BIORANKINGS®
TECHNICAL REPORT SERIES

Repeated Measures Method for Microbial Count Data

Microbiome research is moving to routinely take samples at multiple times in the same individuals making this *longitudinal* or *repeated measures* data. Repeated measures is an attractive experimental design model since it increases the power of detecting group differences, and allows one to detect changes over time. Currently, no formal statistical hypothesis test for microbiome data repeated measures is available even though a ***validated parametric test is likely going to be needed to get a product though the FDA pipeline***.

This technical report introduces a repeated measures analysis method for the microbiome data using the generalized Dirichlet-multinomial model. We start by reviewing the concept of compositional data, explain the challenge of the repeated data analysis, present the method, and illustrate its performance in hypothesis testing using simulated data.

Compositional Data and Dirichlet-Multinomial Distribution

In microbiome studies, subjects' samples are sequenced and microbial communities are made available for analysis as counts of taxa per sample. The data is compositional in its nature. That is, the amount of taxa A is not independent of the amount of taxa B. Once converted to proportions, the total for all taxa adds to 100%. For example, if all but one taxa equal 95%, then the remaining taxon must be 5% of the total. If the abundance of one taxa increases, the abundance of one or more other taxa will decrease.

A standard method of multivariate analysis for analyzing count data is the Multinomial model. However, microbiome data has higher between subject variation (overdispersion) than is expected in the Multinomial model. The between subject variation refers to the difference in microbial composition between different subjects in a sample. Overdispersion is common in ecological data. In the presence of overdispersion, the Dirichlet-Multinomial distribution is used. In 2012, the application of DM distribution to the microbiome data with formal tests hypotheses comparing microbiome in different groups (e.g., healthy/sick) was developed (La Rosa et al., 2012). The method has since been successfully applied to a variety of microbiome data sets (La Rosa et al. 2014, Warner et al. 2016, Blount et al. 2017). Two statistics define the distribution – average proportions of taxa across all subjects (like mean in classical statistics) and overdispersion (a measure of between samples variation, similar to variance in classical statistics).

We have recently extended this model to repeated measures microbiome data.

Repeated Measures Concept and Challenges

Once data is collected at different times on the same subject, the data becomes more complicated to analyze. First, the data within subjects is not independent (*i.i.d.*, Shannon, 2017b) making classical tests that assume independence (e.g., t-tests or analysis-of-variance) incorrect. Second, the analysis needs to consider both between-subject differences and within-subject differences (Anderson, 1984; Timm, 2002) or the statistical results will be wrong.

The within-subject differences are important to understand, and refers to the fact that a person's microbiome composition at any time point is correlated with his/her microbiome at other times (if someone had a lot of taxa A yesterday, they are likely to have a lot today). While some changes will occur within an individual, the tendency to have similar data must be considered.

In classical statistics where the outcome is a number, methods for repeated measures have been well-developed and in use for a long time such as mixed models (Holden et al 2008). However, these methods do not work with longitudinal microbiome compositional data requiring a new model to be developed.

Repeated Measures for Microbiome Data

In microbiome repeated measures data, each time point can be viewed as a table of taxa counts for subjects in each group. Figure 1 shows this for a hypothetical dataset where samples have been collected at 3 time points. The rows of each barplot are the subjects, and the color bars are taxa

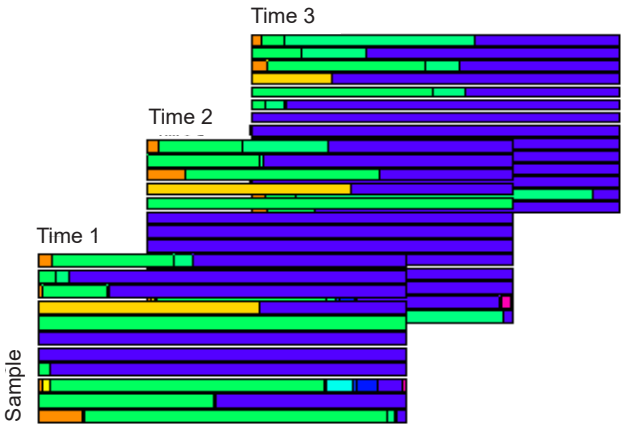

Figure 1

proportions. Note the colors across the time points always represent the same taxa. The figure shows how individual taxa change in subjects indicated by the changing lengths of the colored bars.

Here we describe the **repeatDM** algorithm for fitting microbiome repeated data. **repeatDM** is a test of hypothesis comparing two groups where the microbiome is measured at multiple times in the subjects. It is based on the generalized Dirichlet-Multinomial model (Wilson and Chen, 2007) that does two things necessary for this analysis. First, it accounts for the overdispersion. Second, it accounts for the non-independence of samples from the same subject. In a simple example, suppose patients are randomized to placebo or treatment, and their microbiomes measured multiple times during the study. **repeatDM** allows us to test and calculate a P value for the null hypothesis that the microbiome is the same in the two groups at all time points (i.e., treatment does nothing to change the microbiome composition).

Properties of a Valid Statistical Test

Statistical tests are powerful because we can know the Type I and II error, can calculate sample sizes and power based on the test and the hypothesis, and because they automate the decision making (e.g., P < 0.05 means the groups are significantly different). In the following sections we test **repeatDM** using simulated data to make sure it produces the right results.

When there is no difference between groups (the null hypothesis), the p-value is known to fall anywhere between 0 and 1 at random. If you run the experiment over and over under the null, and plot the P values, they would show a uniform distribution. This is tested in Experiment 1.

When there are differences between groups (i.e., the alternative hypothesis), P < 0.05 will occur at the same rate as the power. If power was set at 80%, then 80% of the P values would be less than 0.05, and 20% above 0.05. This is tested in Experiment 2.

Determining the power and sample size for microbiome studies is discussed in detail for biologists in chpt 6 of Metagenomics for Microbiology.

Simulations

The data was simulated as follows. Consider two groups of subjects, say Treatment and Placebo, measured at 3 time points. Time 1 is baseline, with no difference in the microbiome between the two groups. In Experiment 1 (null), the microbiome stays the same in the two post-baseline timepoints. This is shown in the top 3 plots of **Figure 2** where

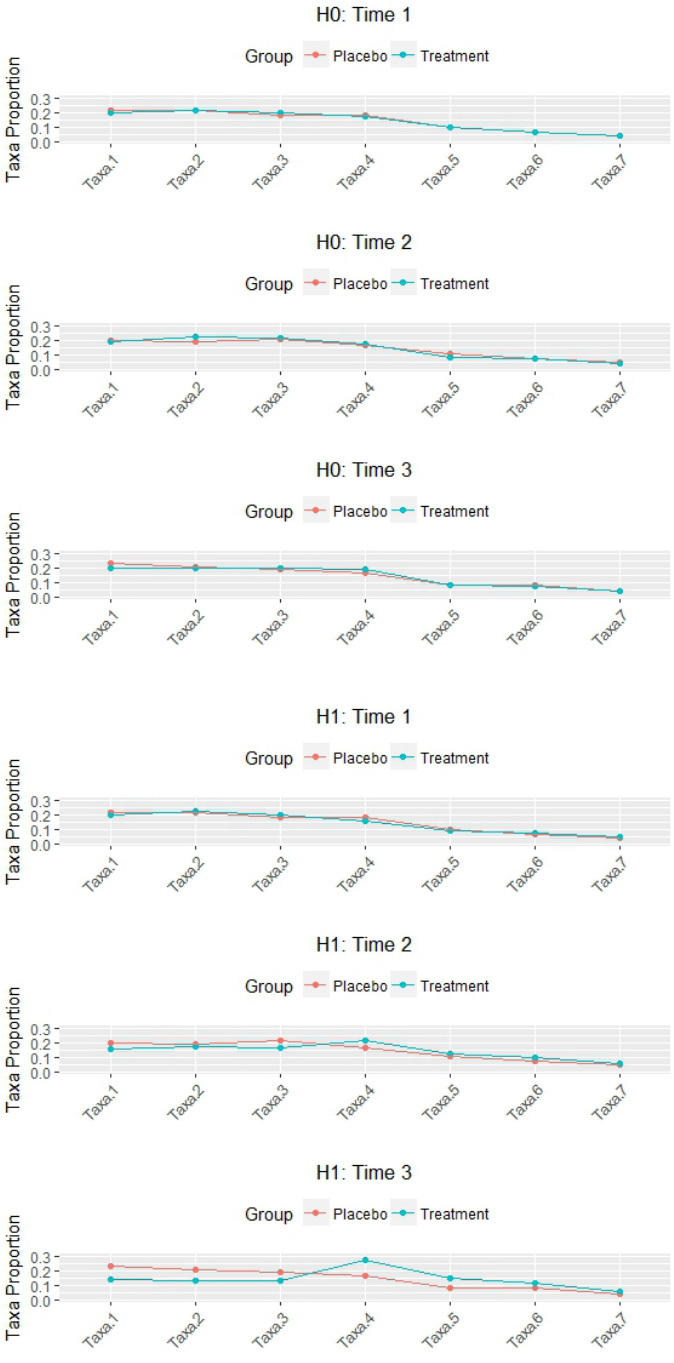

Figure 2

the X axis are the taxa, the Y axis are the proportions of each taxa, the two lines are the groups, and the three plots in the column are the time points. In Experiment 2 (alternative hypothesis), the microbiome changes in the Treatment group (blue line), shown in the bottom three plots of **Figure 2**.

Experiment 1: No Group Difference

The simulation was repeated 10,000 times with three sample sizes of 20, 60, and 100 subjects per group, and the p-value from **repeatDM** recorded. As expected, the P values showed a uniform distribution between 0 and 1 for all three sample sizes (**Figure 3**).

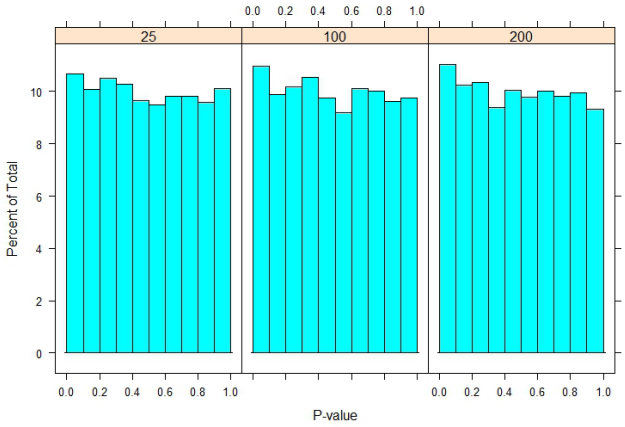

Figure 3

Experiment 2: Group Difference

Under the alternative hypothesis, there is a difference between the two groups. The amount of difference between groups is called the **effect size**, and is defined for microbiome studies in both *La Rosa* and *Izard*. The larger the effect size, the fewer number of samples needed to find a statistical difference.

In Experiment 2 the effect size between the two groups at time 2 and 3 was set at small, medium, and large. Three sample sizes were tested at N = 25, 100, and 200. The process was repeated 10,000 times and the **repeatDM** P value recorded. Power is defined as the proportion of P values < 0.05. **Table 1** shows the power behaving as expected, with power increasing with both an increase in N and/or an increase in effect size.

	Effect size (group difference)		
	Small	Medium	Large
25	51.4%	74.0%	90.8%
100	80.5%	94.3%	99.3%
200	87.2%	97.3%	99.8%

Table 1

Conclusion

In this Technical Report we present a new fully parametric method of analyzing repeated measures microbiome data which is called **repeatDM**. While the theoretical development shows the statistical test is correct (manuscript to appear), this Technical Report shows two simulations which confirm that the test is behaving as expected under the null and alternative hypotheses.

A second Technical Report is being written in collaboration with Rebiotix Inc. where **repeatDM** is applied to real data for illustration.

For more information and software please contact bill@biorankings.com

BioRankings' Technical Report Series

BioRankings' mission is to help biomedical researchers move their technology from the lab to clinical applications using statistically valid analytical tools for efficient study designs, correct data analyses and conclusions, and rigorous and objective decision making for designing follow-up studies and eventual FDA approval.

To help achieve its mission, BioRankings publishes a Technical Report series focused on applying various statistical methods to real data analyses. Written for understanding by scientists and administrators, these reports will provide an intuitive understanding of the analyses leaving the statistical details to other publications.

For more information, contact BioRankings at 314-704-8725 or bill@biorankings.com.

References

Anderson, T. W. (1984). An introduction to multivariate statistical analysis (2nd ed.). New York: Wiley.

Blount et al. 2017 "Changing the Microbiome: Patients with a Successful Outcome Following Microbiota-Based RBX2660 Treatment Trend Toward Human Microbiome Project Healthy Subjects' Profile"

Holden et al. (2008). "Analyzing Change: A Primer on Multilevel Models with Applications to Nephrology", American Journal of Nephrology

Izzard et al. (2014) Metagenomics for Microbiology http://www.sciencedirect.com/science/book/9780124104723

La Rosa et al. (2012) "Hypothesis Testing and Power Calculations for Taxonomic -Based Human Microbiome Data" PlosOne

La Rosa et al. (2014) "Patterned progression of bacterial populations in the premature infant gut" La Rosa et al. Proc Nat Acad Scie

Shannon, W. (2017b). Finding Distinct Subgroups of Samples Using Microbiome Taxa Count Data. Retrieved from http://biorankings.com/Finite_Mixture_Models_DM.pdf

Timm, N. H. (2002). Applied multivariate analysis. New York: Springer.McLachlan, G. J., & Peel, D. (2000). Finite mixture models. New York: Wiley.

Warner et al (2016) "Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: a prospective case-control study" Warner et al. Lancet

Wilson and Chen (2007). "Dirichlet-multinomial Model with Varying Response Rates over Time" Journal of Data science