BIO782P Statistics for Bioinformaticians: Assessment 2

Dataset 3

There was a significant linear relationship between HIV viral load and genetic distance (LM, F(1, 38) = 19.139, p < 0.001). There was a weak positive correlation between the two variables (adjusted R-squared = 0.318). For every unit increase in mean genetic distance, we observed a 1.17 unit increase in viral load.

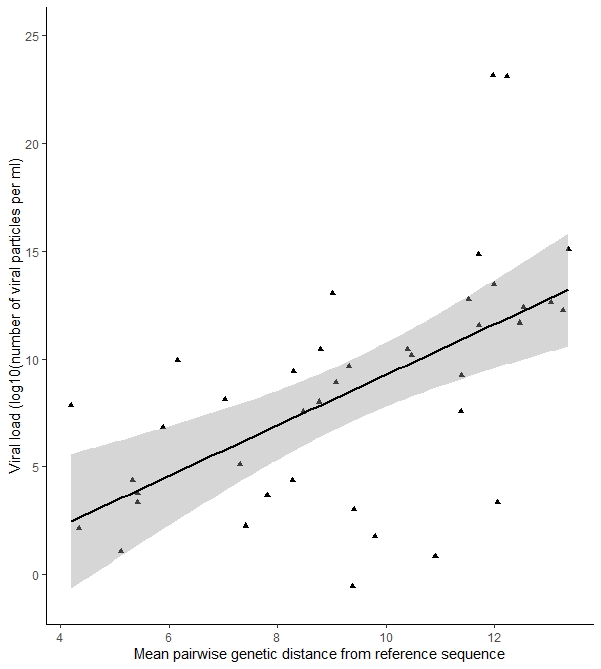
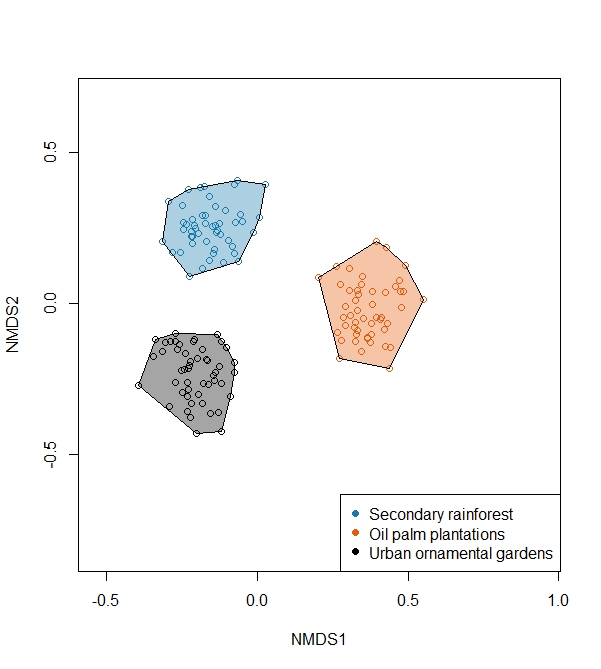


Figure 1. Scatter plot showing the linear relationship between genetic distance and viral load in an individual over 40 weeks. Viral load = -2.433 + 1.170 mean genetic distance. 95% confidence intervals shown in grey.

As the diagnostic plot of residuals vs fitted values did not appear to the naked eye to be homoscedastic, the model was tested for heteroscedasticity (along with linearity, skewness, kurtosis and link function) using the gvlma R package (Pena and Slate, 2019), and all assumptions were satisfied (all *p*-values > 0.112).

We attempted to fit a model to explain viral load that additionally incorporated the explanatory variables CD4 count (as a factor, with levels low and high) and polynomial-transformed Shannon diversity (degree = 3), with two interaction terms, CD4 count:Shannon diversity, and CD4 count:genetic distance, but none of these terms were significant (*p*-values all > 0.11).

Dataset 4

Each land use type differed significantly from each other in terms of insect OTU composition of bat faeces (PerMANOVA: F = 62.737 *p* = 0.002). 46% of the variance in the dissimilarity matrix was attributed to land use type. However, a permutest gave a significant result (p = 0.001) suggesting that one assumption of the PERMANOVA test, homogeneity of variance between groups, was violated.  
Figure 2. Dietary insect composition by land use type. Each small empty circle represents a faeces sample from one bat. NMDS stress = 0.097. PerMANOVA: F = 62.737 *p* = 0.002.

For each pairwise comparison between land use types, 14 of the 25 insect OTUs were deemed “influential” by virtue of the size of their contribution to the Bray-Curtis measures of diversity (table 1). OTUs X5, X9, X10, X15, X19, X22 and X25 were influential in all three pairwise comparisons, and OTUs X2, X7, X8, X11 and X23 weren’t influential in any.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Individual contribution to differences (%) | Cumulative contribution (%) | Mean relative abundance (%) |  |  | Individual contribution to differences (%) | Cumulative contribution (%) | Mean relative abundance (%) |
| X1 | 7.49 | 7.49 | 3.37 |  | X9 | 9.62 | 9.62 | 6.63 |
| X9 | 6.59 | 14.08 | 6.63 |  | X1 | 7.06 | 16.68 | 3.37 |
| X22 | 6.13 | 20.21 | 7.11 |  | X22 | 6.02 | 22.70 | 7.11 |
| X3 | 5.67 | 25.88 | 2.13 |  | X4 | 5.36 | 28.06 | 4.12 |
| X6 | 5.27 | 31.15 | 1.58 |  | X16 | 4.99 | 33.05 | 5.93 |
| X10 | 5.08 | 36.23 | 4.22 |  | X5 | 4.91 | 37.95 | 5.46 |
| X15 | 5.02 | 41.26 | 5.13 |  | X24 | 4.55 | 42.51 | 3.79 |
| X24 | 4.92 | 46.18 | 3.79 |  | X18 | 4.55 | 47.05 | 4.77 |
| X25 | 4.89 | 51.07 | 6.38 |  | X20 | 4.40 | 51.45 | 2.16 |
| X17 | 4.20 | 55.27 | 2.97 |  | X19 | 4.11 | 55.55 | 2.96 |
| X5 | 3.96 | 59.23 | 5.46 |  | X3 | 4.03 | 59.58 | 2.13 |
| X12 | 3.76 | 62.99 | 3.00 |  | X25 | 3.92 | 63.51 | 6.38 |
| X21 | 3.66 | 66.65 | 2.60 |  | X15 | 3.72 | 67.22 | 5.13 |
| X19 | 3.44 | 70.09 | 2.96 |  | X10 | 3.64 | 70.86 | 4.22 |
| **Rainforest:Plantations** | | |  |  | **Rainforest:Gardens** | | |  |
|  |  |  |  |  |  |  |  |  |
|  | Individual contribution to differences (%) | Cumulative contribution (%) | Mean relative abundance (%) |  |  |  |  |  |
| X15 | 8.06 | 8.06 | 5.13 |  |  |  |  |  |
| X19 | 7.06 | 15.12 | 2.96 |  |  |  |  |  |
| X21 | 6.69 | 21.81 | 2.60 |  |  |  |  |  |
| X25 | 6.47 | 28.28 | 6.38 |  |  |  |  |  |
| X6 | 5.72 | 33.99 | 1.58 |  |  |  |  |  |
| X5 | 5.70 | 39.70 | 5.46 |  |  |  |  |  |
| X16 | 5.55 | 45.25 | 5.93 |  |  |  |  |  |
| X18 | 4.29 | 49.53 | 4.77 |  |  |  |  |  |
| X10 | 3.88 | 53.41 | 4.22 |  |  |  |  |  |
| X14 | 3.71 | 57.13 | 5.16 |  |  |  |  |  |
| X13 | 3.65 | 60.77 | 5.89 |  |  |  |  |  |
| X22 | 3.62 | 64.39 | 7.11 |  |  |  |  |  |
| X9 | 3.62 | 68.01 | 6.63 |  |  |  |  |  |
| X4 | 3.54 | 71.55 | 4.12 |  |  |  |  |  |
| **Plantations:Gardens** | | |  |  |  |  |  |  |

Table 1. OTUs that contributed a combined ~70% to Bray-Curtis measures of diversity for each pairwise comparison between land use types. Figures obtained using SIMPER.

References:

Oksanen, J. et al. 2019, *vegan: Community Ecology Package.* R package version 2. 5-6. CRAN, viewed 21 November 2019, <https://CRAN.R-project.org/package=vegan>

Pena, E. and Slate, E. 2019, *gvlma: Global Validation of Linear Models Assumptions.* R package version 1.0.0.3. CRAN, viewed 21 November 2019, <https://CRAN.R-project.org/package=gvlma>

Wickham, et al. 2019, *dplyr: A Grammar of Data Manipulation*. R package version 0.8.3. CRAN, viewed 21 November 2019, <https://CRAN.R-project.org/package=dplyr>

Wickham, H. 2016, *ggplot2: Elegant Graphics for Data Analysis,* 2nd edn, Springer-Verlag, New York.