^4)$ be the groups with different variance s^2 and s^4 , respectively, to introduce heteroscedasticity. In our simulation, we let $s^2 = \{1, 0.8, 0.2\}$ to control the degree of heteroscedasticity in the range from none to large. Finally, we let the sample sizes n_1, n_2 , and n_3 take values of 5, 10, 20, or 40 to generate data with varying total sample size and degree of balance. For each combination of sample sizes and variance, we have performed 1000 simulations of the data for a total of 192,000 datasets. Each dataset has been analyzed using our reference implementation of W_d^* , PERMANOVA (adonis function in R library vegan), and univariate Welch ANOVA (oneway.test in R library stats). For distance-based methods, Euclidean distances have been used. Details of simulation are available as a knitted R Markdown file in Additional file 1. The simulation results comprise the fraction of rejected null hypotheses at $\alpha = 0.05$ by each test (Fig. 1a). A test properly controlling the type I

error is expected to have the fraction of rejections equal to the nominal error rate (0.05). Notice that the proposed W_d^* test, in fact, produces the expected error rates over the entire range of simulation parameters. Similarly to our previous observations in the two-group case, PERMANOVA is not robust to heteroscedasticity when sample size imbalance is present. Observe that whenever the number of observations in the reference group (the one with variance equal to 1) is smaller than that in the less dispersed groups, the fraction of rejections is overly inflated, resulting in higher type I error. Also notice that when there are more observations in the reference group than in others (e.g., $n_1 = 40, n_2, n_3 < 40$), it is hard for PERMANOVA to make the rejections, resulting in approximately zero type I error.

Interestingly, when we compare the raw p values obtained from W_d^* to those from the distribution based asymptotic Welch test, we see a good concordance between the two (Fig. 1b). The variability around the trendline is most likely due to Monte Carlo error associated with permutation testing and small sample size. On the contrary, when PERMANOVA is compared to the distribution-based asymptotic test, the fit is clearly much noisier (Fig. 1c). The concordance is much smaller for tests involving groups with larger degree of heteroscedasticity. The code used to produce the plots in Fig. 1 is available as Additional file 2.

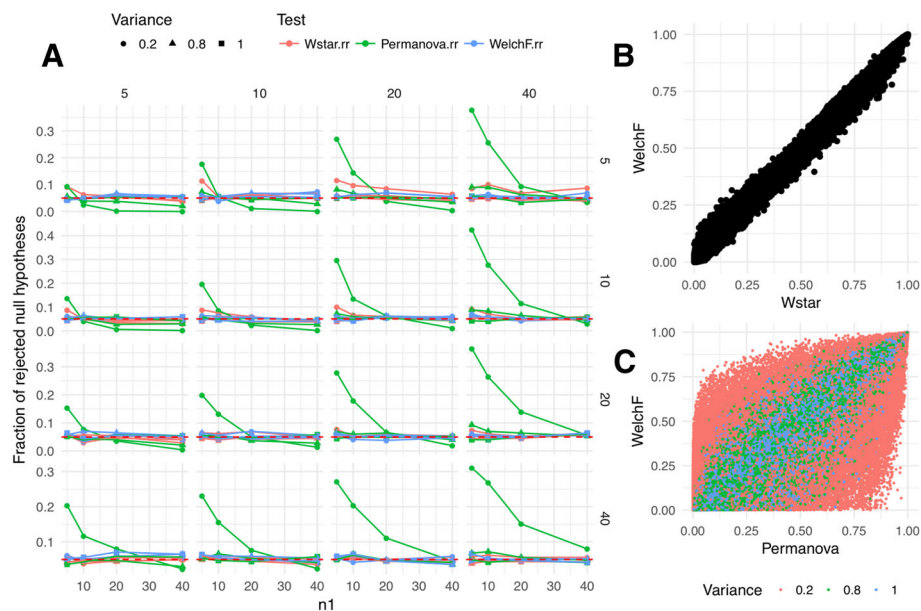


Fig. 1 Evaluation of type I errors of W_d^* and PERMANOVA permutation tests. Simulation under the null hypothesis results for comparison of W_d^* (Wstar), PERMANOVA (Permanova), and distribution-based Welch ANOVA F (WelchF) tests are presented. In panel **a**, we evaluate the fraction of null hypotheses that have been rejected by each test at $\alpha = 0.05$. The subpanels of **a** correspond to simulated datasets with corresponding number of samples in the non-reference groups, with columns corresponding to the least dispersed and rows corresponding to the most dispersed sample. In panel **b**, the raw p values from W_d^* test are plotted against those for the same data with Welch ANOVA F test. In panel **c**, we do the same for PERMANOVA p values and color the points by respective degree of heteroscedasticity in the simulated dataset

Finally, given the equivalence of the W_d^* to T_w^2 for $k = 2$, and the fact that the two-level test is powered similarly to PERMANOVA, we expect the test described in this paper to be of similar power for $k > 2$ as well. The full empirical evaluation of power characteristics for $k > 2$ is hard to achieve in non-superficial setups as most realistic simulation scenarios present an infinite universe for choice of parameters.

Application example: colorectal cancer disparity and microbiome

Extensive scientific literature suggests an important, yet not fully understood role of the intestinal microbiome in the development, progression, and treatment of colorectal cancers (CRC). Several genus level bacterial taxa have been associated with CRC [7] but the role of personal characteristics in influencing the presence of CRC-associated bacteria is not well understood. A few studies have noted marked differences in the microbial environment in the gut of African-Americans (AA) versus others [7–11] (e.g., Caucasian (CA)) and suggested differences in microbial composition among those with and without colorectal polyps and cancer. Others found distinct differences in the microbes populating the proximal and distal colo-rectum [12, 13]. Lower socioeconomic status and western diet have been associated with a lower microbial diversity, especially in the distal colon [14, 15]. Microbial signature approaches have been used for development of diagnostic biomarkers [9, 16–18] or assessing differences in immune gene expression [13]—highlighting the increasing importance of statistical methods to analyze clusters of microbes-genes while also taking into account patient-level variables. The role of the gut microbiome in CRC disparities is likewise poorly understood [19]. Here we use a pilot CRC dataset to demonstrate the utility of W_d^* in uncovering signals potentially missed due to heteroscedasticity.

The Medical University of South Carolina (MUSC) Institutional Review Board approved all study activities. The Cancer Registry at Hollings Cancer Center (HCC) at MUSC was used to identify all cases of CRC. The study population was comprised of a sample of histologically confirmed cases diagnosed between January 1, 2000, and June 30, 2015. Patients were of either AA or CA descent. For each case, we obtained a formalin-fixed, paraffin-embedded tissue blocks from the MUSC Department of Pathology and Laboratory Medicine. DNA was extracted following standard protocols in the laboratory. Briefly, the colonic tissue was transferred to a tube containing lysis buffer (1% SDS, 1 mg/ml Proteinase K, LTE pH 8.0). The solution was incubated at 50°C for 1 h, followed by phenol/chloroform extraction and ethanol precipitation. The quantity and quality of DNA was then determined by running a small aliquot

on a 1% agarose gel and comparing it to a set of DNA standards. The extracted DNA was stored at -80°C . V3 and V4 regions of the 16S rRNA gene have been amplified using 16S Amplicon PCR Forward Primer = 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGC CTACGGGNGGCWGCAG 16S Amplicon PCR Reverse Primer = 5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGACTACHVGGGTATCTAATCC using KAPA HiFi enzyme. The library has been prepared using Nextera XT index kits and sequenced using MiSeq Reagent Kit v3 in a MiSeq instrument. We have analyzed the genera previously reported in a systematic review to be associated with CRC [20]. Jensen-Shannon Divergence distances have been computed between the subjects of Caucasian and African-American races with cancers in distal and proximal locations of their colons. See Additional file 4 for the list of 14 genera retained for this analysis.

We selected a convenience sample from our MUSC cancer cohort of 20 patients (10 AAs, 10 CAs) which we matched on colonic location (proximal, distal) and sex. Of the 20 cases, 6 have been removed due to low sequence count (< 100) within the genera of interest. This resulted in groups with highly unbalanced number of subjects (Table 1). The location of the cancers correlate with outcomes, whereby distal results in the worst. Hence, to examine possible microbial clues into racial disparity the interaction of race and location is important. Due to extremely small pilot-scale sample size, the group unbalance and potential for heteroscedasticity prompt caution with using PERMANOVA for these comparisons (Fig. 2). Visually, we observe the two distal AA observations segregating to the extreme left of the other specimens. On the other hand the centroid of the CA distal observations is in the diagonal quadrant from the centroid AA distal observations. The proximal specimens are not separated on the first principal axis, but a moderate segregation can be observed along the second axis. Note that if primary effects only are considered in this case, the differences in the location of the cancers would not be plausibly different, as the corresponding centroids would lie well within each others' circles of inertia. These qualitative observations are suggestive that the interaction of race and location represented by the four groups of observations harbor differential microbiota, albeit the extremely small sample size

Table 1 Number of the subjects in the colorectal cancer example analysis

Race	Cancer location	<i>N</i>
African-American	Distal	2
	Proximal	3
Caucasian	Distal	5
	Proximal	4

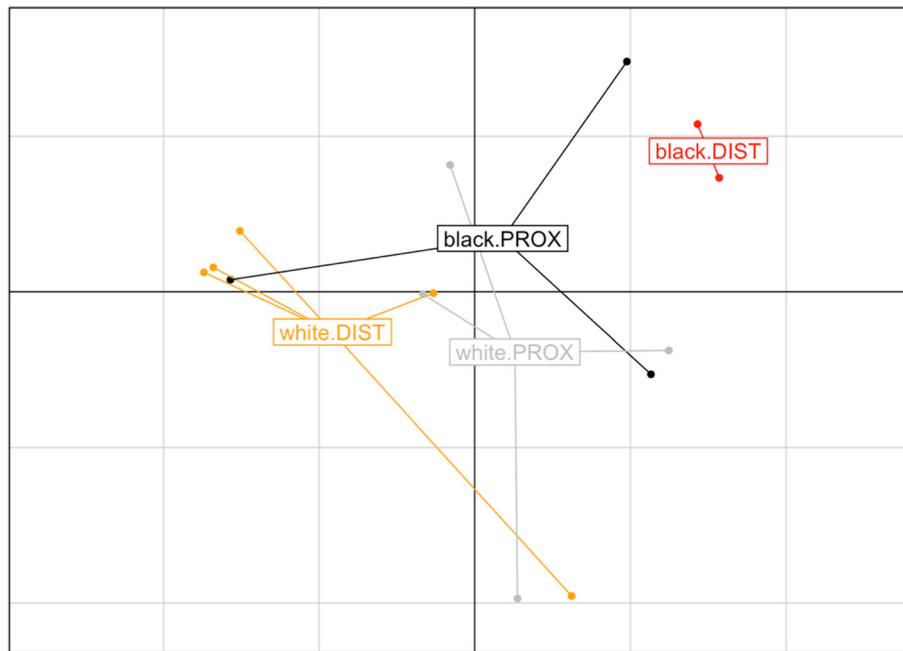


Fig. 2 PCoA plot of the JSD distances between CRC microbiome samples. African-American distal (red) samples appear to be separated on PC1 from the samples in the proximal AA (black) and Caucasian (gray) and Caucasian distal (orange) samples. Likewise, the plot suggest that the multivariate spread may differ dramatically in the compared groups with AA distal samples being most concentrated relative to the other groups

may not be sufficient to deem the observed effect sizes significant.

Nonetheless, the race and location interaction model achieves significance ($p < 0.05$) with W_d^* test (Table 2). Observe that as expected the difference between the distal and proximal cancers alone is not significant, but in combination with race suggest existence of differential interaction. These nuances are not captured by the analyses with PERMANOVA, which yields unremarkable p values. Interestingly, there is a discrepancy in test results for the primary effect of the race at 0.05 significance threshold, which is well within the gray zone of being notable. However, the discrepancy of PERMANOVA and W_d^* in the interaction term is a clear illustration for utility of our method. In the presence of heteroscedasticity and sample size imbalance, one might doubt the result by PERMANOVA. Next, we demonstrate the application of the post hoc procedures described in the methods section. Significance of the interaction term may dictate additional questions about which groups differ from

the rest. We demonstrate the use of one versus all post hoc testing by comparing each group with the rest of the samples (Table 3). As expected, these indicate a significant difference ($p < 0.05$) in the microbiome of the AA distal CRC samples from the rest, and a trend for difference of the Caucasian distal samples. Note that the interpretations of these results might differ if multiple comparison issues are taken into account. Due to the pilot nature of these data, we do not perform any formal corrections, as our goal is to determine the plausibility of significant differences, which are to be evaluated in appropriately sized datasets where power is not a concern.

Epidemiological literature indicates that AA and CA have notable differences in the prevalence of colorectal neoplasia in the proximal and distal colorectum at both the precancerous [21–24] and invasive stages [25]. Numerous lifestyle and dietary factors associated with dysbiosis (e.g., red-meat intake, sedentary

Table 2 Significance of the primary and interaction effects by PERMANOVA and W_d^* tests

Covariate	PERMANOVA p value	W_d^* p value
Race	0.064	0.047
Location	0.907	0.908
Race and location	0.282	0.037

Table 3 One versus all post hoc comparisons of the interaction terms

Group	T_w^2 statistic	W_d^* p value
AA distal	8.88	0.039
CA distal	1.93	0.075
AA proximal	0.36	0.936
CA proximal	0.70	0.665

lifestyle, heavy alcohol use, western diet) are strongly associated with the risk of distal colorectal cancer [26–30]. A recent study reported that blacks compared to whites had a greater abundance of sulfidogenic bacteria in the normal colonic mucosa which correlated with higher intakes of fat, protein, and meat per day [31]. Overall, the racial differences we observed in microbial patterns in the CRCs by colonic location may reflect differences in modifiable lifestyle and dietary factors.

The data and R Markdown for this application is included in Additional files 3 and 4.

Discussion and conclusion

Community-wide analyses where the entire microbiome is modelled as a response variable of one or more factors has become a standard first line of analysis technique in the field. These techniques address the question of overall aggregate changes in the microbiome in response to explanatory variables without the need to model each individual microbiome constituent. PERMANOVA [1] has been one of the most dominant tools for such analyses, although the potential for confounding of location and dispersion effects has been recognized for a long time [32, 33]. The W_d^* method closes the gap by explicitly accounting for the differences in multivariate dispersion in the data tested, which has been shown to be associated with adverse statistical properties in PERMANOVA [4]. Current heteroscedasticity-aware methodologies allow for modeling multi-level factors, stratification, and multiple post hoc testing scenarios. Although in many applications the differences in statistical decisions made on the basis of PERMANOVA and W_d^* may remain unchanged, the principled guarantees of being correct in wider range of scenarios provided by the latter might be important for practitioners. Although originally developed for discrete-valued covariates, PERMANOVA remains a viable analysis option for continuous covariates as well when multivariate regression-like formula are utilized [34]. However, the effect of heteroscedasticity has not been rigorously evaluated or addressed for such analyses. To be fair, heteroscedasticity with continuous covariates is an issue that does not have a generic statistical solution applicable in most cases. A more cautious analysis involving continuous covariates may require corroboration with discretized independent variables by W_d^* , but has to also account for potential statistical power issues pertaining to discretization.

A major limitation of most community-wide analyses is that those often do not yield a natural unified framework for evaluation of taxon-level effects. Currently, methods that have this unifying ability are emerging [35]. None of these, however, are evaluated for robustness with heteroscedastic data yet.

Additional files

Additional file 1: Knitted HTML R Markdown document detailing the steps of producing the simulation datasets and running each test to evaluate the Type I error performance of W_d^* relative to PERMANOVA and asymptotic Welch F test. (HTML 17 kb)

Additional file 2: Knitted HTML R Markdown document containing the code used to produce Fig. 1. (HTML 2974 kb)

Additional file 3: R Data file containing the R package phyloseq object with data for the application example. The object includes the genus level abundance tables, sample data containing designations of the race and CRC location, and taxonomic table for the data. (RData 8 kb)

Additional file 4: Knitted HTML R Markdown document detailing application example analyses. (HTML 842 kb)

Acknowledgements

The authors would like to thank ZhengZheng Tang for early input in this work.

Funding

AVA and BH are supported by NIH/NLM R01 LM12517, AVA and KW are supported by Medical University of South Carolina College of Medicine Enhancing Team Science Award. AVA is supported by NIH/NCI U54 CA210962. The project described was supported by the NIH/NCATS UL1 TR001450.

Availability of data and materials

All data, software and other materials are available at <https://github.com/alekseyenko/WdStar>.

Authors' contributions

AVA has conceived the method, derived the test statistic, developed reference implementation in R statistical programming language, wrote the manuscript, and performed the data analysis; BH has implemented code for restricted permutations; KW has designed original study on CRC and collected and organized tissue and DNA samples; CV has generated 16S rRNA gene sequencing data. All authors have reviewed and approved the final manuscript.

Ethics approval and consent to participate

The human subjects component of this research has been approved by the Medical University of South Carolina (MUSC) Institutional Review Board.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Program for Human Microbiome Research, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA. ²Biomedical Informatics Center, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA. ³Department of Public Health Science, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA. ⁴Department of Microbiology and Immunology, Medical University of South Carolina, 173 Ashley Avenue MSC 509, 29425 Charleston, SC, USA. ⁵Department of Oral Health Sciences, Medical University of South Carolina, 135 Cannon Street MSC 200, 29425 Charleston, SC, USA.

Received: 28 August 2018 Accepted: 11 March 2019

Published online: 01 April 2019

References

- Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001;26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
- Redel H, Gao Z, Li H, Alekseyenko AV, Zhou Y, Perez-Perez GI, Weinstock G, Sodergren E, Blaser MJ. Quantitation and composition of

- cutaneous microbiota in diabetic and nondiabetic men. *J Infect Dis*. 2013;207(7):1105–14. <https://doi.org/10.1093/infdis/jit005>.
3. Cox LM, Cho I, Young SA, Anderson WH, Waters BJ, Hung SC, Gao Z, Mahana D, Bihan M, Alekseyenko AV, Methe BA, Blaser MJ. The nonfermentable dietary fiber hydroxypropyl methylcellulose modulates intestinal microbiota. *FASEB J*. 2013;27(2):692–702. <https://doi.org/10.1096/fj.12-219477>.
 4. Alekseyenko AV. Multivariate welch t-test on distances. *Bioinformatics*. 2016;32(23):3552–8. <https://doi.org/10.1093/bioinformatics/btw524>.
 5. Alekseyenko AV, Perez-Perez GI, De Souza A, Strober B, Gao Z, Bihan M, Li K, Methé BA, Blaser MJ. Community differentiation of the cutaneous microbiota in psoriasis. *Microbiome*. 2013;1(1):31. <https://doi.org/10.1186/2049-2618-1-31>.
 6. Welch BL. On the comparison of several mean values: An alternative approach. *Biometrika*. 1951;38(3-4):330–6. <https://doi.org/10.1093/biomet/38.3-4.330>.
 7. Yazici C, Wolf PG, Kim H, Cross TL, Vermillion K, Carroll T, Augustus GJ, Mutlu E, Tussing-Humphreys L, Braunschweig C, Xicola RM, Jung B, Llor X, Ellis NA, Gaskins HR. Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut*. 2017;66(11):1983–94. <https://doi.org/10.1136/gutjnl-2016-313321>.
 8. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K, Gaskins HR, O'Keefe SJ. Diet, microbiota, and microbial metabolites in colon cancer risk in rural africans and african americans. *Am J Clin Nutr*. 2013;98(1):111–20. <https://doi.org/10.3945/ajcn.112.056689>.
 9. Brim H, Yooseph S, Lee E, Sherif ZA, Abbas M, Laiyemo AO, Varma S, Torralba M, Dowd SE, Nelson KE, Pathmasiri W, Sumner S, de Vos W, Liang Q, Yu J, Zoetendal E, Ashktorab H. A microbiomic analysis in african americans with colonic lesions reveals streptococcus sp.vt162 as a marker of neoplastic transformation. *Genes (Basel)*. 2017;8(11). <https://doi.org/10.3390/genes8110314>.
 10. O'Keefe SJ, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, Posma JM, Kinross J, Wahl E, Ruder E, Vippera K, Naidoo V, Mtshali L, Tims S, Puylaert PG, DeLany J, Krasinskas A, Benefiel AC, Kaseb HO, Newton K, Nicholson JK, de Vos WM, Gaskins HR, Zoetendal EG. Fat, fibre and cancer risk in african americans and rural africans. *Nat Commun*. 2015;6:6342. <https://doi.org/10.1038/ncomms7342>.
 11. Bridges KM, Diaz FJ, Wang Z, Ahmed I, Sullivan DK, Umar S, Buckles DC, Greiner KA, Hester CM. Relating stool microbial metabolite levels, inflammatory markers and dietary behaviors to screening colonoscopy findings in a racially/ethnically diverse patient population. *Genes (Basel)*. 2018;9(3). <https://doi.org/10.3390/genes9030119>.
 12. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, Peterson SN, Snesrud EC, Borisov GG, Lazarev M, Stein E, Vadivelu J, Roslani AC, Malik AA, Wanyiri JW, Goh KL, Thevambiga I, Fu K, Wan F, Llosa N, Housseau F, Romans K, Wu X, McAllister FM, Wu S, Vogelstein B, Kinzler KW, Pardoll DM, Sears CL. Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci U S A*. 2014;111(51):18321–6. <https://doi.org/10.1073/pnas.1406199111>.
 13. Flemer B, Herlihy M, O'Riordain M, Shanahan F, O'Toole PW. Tumour-associated and non-tumour-associated microbiota: Addendum. *Gut Microbes*. 2018;1–5. <https://doi.org/10.1080/19490976.2018.1435246>.
 14. Miller GE, Engen PA, Gillevet PM, Shaikh M, Sikaroodi M, Forsyth CB, Mutlu E, Keshavarzian A. Lower neighborhood socioeconomic status associated with reduced diversity of the colonic microbiota in healthy adults. *PLoS One*. 2016;11(2):0148952. <https://doi.org/10.1371/journal.pone.0148952>.
 15. Zinocker MK, Lindseth IA. The western diet-microbiome-host interaction and its role in metabolic disease. *Nutrients*. 2018;10(3). <https://doi.org/10.3390/nu10030365>.
 16. Liang Q, Chiu J, Chen Y, Huang Y, Higashimori A, Fang J, Brim H, Ashktorab H, Ng SC, Ng SSM, Zheng S, Chan FKL, Sung JY, Yu J. Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. *Clin Cancer Res*. 2017;23(8):2061–70. <https://doi.org/10.1158/1078-0432.Ccr-16-1599>.
 17. Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res (Phila)*. 2014;7(11):1112–21. <https://doi.org/10.1158/1940-6207.Capr-14-0129>.
 18. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Bohm J, Brunetti F, Habermann N, Herczeg R, Koch M, Luciani A, Mende DR, Schneider MA, Schrotz-King P, Tournigand C, Tran Van Nhieu J, Yamada T, Zimmermann J, Benes V, Kloor M, Ulrich CM, von Knebel Doeberitz M, Sobhani I, Bork P. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. 2014;10:766. <https://doi.org/10.15252/msb.20145645>.
 19. Wallace K, Lewin D, Sun S, Spiceland C, Rockey D, Alekseyenko A, Wu J, Baron J, Alberg A, Hill E. Tumor-infiltrating lymphocytes and colorectal cancer survival in african american and caucasian patients. *Cancer Epidemiol Biomark Prev*. 2018;27(7):755–61. <https://doi.org/10.1158/1055-9965.EPI-17-0870>.
 20. Borges-Canha M, Portela-Cidade JP, Dinis-Ribeiro M, Leite-Moreira AF, Pimentel-Nunes P. Role of colonic microbiota in colorectal carcinogenesis: A systematic review. *Revista Española de Enfermedades Digestivas*. 2015;107(11):659–71. <https://doi.org/10.17235/reed.2015.3830/2015>.
 21. Corley DA, Jensen CD, Marks AR, Zhao WK, de Boer J, Levin TR, Doubeni C, Fireman BH, Quesenberry CP. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: implications for screening and quality programs. *Clin Gastroenterol Hepatol*. 2013;11(2):172–80. <https://doi.org/10.1016/j.cgh.2012.09.010>.
 22. Friedenberg FK, Singh M, George NS, Sankineni A, Shah S. Prevalence and distribution of adenomas in black americans undergoing colorectal cancer screening. *Dig Dis Sci*. 2012;57(2):489–95. <https://doi.org/10.1007/s10620-011-1952-z>.
 23. Lebwahl B, Capiak K, Neugut AI, Kastrinos F. Risk of colorectal adenomas and advanced neoplasia in hispanic, black and white patients undergoing screening colonoscopy. *Aliment Pharmacol Ther*. 2012;35(12):1467–73. <https://doi.org/10.1111/j.1365-2036.2012.05119.x>.
 24. Lieberman DA, Williams JL, Holub JL, Morris CD, Logan JR, Eisen GM, Carney P. Race, ethnicity, and sex affect risk for polyps >9 mm in average-risk individuals. *Gastroenterology*. 2014;147(2):351–8145. <https://doi.org/10.1053/j.gastro.2014.04.037>.
 25. Xicola RM, Gagnon M, Clark JR, Carroll T, Gao W, Fernandez C, Mijic D, Rawson JB, Janoski A, Pusatcioglu CK, Rajaram P, Gluskin AB, Regan M, Chaudhry V, Abcarian H, Blumetti J, Cintron J, Melson J, Xie H, Guzman G, Emmadi R, Alagiozian-Angelova V, Kupfer SS, Braunschweig C, Ellis NA, Llor X. Excess of proximal microsatellite-stable colorectal cancer in african americans from a multiethnic study. *Clin Cancer Res*. 2014;20(18):4962–70. <https://doi.org/10.1158/1078-0432.CCR-14-0353>.
 26. Cong YJ, Gan Y, Sun HL, Deng J, Cao SY, Xu X, Lu ZX. Association of sedentary behaviour with colon and rectal cancer: a meta-analysis of observational studies. *Br J Cancer*. 2014;110(3):817–26. <https://doi.org/10.1038/bjc.2013.709>.
 27. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, Negri E, Straif K, Romieu I, La Vecchia C, Boffetta P, Jenab M. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol*. 2011;22(9):1958–72. <https://doi.org/10.1093/annonc/mdq653>.
 28. Kunzmann AT, Coleman HG, Huang WY, Kitahara CM, Cantwell MM, Berndt SI. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Clin Nutr*. 2015;102(4):881–90. <https://doi.org/10.3945/ajcn.115.113282>.
 29. Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer*. 2009;124(10):2406–15. <https://doi.org/10.1002/ijc.24191>.
 30. Mehta RS, Song M, Nishihara R, Drew DA, Wu K, Qian ZR, Fung TT, Hamada T, Masugi Y, da Silva A, Shi Y, Li W, Gu M, Willett WC, Fuchs CS, Giovannucci EL, Ogino S, Chan AT. Dietary patterns and risk of colorectal cancer: Analysis by tumor location and molecular subtypes. *Gastroenterology*. 2017;152(8):1944–19531. <https://doi.org/10.1053/j.gastro.2017.02.015>.
 31. Yazici C, Wolf PG, Kim H, Cross TL, Vermillion K, Carroll T, Augustus GJ, Mutlu E, Tussing-Humphreys L, Braunschweig C, Xicola RM, Jung B, Llor X, Ellis NA, Gaskins HR. Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut*. 2017;66(11):1983–94. <https://doi.org/10.1136/gutjnl-2016-313321>.
 32. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*. 2006;62(1):245–53. <https://doi.org/10.1111/j.1541-0420.2005.00440.x>.
 33. Warton DI, Wright ST, Wang Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol Evol*. 2012;3(1):89–101. <https://doi.org/10.1111/j.2041-210X.2011.00127.x>.

34. Zapala MA, Schork NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc Natl Acad Sci U S A*. 2006;103(51):19430–5. <https://doi.org/10.1073/pnas.0609333103>.
35. Satten GA, Tyx RE, Rivera AJ, Stanfill S. Restoring the duality between principal components of a distance matrix and linear combinations of predictors, with application to studies of the microbiome. *PLoS One*. 2017;12(1):0168131. <https://doi.org/10.1371/journal.pone.0168131>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

