**Multivariate Analysis Block Project**

Note to the instructor that this is a demonstration of the type of assignment that students have completed in an advanced biological analysis course. It is just provided here as an example of something you could build off with the dataset project provided. It used a different dataset than the one that is used in this project – but the assignment could be adapted to use the included dataset. It also includes information specific to the campus where I teach that will not be relevant (e.g., presentation at Student Academic conference, use of CSE citation style guide, use of JASP stats package to conduct a two-way ANOVA).

**I. Introduction**

In this block-long project you will develop your researcher, thinker and communicator skills by exploring an ecological dataset. You will develop a research hypothesis/es, gain an understanding of how the data are collected and archived, run analysis to test your statistical and research hypotheses, and write-up and present a poster on your research findings. This project provides you with the opportunity to immerse yourself in the process of biological analysis by engaging first hand with this dataset and envisioning how you could replicate and build on it in the field.

Each group (2 students/group) will be responsible for completing a research poster that will be presented during the last day of class during our ‘Advanced Biological Analysis’ Symposium that will be held in the Roger Epp Room. Your poster will be projected on the class screen (rather than you having to print it out). You also have the option to present your poster at the Student Academic Conference. Final project components include:

1. Two preparatory assignments (information will be posted on eClass on each of these assignments separately from this document). These assignments will include some information that will become part of your final poster. Be sure to incorporate feedback from these assignments when constructing your poster!
2. Peer review of another group’s poster (this will be done individually).
3. Research poster presented to the class (that incorporates feedback from peer review).
4. Electronic evaluation of group members - will be provided on eClass. It is important to fairly evaluate the contributions of each group member – as it is expected that ALL members should be contributing equally, but this does not always happen so this gives you an opportunity to let your instructor know how working in a group was for your project. Marks will reflect the contributions of group members.

**II. Due Dates and Grade Values (Table 1)**

Table 1. List of due dates for AUBIO 315 Project

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **Due Date** | **Grade Value (% of final course grade)** | **Additional Details** |
| Preparatory Assignment #1 | Friday Sept 7 | 10% | Assignment info will be posted on eClass. Submit assignment on eClass. |
| Preparatory Assignment #2 | Thursday Sept 13 | 10% | Assignment info will be posted on eClass. Submit assignment on eClass. |
| Peer Review Submission | Monday Sept 17—before am class | n/a | Submit a draft poster on eClass. You will then be assigned another poster to provide feedback on. |
| Peer Review Due | Monday Sept 17—due at the start of pm class | 5% | Download peer review form from eClass and submit your completed evaluation form on eClass. Anne will distribute feedback. |
| Poster submission and presentation of poster | Tuesday Sept 18 – due at the start of class | Poster (30%) Presentation (10%) | Submit your completed poster on eClass prior to start of am class. See detailed evaluative rubric provided in Table 2 below. |
| Group member feedback | End of day Tuesday Sept 18 | n/a – but required | Complete on eClass |
| Student Academic Conference (SAC) Presentation | TBD (optional) | 2% bonus on your grade if you present at SAC | Information on abstract submission will be provided to those who are interested in presenting at SAC. Contact Anne for more information. |

**III. Instructions for Project Components**

**A. Preparatory Assignments.**

**Preparatory Assignment Part I. (16 pts)**

**Background for your project (Complete individually – but you can search for articles together with your teammates).**

Answer the following assignment components in Part I based on **one** peer-reviewed primary literature source (journal article) that is relevant to your dataset. (Hint: Start with the Biology Library Guide - [https://guides.library.ualberta.ca/augustana/biology)](https://guides.library.ualberta.ca/augustana/biology%29). The articles that you and your teammates select will be useful to you in developing your research hypothesis for your own study so choose them carefully and make sure that they are relevant (e.g., geographical context, type of organism(s) being studied, types of treatments/grouping variables being tested). Consider the impact factor of your article too (why is it important to consider the impact factor of a journal when selecting an article from it?).

1. \_\_\_(1) List the article in CSE format (Exception: you do not have to abbreviate journal names). In addition provide the impact factor for the journal (be sure to include what time interval it captures).(0.5 deduction for each error)
2. \_\_\_(2) Write out the research question/objective of the study, along with the associated research hypothesis. (Even if they did not explicitly provide a research hypothesis you can still determine this from the information provided in the study). If they had multiple questions and hypotheses – just pick one.
3. \_\_\_(2) Briefly describe the type of study design they used.
4. \_\_\_ (2) List the variables the researchers controlled for in their study. Have they been explicit about identifying control variables – you may have to dig into the methods to determine this? Identify/propose at least four variables that they should have controlled for in the study even if it wasn’t part of their design.
5. \_\_\_ (2) What is the statistical scope of inference for the study?
6. \_\_\_ (4) List an analytical (not descriptive!) statistical test performed in the study. For that test – identify which variables were included as the response/dependent and explanatory/independent variables. Create an entity-level metadata table that describes these variables. Your table should include information on type of variable (continuous, categorical and response/explanatory), units (if applicable), and a description field (provides information so that a reader would know what that variable is in the context of the study being conducted).
7. \_\_\_ (3) Design a sample datasheet that you could use to collect the raw data for the test in #6 above – make sure to include identifying project-level information that should also be entered on the datasheet, and any other information that you think should be included.

**Part II. Complete the following components with your teammate(s) based on your project dataset – one submission per group.**

**Total (14 pts)**

**Hypothesis development**

1. (4) Write out a research hypothesis/prediction statement for your study dataset based on the studies that you have collectively researched as a team and any additional research that your team has conducted (Hint: remember the ‘if, then, because’ statements from AUBIO 253).

**Biodiversity measures**

1. (10 pts total; 2 points for table formatting/organization) Pick two study units/sites/plots from your study and quantify their diversity as describes in i-iv below (Note: in your poster you should present the diversity measures for all of your sites – providing measure of central tendency and spread for each treatment/category/group that you are comparing). Be explicit about which two study units/sites/plots you are comparing. **Present your results in a properly formatted table(s)**.

i) (2) Calculate the species richness for each of the two sites.

ii) (2) Calculate alpha diversity using the Shannon index for each of the two sites.

iii) (2) Calculate alpha diversity using the inverse Simpson dominance index for each of the two sites.

v) (2) Calculate beta diversity/species turnover between the two sites.

vi) Attach the file where you have done your calculations for instructor to review.

**Submission:**

Submit both Part I (one per person) and Part II (one per group) on eClass prior to the 9 am class on Monday September 10th. **Please be sure to NUMBER each part of your assignment so it is clear to me what part of the assignment your answer refers to.** **When submitting Part I be sure to include a copy of the article that you used to answer the questions, and when submitting Part II be sure to include the spreadsheet file you used to do your calculations with.**

**Block-long Project Preparatory Assignment #2 (New due date: Friday September 14, 9 am)**

Complete this assignment with your block project teammate(s); value: 10% of your final grade.

The goal of this assignment is to develop your research and thinker core academic skills by having you complete univariate and multivariate statistical analyses on a project dataset that you have been given. This will be a dataset that you will continue to explore in class on Friday after you have submitted your assignment. You should use information from this assignment (and feedback you receive on it) when developing your final poster that you will present at the Advanced Biological Analysis Symposium. For example – consider whether your ordination is presented in an appropriate way, along with the output of your statistical tests. Even though your project has a different dataset, the types of statistical analyses that you are conducting are very complimentary/similar.

**Background Information/Metadata**

Brief description: In this study dataset you are exploring the potential effects of wellsites (both those no longer used (plugged and abandoned) and currently producing across a series of age classes) on the plant community of grasslands in the Pawnee National Grasslands in Colorado, USA. Sites were sampled in 2014. You are interested in exploring whether the plant community differs between these different groups of wellsites, along with whether the effects differ based on the distance of the grassland plant community from the wellsite (20, 50, and 100 m distance from the well pad).

**Main Matrix** file contains an ID field (combination of site number, production group and distance) and then a column for each species that was recorded (using seven letter species codes). Note that species have already been relativized to range from 0-1 and rare species (only found in one site n=27) have been removed.

Information on variables in the **second matrix** are provided in Table 2. A description of the species codes and detailed species attribute is provided in Supplementary Table 1 (Excel file posted on eClass).

Table 2. Information on variables found in second matrix

|  |  |
| --- | --- |
| Variable | Description |
| ID | Unique identifier for each site that matches up with the ID in main matrix |
| Group | Production groups for plots adjacent to well pads that are either inactive or have been active for several different time periods PA = Plugged and abandoned in the 1980s, PR1 = Producing since 1980-1990, PR2 = Producing since 2000-2005, PR3 = Producing since 2006-2013 |
| Prod | this is a binary code, producing= 1, plugged and abandoned= 0 |
| Distance | 100 m, 50 m and 20 m from the wellpad |
| PxDcombined | This is production group x distance combination variable |
| cover | % mean abundance of non bare ground |
| bareground | % mean abundance  of bare ground |
| D | 1-Simpson’s Dominance  |
| D’ | Inverse Simpson’s Index |
| H | Shannon diversity |
| nativeS | The richness of species that are native to the area |
| nonnativeS | The richness of species that are introduced to the area |

****

****

**I. (18 pts) Univariate Analyses**

Perform a two-way ANOVA that examines the effects of **group** and **location** on **non-native species richness** and answer the following questions.

1. (3) Justify whether your data meet the assumptions and conditions of the ANOVA test (you can provide statistical output that provides evidence).
2. (2) Report your results from the two-way ANOVA (F, df, P-value) – don’t just copy and paste the table from the statistics program. Be clear about what your response variable is (e.g.,. is it transformed?)
3. (3) List the null hypotheses for your ANOVA and state what your conclusions are in relation to your null hypotheses for each of them.
4. (2) Conduct appropriate post-hoc tests and report your findings. Be explicit about what family-wise correction factor you used (pick one but do NOT use Scheffe). YOU CAN JUST COPY AND PASTE YOUR TABLE FROM JASP/SPSS HERE RATHER THAN CREATE A NEW ONE.
5. (3) Create an appropriate type of graph of non-native richness that compares the effects of group and location. Be sure that your graph includes all necessary components (e.g., stand alone figure caption, label axes and units, legend away from the x-axis, etc.) and that it is clear to your reader which groups and/or locations are significantly different from one another. -0.5 for every error.
6. (3) In your own words state your conclusions from the two-way ANOVA referring to your figure.
7. (2) Would you be comfortable making inferences about cause and effect from this dataset? Why or why not?

**II. (19 pts) Multivariate analyses.**

**You will perform a NMS ordination, MRPP and Indicator Species Analysis to explore the effect of group on the plant community as a whole. Note that in this example the species data have already been relativized on scale of 0-1 for you.**

1. (2) Run an NMS autopilot (quick and dirty) and copy and paste your scree plot from your preliminary ordination runs here. How many dimensions does it suggest your final NMS solution/ordination should be?
2. (2) Do you need to transform your data for these analyses? Why or why not?
3. (6) Run your final ordination and display a **2-dimensional solution** (even if PC-ORD suggests a 3-D solution – let’s stick with a 2-D one to keep things less complicated!) for your NMS ordination – be sure to include the amount of variation explained for each axis (2), a properly labeled legend (i.e., codes of 1, 2, 3 etc are not very useful so make sure your legend labels are meaningful)(1), and properly formatted figure caption – including referring to what distance measure you used and the method of ordination used (2). **Please also submit your graph row and graph column files separately with your assignment.**
4. (2) In your own words describe the patterns visualized in the ordination plot.
5. (1) Why can’t you run a perMANOVA with this dataset?
6. (2) Because you can’t run a perMANOVA, instead run two MRPPs to compare the wellsite groups and locations separately. In your own words state your conclusions from each MRPP (Be sure to include your A and P-values (you do not have to adjust them for this assignment).
7. (2) Run an Indicator Species Analysis and provide a table that lists (only) the significant indicator species for each group. Use the raw P-values that are provided in the Monte Carlo test of significance.
8. (2) State your conclusions from your indicator species analysis.

**Submission:**

Submit electronic Word document (one per group) on eClass prior to the 9 am class on Friday September 14th. **Please be sure to NUMBER each part of your assignment so it is clear to me what part of the assignment your answer refers to.** **Please also submit your graph row and graph column files separately with your assignment along with your statistical output file from the univariate analysis (e.g., JASP results file).**

**B. Peer Review (done individually)**

Peer review is an important part of the scientific process. When a scientist publishes an article in a refereed journal it has already undergone peer review by at least two scientists and a journal editor. The authors have had to incorporate feedback and make changes to their article based on the feedback that they receive from their peers. This process can greatly enhance the quality of a poster (or a paper!) by showing the authors areas where information should be clarified and it can also identify issues in the paper. Thus, providing feedback (and receiving it too) from your peers is a valuable way to learn more about how to construct an effective poster. This is your chance to take advantage of peer review by having your poster reviewed by a colleague, and in turn review someone else’s poster. You will be amazed at how much easier it is to identify issues and challenges in other people’s work compared with your own. This is a required part of the assignment. A set of peer review guidelines will be posted on eClass that will be used to evaluate posters. I strongly encourage you to review the guidelines and apply them to your own poster before submitting it for peer review. You will have the morning class session to complete the feedback on the poster you review (see due dates in Section I above). You and your poster will benefit from this experience and you can expect that your grade will reflect the peer review having improved the quality of your poster. You will be evaluated on the quality of the peer review that you provide to your fellow student.

**C. Scientific Poster**

**i) Poster.** Prepare your scientific poster using the MS Powerpoint template provided on eClass. Because you will be presenting the poster in class on the projector it should be in landscape lay-out (wider than long). The template has already adjusted the size to the appropriate size if you are planning on presenting at SAC (optional but encouraged). Remember to personalize the poster and make it your own. I recommend following instructions in Purrington’s *Advice on designing scientific posters* available at <http://www.swarthmore.edu/NatSci/cpurrin1/posteradvice.htm>. In addition, Pechenik’s (2013) suggestions from chapter 12 should prove valuable as well (posted on eClass). Remember that the goal of a poster is not to squish as much information as possible – you want to tell a story in a way that attracts people to want to look at your poster, rather than be scared off by information overload and cluttered disarray! Bullet form is perfectly acceptable in a poster! Be sure to provide a clear, logical and well-structured narrative in your poster that flows well.

For your in-class presentation of your poster you can just project a virtual copy of your poster using powerpoint. Be sure to upload a copy of your final poster to eClass prior to the start of the morning class so that it is ready for you to showcase it during our Advanced Biological Analysis Symposium. This will be the version of your poster that I will grade too.

**Sections to include in your poster (and be sure to incorporate feedback from peer review)**

**Introduction**

Should include background information about the topic that you are exploring that sets the stage for the experiment you are conducting and why it is relevant. You should incorporate background information from a minimum of **three** primary literature sources (Wikipedia is not a primary literature source! - <http://guides.library.ualberta.ca/augustana_biology> ). What is the objective of the study? You should clearly state your objective(s) and what the hypothesis/es you are testing is/are. Your introduction should end with you providing a research hypothesis/es related to your objective(s). Remember that a research hypothesis is different from a null hypothesis – this is not the place for your null hypothesis/es.

**Methods**

Briefly describe the methods used. Unlike in a research paper where you would describe the methods in enough detail that a reader could replicate your experiment, in a poster you want to be more brief with your methods. You can use the methods taken from your metadata files as a starting point, but be sure to put them into your own words and be selective about what is appropriate to include. You also need to tell your reader what types of statistical analysis you did. This is where your null hypothesis/es belong.

**Results**

Report the results in a tidy and concise Results section. Your results section must include **figures** (and tables may be appropriate too**)** that you refer to in the written portion of your results section. Be sure that you have fully exploited the full potential of the dataset with your analyses that you present. Your results section should include alpha and beta diversity values (if appropriate), your ordination graph(s), as well as any other statistical analyses conducted as appropriate (e.g., one or two-way ANOVA, perMANOVA, MRPP/MRBP, ISA, etc.). Remember, NO INTERPRETATION HERE! Be sure to include appropriate captions in the proper locations for each table and figure. HINT: Look back to your class notes and in-class assignments and feedback on them to ensure you have included all the necessary information!

**Discussion**

A brief discussion of your findings put in the context of primary literature (minimum of three peer-reviewed articles). Was your hypothesis rejected or accepted? What would be the next steps based on your findings?

This is the place where you explain the findings from your results section. Do not just restate your results instead discuss whether your findings supported or rejected your research hypothesis/es? Why do you think this is – what are possible explanations for your findings or lack thereof? How does this connect with other findings related to the topic you are examining? Cite at least three primary literature sources in this section. You should also discuss sources of potential error in this section and what future research might be valuable to build on the research that you conducted in this study.

**Conclusions**

What are the main take-home messages of your study? Synthesize them here. This should be short and sweet and just hit the high points from your study!

**References**

Also remember to use the CSE Quickguide posted on the library website – you need to follow CSE format when citing articles in your poster – the only exception to this is that I consider it acceptable to NOT abbreviate journal names (I don’t see the point of this – I am not interested in checking to see if you used the correct abbreviations – it wastes both our time imho!)

**ii) Presentation**. Your group will complete a **five-seven minute** talk in class during our Advanced Biological Analysis Symposium that highlights the key findings of your poster (you will use your poster as the backdrop – it will be posted on the projector during class). Each group member should be an equal partner during the oral presentation! See the detailed rubric provided below. Leave us in awe of your poster and presentation of it! Respect the time limits and be enthusiastic in your delivery of your presentation about the poster – you have worked hard – now is your opportunity to showcase what you found. Make sure you are engaging with your poster as you talk – but don’t talk at your poster – talk to your audience. Also remember that your audience will be familiar with the statistical techniques you used so you shouldn’t spend a lot of time talking about the techniques – use that time instead to talk about the results of your statistical test. Remember to have fun. There will be additional time for questions after your presentation.

**D. Group Member feedback**

Please provide feedback on your experience working with your team member by completing the evaluation form that will be posted on eClass. Be sure to complete it twice – once to evaluate yourself and a second time to evaluate your partner. Only complete it after you have finished your presentation at the symposium on the last day of class.

|  |  |  |
| --- | --- | --- |
| **Section** | **Grading Rubric for Poster and Presentation (value in parentheses is maximum value for that category)** |  |
| **Content and Organization of Ideas** | Title is clear and representative of the purpose or key finding of the study. Free of jargon and not overly lengthy: Alluring |  |
| Objective(s) clearly stated and easy to locate, show how the research questions/hypotheses related to the objective(s) are investigated. |  |
| Appropriate information in well-organized sections (introduction, methods, results, discussion, conclusions, references) |  |
| Analyses in results section are appropriate to the dataset and have exploited the full potential of the dataset |  |
| **Unacceptable (0-9)** | **Fair (10-13)** | **Good (14-17)** | **Excellent (18-20)** | **Total** |
| Poor content quality, missing key information about most topics. Lacks grasp of information and appears thrown together. No connectivity between sections.  | Content missing some key information (e.g., missing sections of results; tables and figures w/out supporting text) or info placed in wrong section. Some areas covered adequately but other have not. Poor connectivity b/w sections (e.g., new ideas in results) or repetition between sections. | Good content. For most part complete, however there is room for adding some key information in at least one of the main areas of focus (e.g., introductory context is not well developed). Room for improvement in connectivity between sections. | Outstanding content. Significant background work, easy to follow, and interesting. Demonstrates in-depth knowledge of topic. Excellent connectivity (e.g., everything in introduction matches up with all other sections). Table and figures are self-sufficient.  |  |
|  | **Unacceptable (0-4)** | **Fair (5-6)** | **Good (7-8)** | **Excellent (9-10)** |  |
| **Poster Design** | Poster is visually unappealing, little thought in designing it, poster very busy and no use of pictures/figure. Font is too large or small, too much text – cluttered, hard to read (e.g., bad colour scheme). Lots of typos. CSE not used. No peer review feedback incorporated. | Content is poorly organized, poster choppy, not well thought out, little use of animations, pictures/figures, colour etc. font size often okay but also sometimes too small or big. Some badly cluttered information on poster, several typos. CSE used but riddled with errors. Some changes made from peer review. | Content is generally well organized. Some use of lists/bullets, but not consistently, generally poster looks good but there is room for improvement. Good use of pictures (including good resolution) and figures to enhance poster info. Poster uses good font size, some room for improvement (e.g., some sections with excessive text), only few typos, including CSE references. Feedback from peer review primarily incorporated (where appropriate). | Professional appearance. Content well organized: good use of titles and headings, bulleted lists as appropriate. Excellent use of (good resolution) pictures and figures to enhance poster. Poster concise, uncluttered, easy to read, appropriate font size, free of CSE formatting issues, typos, and grammatical errors. Innovative! Thoughtful attention to peer review feedback. |  |
|  | **Unacceptable (0-4)** | **Fair (5-6)** | **Good (7-8)** | **Excellent (9-10)** |  |
| **Oral Presentation Impression/Interactions with Audience** | No eye contact, voice too low to be heard, poor tone and style. Read from prepared sheets all/most of time. Student not prepared, presentation is choppy, lacks flow. | Little eye contact, poor voice projection, poor tone and style. Use sheets but do not read 100 % of time (≈ 50 %). Student is prepared but can’t answer questions very well. Apathetic. | Good enthusiasm and engagement with audience (e.g., good eye contact, voice projection, and tone and style). However still has some reliance on the use of notes. Student is well prepared and can answer questions well. . | Excellent engagement with audience: everyone wants to come hear what you say b/c of your passion (maintains interest). Very effective at responding to questions and generating further ones. Integrating knowledge well (e.g., can address the rationale for what you were studying). Not reading from notes. |  |
| Total Value: (40 % of your final course grade) | **Total** |  **/40** |