**PC Version**

**Introduction**

 The purpose of this exercise is to familiarize you with some basic methods and standards for presenting scientific data. This exercise will walk you through conducting basic data manipulations, presenting data in a table, and generating several figures using Excel. You will need these skills not only for this course but also throughout your academic career and I can guarantee you’ll get higher marks on your assignments if you pay close attention! To make this exercise more interesting we’ll be using an actual dataset from real research[[1]](#footnote-1). The overall goal of this research was to determine if demographic parameters were influencing the genetic diversity of walleye populations.

Walleye are nocturnal top predators in many North America lakes. They are also highly sought after sport fish so they are closely monitored and heavily stocked. With this in mind, the researchers gathered genetic data from walleye populations with varying demographic characteristics and stocking intensities. The objective was to identify possible correlations between demographic factors, such as population size and sex ratios, with genetic diversity. The genetic diversity data will be given to you since such analysis is outside the scope of this exercise. We will focus on population level heterozygosity (HE) and allelic richness (AR; basically the number of alleles occurring per locus in each population). But you will have to calculate several demographic variables. Next, you’ll assess the influence of stocking on genetic diversity in two ways. First you’ll take a qualitative approach and see if genetic diversity different between three classes of lakes. You’ll do a few summary statistics and make a table and bar graph of your results. Next, we’ll take a more quantitative approach and look at the correlation between stocking intensity and genetic diversity. Finally, you’ll produce a brief write-up interpreting your results and referencing your figures.

**Step 1 - Filling in the Table**

1. Open ‘Walleye\_Student dataset.xlsx’ on your computer[[2]](#footnote-2). There’s a dataset prepared for you on the ‘Data’ tab. Unfortunately, I spilt coffee on my computer and it deleted the data for the sex ratio (*M/F*) and population estimates (*PE*). Good news, you get to generate the data for these two columns.
2. Click on the ‘**Demographic data**’ tab. This is the raw data that we’ll generate our demographic variables from. Population estimates of each sex based on mark-recapture data from state agencies.
3. Type **=sum(E3:F3)** into cell G3, this tells excel to add up the contents of these cells. Next, you can get excel to auto fill the rest of that column by highlighting that cell and double clicking on the small square on the lower right corner.🡨 that small square!
4. Most statistical analysis requires normally distributed data. These data are not normally distributed so we’ll fix this by transforming the data. You’ll want to take the natural log (ln) of each populations estimate. Click on cell H3. **Type =ln(G3)** then do the double click on square auto fill trick.
5. Now we want to use the numbers of males and females to calculate a sex ratio. Easy enough, just click on **I3** and type **=** now click on the number of males in **E3**, hit the / and click on the number of females in **F3**. You just built a formula (**=E3/F3**) that will divide the number of males by females, giving you the sex ratio. Do the double click square auto fill trick (man, I need to come up with a better name for that) and viola you’ve done it for all the data.
6. Okay, here’s a trick that will save you a ton of time it’s called **pivot tables**. Highlight the demographic data (**B2:I61**) under inserts 🡪 click PivotTable and okay.
7. Click and drag **Lake** into the rows this will generate a title for each row. Next drag **Ln(Sum)** to the **Values** box. Then click on the drop down arrow and value field setting to the right of **Sum of ln(…** this allows us to toggle what this field calculates, the default is to add up all the numbers (excel really likes to add). Select average 🡪 click OK.
8. Drag **Ln(Sum)** one more time into that box, click the drop down arrow and value field setting again and select **Stdev**. Now we have the mean and standard devastation of the ln(PE). Do the same thing for the sex ratio (M/F).
9. You’ll want to paste this data into our Table 1 but first we have to deal with an existential crisis, the lakes are in a different order in Table 1 than our pivot table! Highlight the top row (**B3:L3**) of Table 1 and click the sort and filter, then the filter button  found in the top ribbon. You’ll notice by the column label there’s a little drop down arrow; this will let you sort the whole dataset by that column. Click on that arrow next to **Lake** and click on ‘sort A to Z’. For text, this will alphabetize the data for numbers it’ll put them in order from low to high. Click on the filter button again to get rid of the arrows. \*\*Name your sheet – what you want to call it\*\*
10. Now we can paste our data into the table. Highlight from **B4:B18** for the mean PE and right click and hit copy (can also use **control c** or **command c** on a mac). Paste the data into the right spot in Table 1 by right clicking🡪 paste special🡪 paste values. This way it doesn’t mess with our nice formatting. Do the same for the mean sex ratio.
11. You’ll notice that there’s a ton of ugly decimal places in our pasted data. Remember the importance of significant figures. I can’t stress this enough, the number of decimal places presented in a table should reflect the accuracy of the data and not just the mathematics of generating the number! Here we have averages from population estimates, which only had 3 or 4 figures so we’ll want to use 3 significant figures. Highlight the data you just pasted in and decrease the number of decimal places you’re showing with the  button in the ribbon.

Now you have a completed and properly formatted table. You’ll need to make a table in just a bit so lets list a few requirements:

* Caption: The table/figure is naked with out one. Provide enough information to interpret table but do not interpret or discuss the results.
* Significant figures: use the appropriate number and be consistent.
* Variable name really long? Use an abbreviation and define it in the caption.
* It’s nice to have a line on the top and bottom of the first row to delineate the column names and a single line at the bottom.
* Generally, you won’t present raw data. Tables are for summaries (usually mean and standard deviation).
* Formatting should match text (i.e., Times New Roman and usually 12 point font but sometimes you’ll need to decrease this or present the table in landscape).

**Step 2 - Using stocking codes to make a bar graph of genetic diversity**

1. Now we’ll want assess if stocking is impacting genetic diversity in these populations. We’ll first do this by summarizing genetic diversity data in each stocking code. These are subjective codes assigned to each lake by management agencies and are used in management decisions about stocking.
* NR = natural reproduction only
* C-NR = self-sustaining but still stocked
* C-ST = dependent on stocking but some natural reproduction
1. Click on the ‘Stocking bargraph’ tab at the bottom. This is where we’ll summarize our genetic data for each code in both a table and figure.
2. You’ll need to setup a table to present the mean and standard deviation of HE and AR for the populations in of the stocking codes. Type ‘Code’ in B3, Heterozygosity in C2, ‘Allelic richness’ in E2, ‘Mean’ in C3 and E3, ‘Stdev’ in D3 and F3, C-NR in B4, C-ST in B5, and NR in B6. \*\*Name sheet again by double clicking the tab\*\*
3. Highlight the data set in Table 1 (**B3:L18**) and generate another PivotTable.
4. Use the same procedure as before but this time the **Rows** will be “**Stocking code**” and you’ll generate a mean and standard deviation for HE and A­R. Refer to previous section.
5. Copy this data and paste the values into the table you just made in the Stocking bar graph tab.
6. Select all the data (B4:F6) and drag it down one row to make some empty space. Now select the NR row and cut and paste it into the empty row you just made. Now the data is organized from natural to stocked.
7. Now, you’ll make this table more presentable. Highlight C2:D2 and click  Do the same for E3:F3. Now highlight B2:F3 and click the borders button  Select “Top and bottom boarders”. Then highlight B6:F6, click the borders button again and click “Bottom border”. Highlight and click the center text button .
8. Highlight all these numbers and reduce the decimal places so only 3 are shown. Also highlight the whole table and change the Font to Times New Roman.
9. Okay, we have a nice looking table. Make a caption for this table (captions appear above tables and below figures, just convention)

**Making a bar graph**

1. In the Stocking bar graph tab, highlight the stocking code and HE (**B4:C6**), under the **Inserts** click column chart, and select **Clustered Column**.
2. Okay, we have a crummy bar graph, excel calls this a column chart but they really don’t know what they are talking about. Lets make this presentable. First, delete the title, this just wastes space and for the most part you’ll present this information in the figure caption.
3. Next, delete the gridlines; they are not needed for most figures.
4. Add line 🡪 double click axis 🡪 paint can (fill lines) 🡪 click line from auto to solid line. Now change color to black. For x axis, double click and navigate to line 🡪 change to black.
5. Right click on the bars and select **Format data series**. Under **Fill**, change this from automatic to **no fill**. Under **border, solid line** change automatic to black.
6. Now, add error bars. Under the **Chart Tools** in the design tab add new chart element, click **Error Bars** then select **More**. Now select **Custom**🡪click on **Specific Value**🡪Click on the  box to the right and highlight the Stdev of HE (**D4:D6**) for both positive and negative error values 🡪 OK. Under display, make sure **Both** is highlighted and click **OK** again.
7. Lets adjust the Y-axis. Double click on the axis to bring up the options menu (may already be there). Under **axis options,** change minimum to 0.75 and Major unit to 0.01. Under **Number** deselect **Linked to source** and reduce decimal places to 2.
8. Under **Add chart** element click **Axis**🡪**Primary Vertical Axis**. Rename this Heterozygosity and change it from bold to regular font. You should also select the whole figure and change the font to Times New Roman.
9. Now you need to come up with a caption for this figure, maybe something like: Figure 1: Mean heterozygosity for walleye populations in Northern Wisconsin with natural reproduction (NR, n=5), stock augmentation (C-NR, n=5), and stocking dependence (C-ST, n=5). Error bars show standard deviation.
10. Whew, now we can see the mean HE for each stocking code and the standard deviation associated with this estimate. We can see that the Stdev overlaps for NR and C-NR but C-ST is much higher and does not overlap. This IS NOT a formal statistical analysis but gives a pretty good indication that C-ST lakes have a higher HE then NR and C-NR, which did not differ from each other.
11. Now do the same thing for AR and see what happens (Skip this if your low on time).

**Step 3 - Using stocking intensity to make a scatterplot of genetic diversity**

1. For our last analysis we’ll be looking at a regression between the sex ratio (independent variable, always on x-axis) and AR (dependent variable, always on the y-axis).
2. In the Data tab, highlight the AR data (E4:E18). Click **Insert🡪Scatter🡪Marked Scatter**.
3. **Right click chart 🡪 Select Data 🡪** click **edit**  box next to the X-value **option 🡪** select the **M/F data (I4:I18) 🡪** click **OK**.
4. Now for the formatting. Click on and delete “title” and delete the gridlines again.
5. Right click on the **Y-axis 🡪 Format Axis 🡪** Change **major unit** to 0.5. Next click **fill** in **line** 🡪 change line from **auto to solid** (refer back for more details)
6. Right click on X-axis 🡪 **Format Axis** 🡪 Change major unit to 2.5. Next click **Number 🡪 unselect “Linked to source”** this will allow you to adjust the decimal place to 1. Follow #5
7. Right click on the data points 🡪**Format Data Series** 🡪 **line and fill** 🡪 **marker 🡪 fill 🡪 no fill 🡪 Change from Automatic to No Fill**. Next, **Border 🡪** Change from **Automatic to Solid line**. Finally, **Marker Options** 🡪 change from Auto to built in 🡪 select size and symbol.
8. Now add axis labels. Click **chart tools** 🡪 **design** 🡪 **Add new element** 🡪 **Primary horizontal** 🡪 **title below axis**. Under **Chart Layout** select **Axis** 🡪 **Change Axis Title** to something sensible like “Sex ratio (M/F)”.
9. Do the same thing for the vertical axis and label it “Allelic richness”.
10. Select the whole graph again. In the **Home menu**🡪change font to **Times New Roman** (excel defaults with bold axis labels, it’s annoying).
11. Now we’ll do an informal statistical analysis on the data. Right click on the data points 🡪 **Add trend line** 🡪 Click on “**Display equation on chart**” and “**Display R-squared value on chart**”. This will run a linear regression showing the relationship between AR and M/F (The R2 indicates the amount of variation in AR that is explained by variation in M/F).
12. Move the resulting equation to the lower left hand corner of the figure and manually decrease the decimal places to 2. \*\* Manually adjust decimals by deleting and change color of regression line\*\*
13. Note: there’s no P-value here! We’d need to run a more formal analysis to really tell if this relationship is significant or not. For the purposes of this exercise, we can assume this is significant (I already did the analysis and can tell you that it is significant).
14. Finally, we have a good-looking figure! Now we NEED a caption. Maybe something like this: Figure 3: Linear regression (dashed line) between allelic richness and male to female sex ratio for 15 walleye populations in northern Wisconsin (*P-value* < 0.05). (Assuming 2 Fig. Previously completed)
15. Your work with the figure is now done! Redo this whole process to make a graph showing the relationship between another demographic parameter and either AR or HE (Or skip this if you are running low on time).

**Step 4 - Brief write up with figures referenced properly!**

* Now we’ll present and comment on our results. You will ALWAYS strive for **clarity** and **brevity** in scientific writing.
* You just copy and paste your tables and figures from excel into word. Make sure you either type or paste your caption as well. Don’t forget that tables go first with their caption above them and figures go second with their captions below.
* Generally, tables and figures appear at the end of your write-up and in that order that they are discussed.
* You will use parenthesis at the end of the first sentence that makes reference to your table or figure.
* You **don’t** want to write things like: Figure 3 shows the relationship between allelic richness and sex ratio. This is self-evident and any reader should have been able to infer that from the figure (kind of the whole point of having a figure).
* Rather write about the pattern you discern from the figure and use a reference to support the statement like this: Our study documented a connection between demographic factors and genetic diversity in walleye populations. Populations with highly skewed sex ratios exhibited decreased genetic diversity possibly owing to a restricted effective population size (Figure 3).
* Okay, your turn. Construct a brief write-up using Table 2 and two figures of your choice.
1. Waterhouse MD, Sloss BL, Isermann DA (2014) Relationships among Walleye Population Characteristics and Genetic Diversity in Northern Wisconsin Lakes. Trans Am Fish Soc 143:744–756. [↑](#footnote-ref-1)
2. This tutorial was assembled using Microsoft Excel 2013 (PC), so your version may look a bit different but all the instructions should translate [↑](#footnote-ref-2)