

Creating and Analyzing a Scatterplot Using Excel

STUDENT LEARNING OBJECTIVES

Students who successfully complete this lab activity will be able to:

- Construct a scatterplot to visualize the relationship between two variables; and
- Evaluate the strength and significance of the relationship between two variables.

BACKGROUND

Ecologists study the relationships between organisms and other organisms (biotic factors) or their surrounding environment (abiotic factors). In other words, ecologists study the link between biotic (e.g., genetic diversity) and abiotic (e.g., temperature) factors or biotic (e.g., genetic diversity) and other biotic (e.g., predation) factors. Often, ecologists will construct a scatterplot to visually represent these relationships. Additionally, ecologists analyze scatterplots using a statistical method called a Pearson Correlation analysis. The Pearson correlation analysis provides the ecologists a measure of the relationship's strength (e.g., weak or strong), as well as a measure of the significance of the relationship. However, several assumptions must be met in order to use the Pearson correlation analysis. First, the sample must be normally distributed (e.g., a bell curve). Second, the relationship between two variables must be linear ($y = mx + b$), where y is the dependent variable; x is the independent variable; m is the slope of the line; and b is the y -intercept.

Today, you will use your data from the natural selection lab to plot a scatterplot and evaluate the strength and significance of each line using a Pearson correlation analysis.

INSTRUCTIONS—THINK ABOUT YOUR DATA

1. In the space below, list the variables that you plan to use in the scatterplot.

2. Where would each of the above variables be placed on the figure?

x-axis =

y-axis =

INSTRUCTIONS—CREATING THE SCATTERPLOT

1. Obtain the paper copy of your results from the Natural Selection lab.

2. Open a "Blank Excel Spreadsheet."

3. Click on the "File" tab.

4. Click on "Save As" and save your Excel spreadsheet to your W drive or the desktop using the following format.

a. Last Name BIO 112 Lab—NS Analysis

5. Click in Cell A1 and type in the name of your x-axis variable.

6. In Cells A2 through A9, type in the values of your x-axis variable.

a. If you have fewer than seven generations, then only enter the data up to the last generation completed.

7. In Cells B1 through K1, type in the name of your y-axis variables.

8. In Cells B2 through K9, type in the values of your y-axis variables (enter only the after reproduction values).

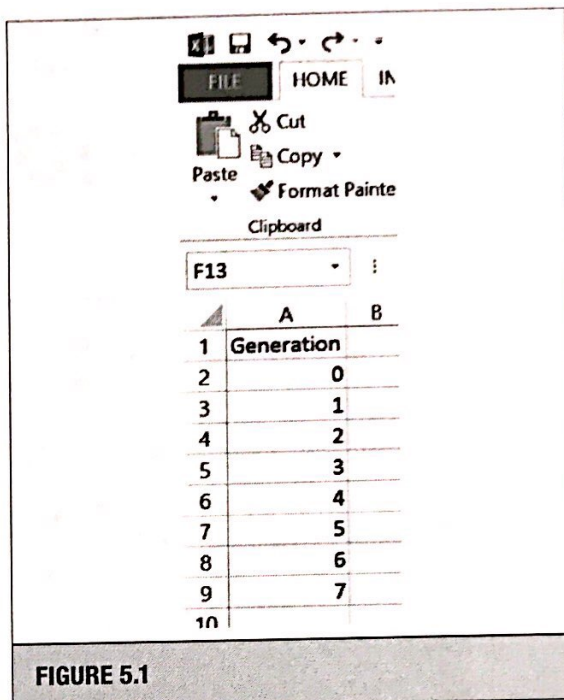


FIGURE 5.1

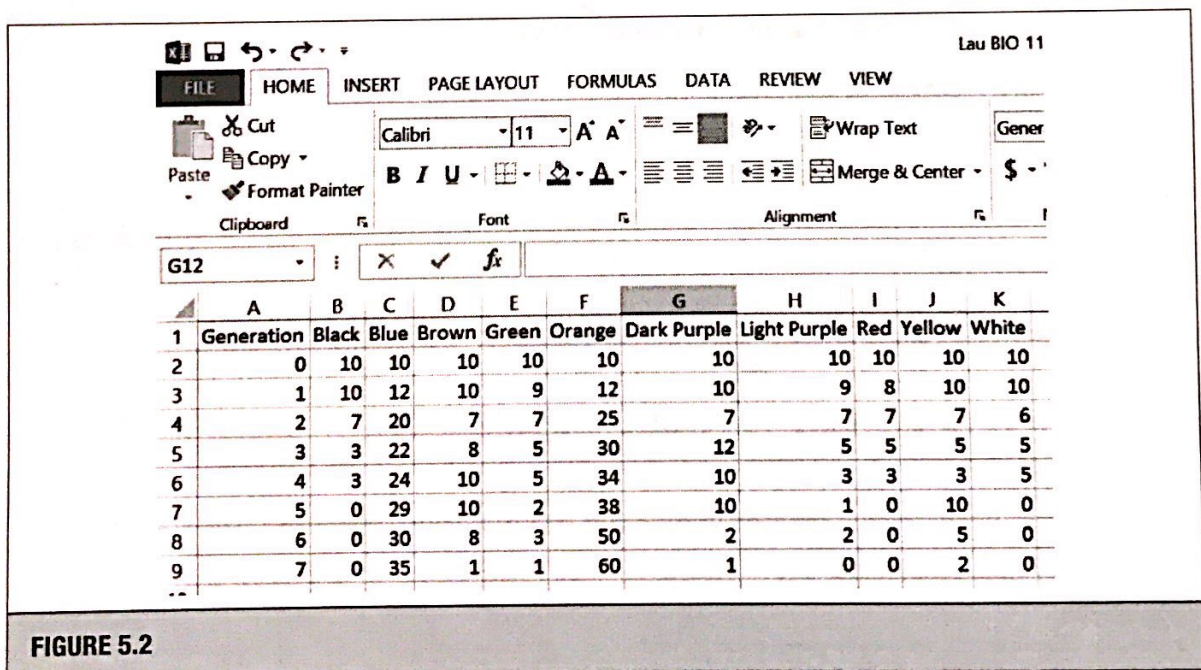


FIGURE 5.2

9. Place your cursor on the faint, grey line between the A and B in the column heading. Double click on this line between each column to change the column width. You will need to complete this trick periodically throughout this lab.
10. Click and hold on Cell A1. Then, drag your cursor to Cell K9 and release. This technique selects all of your data. You should see a faint, green line around all of your data.
11. Click on the "Insert" tab.
12. Move your cursor above the "Scatter" icon and click. Then, click the "Scatter with Smooth Lines and Markers" icon.

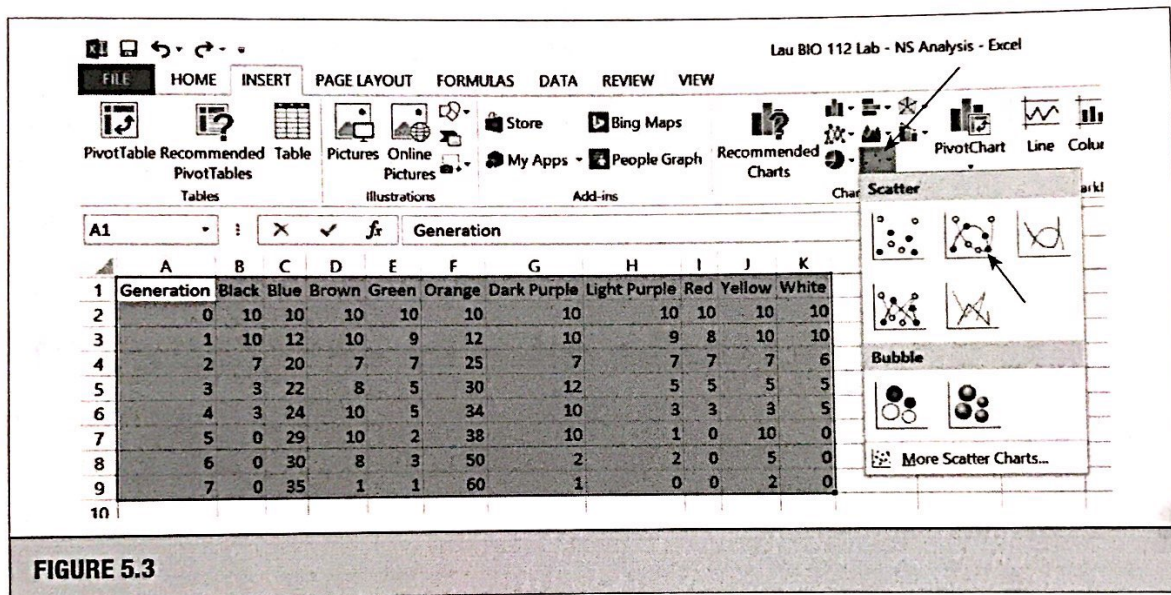


FIGURE 5.3

13. Your figure might look something like the below figure.

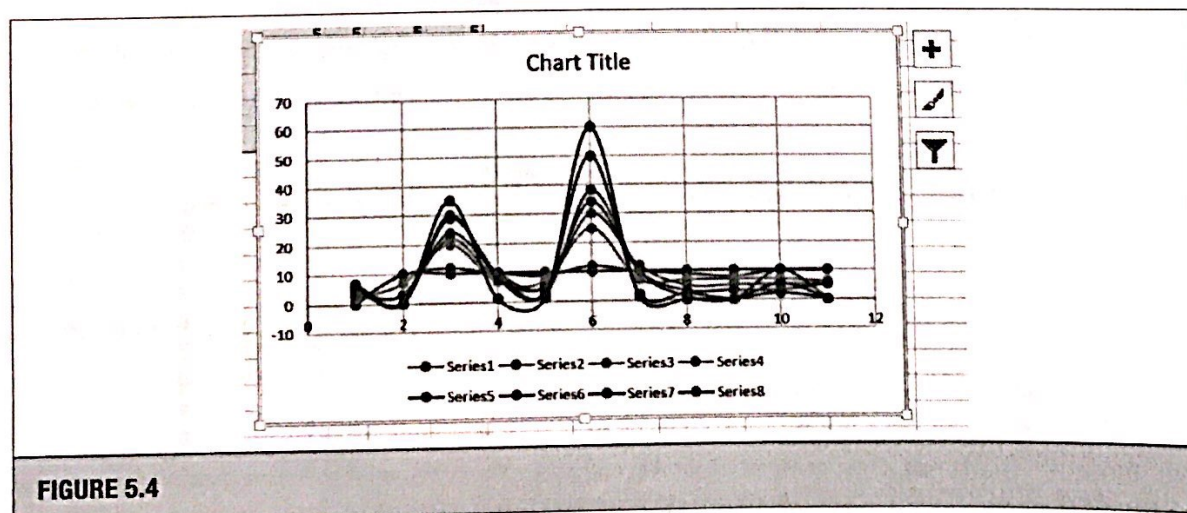
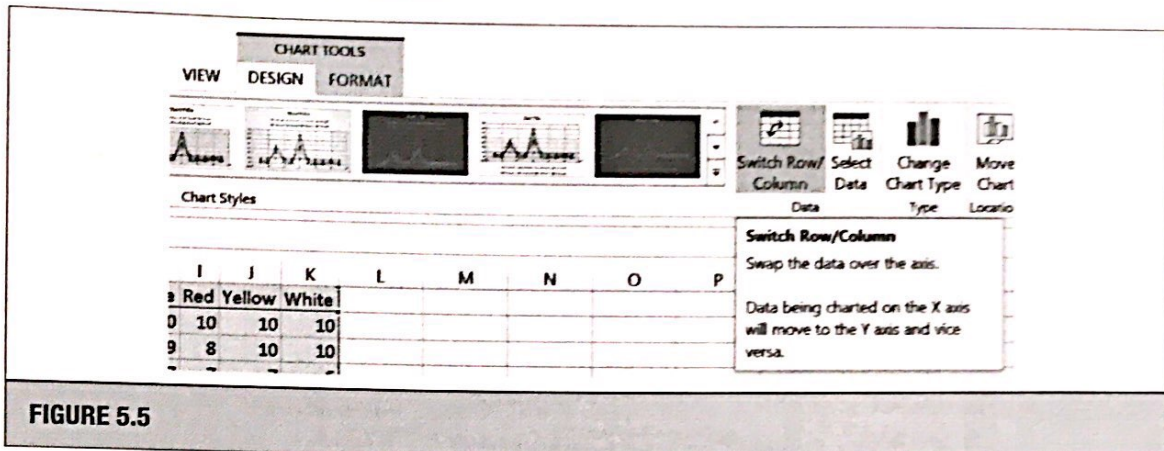
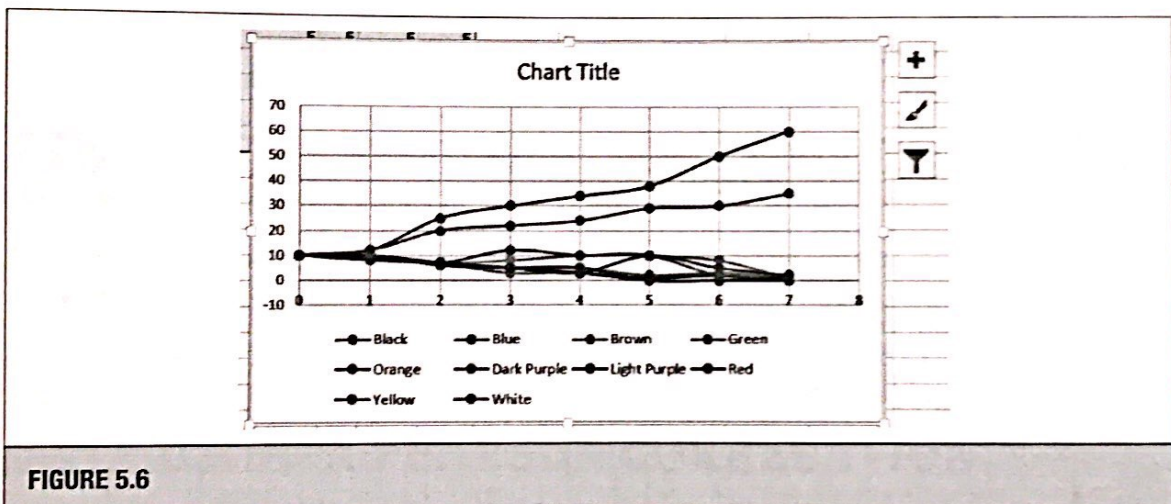


FIGURE 5.4

14. Click on the "Design" tab under the "Chart Tools" tab. If your "Chart Tools" tab is missing, click in the white space in the figure. The "Chart Tools" tab should reappear.
15. Click on "Switch Row/Column."



16. Your figure should now look something like the figure below.



17. Click the "Save" icon (little blue disk at the top left of the Excel screen).
18. Click on the "Chart Title" and Delete. In the space below, explain why we deleted the "Chart Title."
19. Now we need to add axis titles. Make sure that the "Chart Tools" tab is visible.
20. Click "Add Chart Element." A new dialog box should appear.
21. Hover your cursor over "Axis Titles" and then click on "Primary Horizontal" (x-axis).
22. Type in your x-axis title.
23. Follow the same method to add the "Primary Vertical" axis title (y-axis).

24. You may need to adjust the y-axis scale because the y-axis shows negative values or values above 100%. Double click on the y-axis numbers. A new dialog box should appear in the right-hand margin called "Format Axis." Adjust the y-axis minimum to 0% and the maximum to 100%.

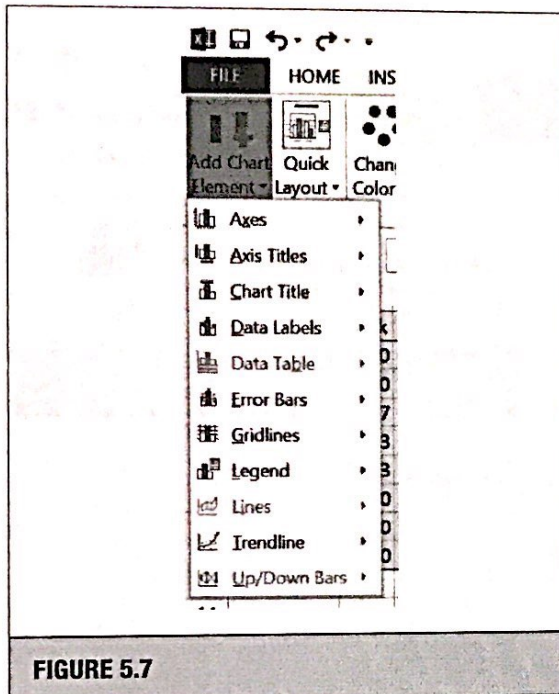


FIGURE 5.7

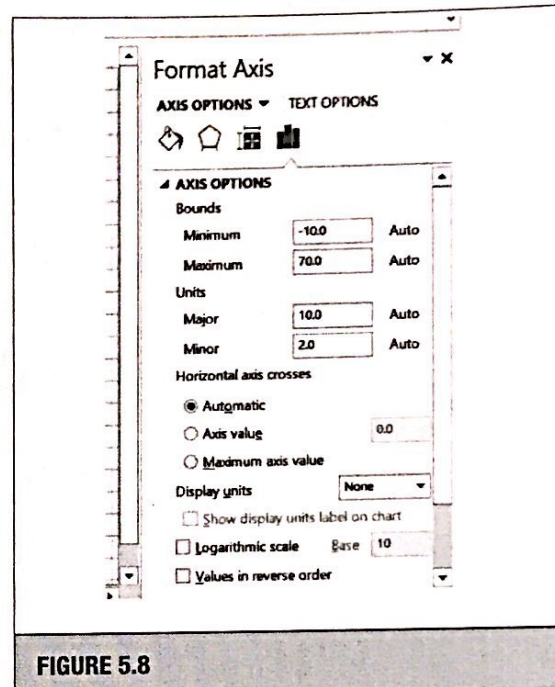


FIGURE 5.8

25. Change the "Minimum" value to 0. Change the "Maximum" value to 100.
26. Hit "Enter."
27. Finally, click once on the horizontal axis lines in the figure. The lines are selected if you see circles at both ends of the lines.

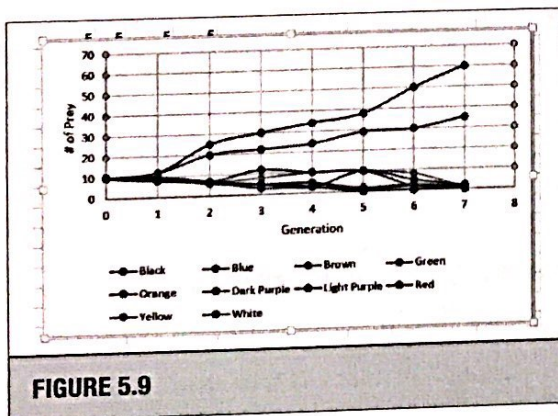


FIGURE 5.9

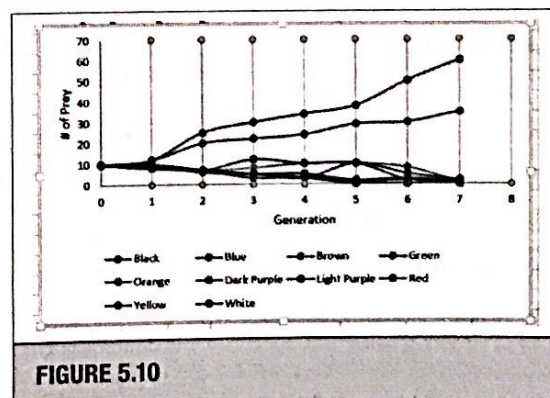
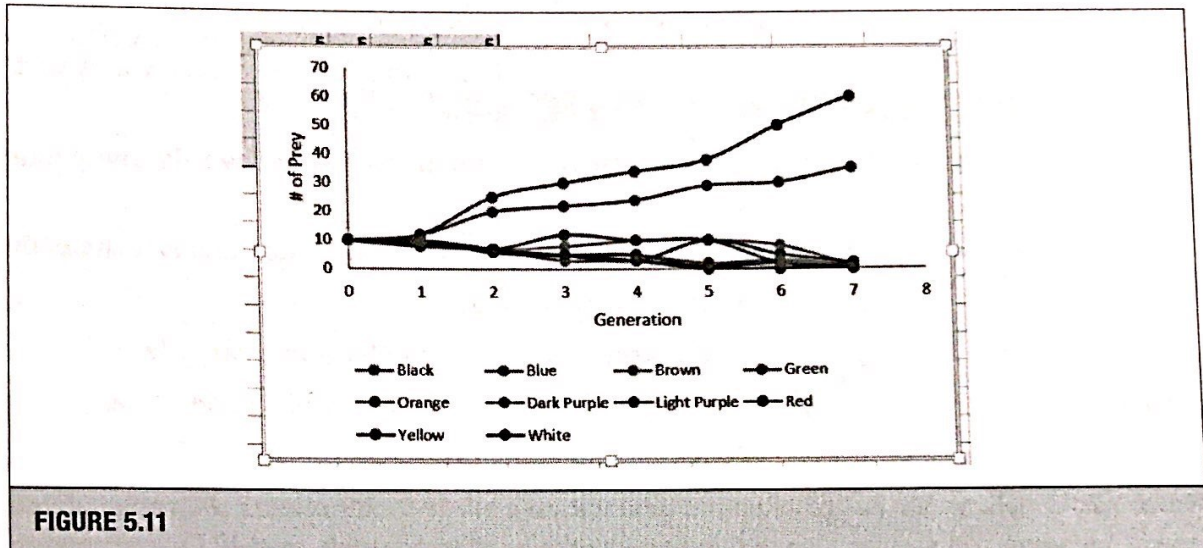


FIGURE 5.10

28. Hit the "Delete" button.
29. Click once on the vertical axis lines in the figure.
30. Hit the "Delete" button.

31. In the space below, explain why both the horizontal and vertical lines were deleted.
32. Click the "Save" icon (little blue disk at the top left of the Excel screen).
33. Your final figure should look something like the below image.



34. You can also move the legend and adjust the figure size. Ask your instructor to assist you with these adjustments.

INSTRUCTIONS—ANALYZING THE DATA USING A PEARSON CORRELATION

1. Click and hold on Cell A1. Then, drag your cursor to Cell K9 and release to select all of your data.
2. Hit "CTRL C" to copy the data table.
3. Click on the "Sheet 2" tab at the bottom of the screen.
4. Click once in Cell A1.
5. Hit "CTRL V" to paste the data table into the new spreadsheet.
6. Click the "Save" icon (little blue disk at the top left of the Excel screen).

7. Click once in Cell A11 and type in Pearson's r.

a. Pearson's r provides the strength of the relationship (or correlation) between two variables. In this example, we will calculate 10 Pearson's r values—one value per chip color. **Note:** Ignore the sign of the Pearson's r value to determine correlation strength.

i. $r \geq 0.70$ indicates a strong correlation

ii. $0.30 < r < 0.70$ indicates a moderate correlation

iii. $r \leq 0.30$ indicates a weak correlation

b. The Pearson correlation also provides the direction of the correlation

i. A positive value indicates a positive correlation. In other words, the y-axis variable increases as the x-axis variable increases.

ii. A negative value indicates a negative correlation. In other words, the y-axis decreases as the x-axis variable increases.

8. We will now calculate Pearson's r using Excel. The format of the command is technically

`=PEARSON(array1, array2)`

where the array1 = the x-axis variable and array2 = the y-axis variable

9. First, let's calculate Pearson's r for the first color. Click once in Cell B11 and type

`=PEARSON(A2:A9, B2:B9)`

Note: A2:A9 selects the range of your x-axis variable while B2:B9 selects the range of your y-axis variable. You should also notice that the range of your x-axis variable is outlined in blue, and the range of your first y-axis variable is outlined in green (see Figure 5.12).

PEARSON		X	✓	f _x	=PEARSON(A2:A9, B2:B9)							
	A	B	C	D	E	F	G	H	I	J	K	L
1	Generation	Black	White	Red	Green	Blue	Yellow	Orange	Purple	Pink	Brown	
2	0	10	10	10	10	10	10	10	10	10	10	
3	1	10	10	8	9	12	10	12	10	9	10	
4	2	7	6	7	7	20	7	25	7	7	7	
5	3	3	5	5	5	22	5	30	12	5	8	
6	4	3	5	3	5	24	3	34	10	3	10	
7	5	0	0	0	2	29	10	38	10	1	10	
8	6	0	0	0	3	30	5	50	2	2	8	
9	7	0	0	0	1	35	2	60	1	0	1	
10												
11	Pearson's r	=PEARSON(A2:A9, B2:B9)										
12												

FIGURE 5.12

10. Hit "Enter."
11. Now, calculate Pearson's r for all colors. Note that array1 will always be A2:A9, while array2 will change depending on the color and column.
12. We now need to format all of the decimals to three decimal places.
 - a. Select all of your Pearson's r values.
 - b. Click the "Home" tab.
 - c. Click on the down arrow next to "General."
 - d. Select "Number."

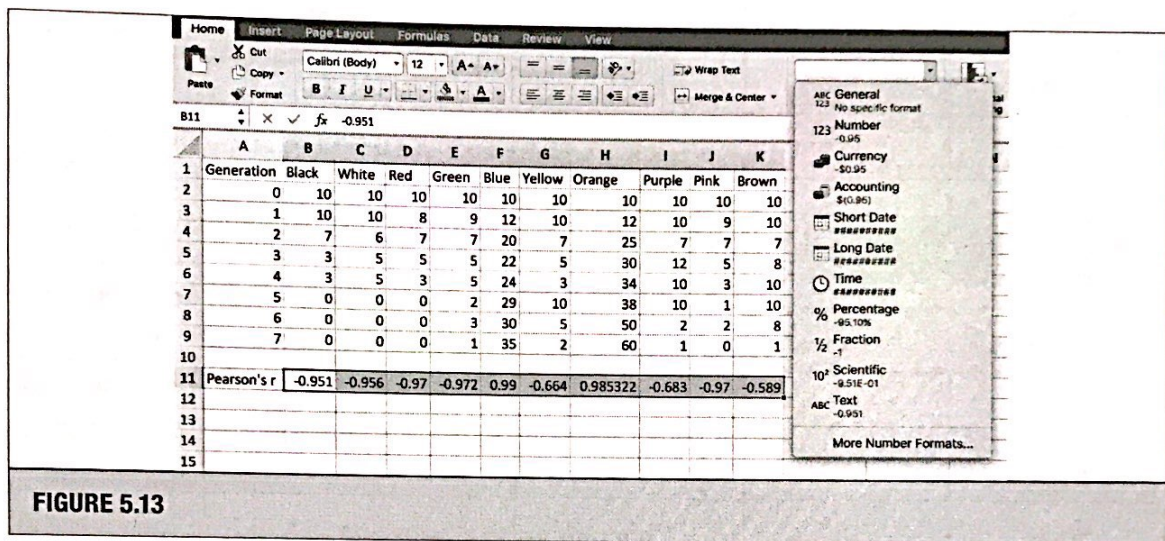


FIGURE 5.13

- e. Use the two icons below to change the decimal places to 3.

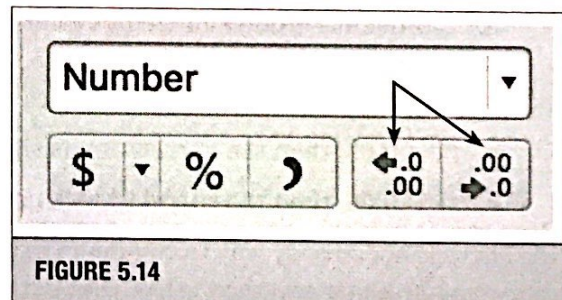


FIGURE 5.14

13. Fill in the first three columns in Table 5.1 below using your calculated Pearson's r and the information in 7 a and b above.

TABLE 5.1 Summary of the relationship strength, direction, and significance between the generation and the number of individual chips per color.					
Color	Pearson's r	Strength (S, M, or W)	Direction (+ or -)	P	Significant (Y or N)
Black	0.961	S	+	0.002	Y
Blue	-0.655	W	-	0.158	N
Brown	-0.828	W	-	0.042	Y
Green	-0.655	W	-	0.158	N
Orange	-0.655	W	-	0.158	N
Dark Purple	-0.686	W	-	0.133	N
Light Purple	-0.828	W	-	0.042	Y
Red	-0.655	W	-	0.158	N
Yellow	-0.655	W	-	0.158	N
White	-0.828	W	-	0.042	Y

14. Next we need to determine whether or not the correlation is significant.
- Significance is determined by calculating P or the probability that the correlation was due to random variation or chance. We set P to ≤ 0.05 , which indicates that there is a less than or equal to a 5% chance that the correlation is due to random variation.
 - If $P \leq 0.05$, then the correlation is significant.
 - If $P > 0.05$, then the correlation is not significant.
15. There is not an easy way to command Excel to calculate P. We must first convert Pearson's r to a t-statistic and then relate that t-statistic to significance. You will learn more about these terms in your upper-division course (e.g., ecology or biostatistics).
16. We will now turn Pearson's r into a t-statistic using Excel. We must also calculate the absolute value (removing negative values) of the t-statistic. The format of the command is technically

$$=ABS((\text{Pearson's } r \text{ value} * \text{SQRT}(\text{sample size}-2)) / (\text{SQRT}(1 - \text{Pearson's } r^2)))$$

where ABS calculates the absolute value

* is "multiplied by"

^ is "taken to the power of"

17. Click once in Cell A12 and type Turning r into t.
18. First, let's convert the first relationship (Pearson's r) into a t-statistic. The sample size of our study is eight (seven generations plus the starting population). **Note:** The sample size may change based on how many generations you completed.

19. Click once in Cell B12 and type

$$=ABS((B11*SQRT(8-2))/(SQRT(1-B11^2)))$$

20. Hit "Enter."

21. Calculate the t-statistic for all 10 relationships. There is a quick trick.

- Click and hold the black square at the bottom right corner of your first t-statistic.
- Drag this box to Cell K12. This trick will tell Excel to use the same command but use the corresponding Cells (i.e., the corresponding Pearson's r values). Check Excel's work by clicking on each cell.

10			
11	Pearson's r	-0.951	-0.956
12	Turning r into t	7.5	8.0
13			

FIGURE 5.15

22. Format the t-statistic to one decimal place.
23. Now let's determine whether or not the Pearson's correlation is significant by relating the t-statistic to P.
24. Click once on Cell A13 and type "Is t significant?"
25. We will now relate the t-statistic to a P-value using Excel. The format of the command is technically

$$=TDIST(t\text{-statistic}, df, 2)$$

where df is the degrees of freedom, which is the sample size minus two

Two is a two-tailed distribution (you will learn more about this in upper-level classes).

26. Click once on Cell B13 and type

$$=TDIST(B12, 6, 2)$$

27. Hit "Enter."

28. Calculate the P-value for all relationships and format the decimal to three places. If your p value is reported by Excel as a zero, you should present it in your lab report as " $p < 0.001$ " because often Excel is simply rounding off a very small number.

10				
11	Pearson's r	-0.951	-0.956	-0.956
12	Turning r into t	7.5	8.0	10.0
13	Is t significant?	=TDIST(B12, 6, 2)		
14				

FIGURE 5.16

29. Your table should look something like Figure 5.17.

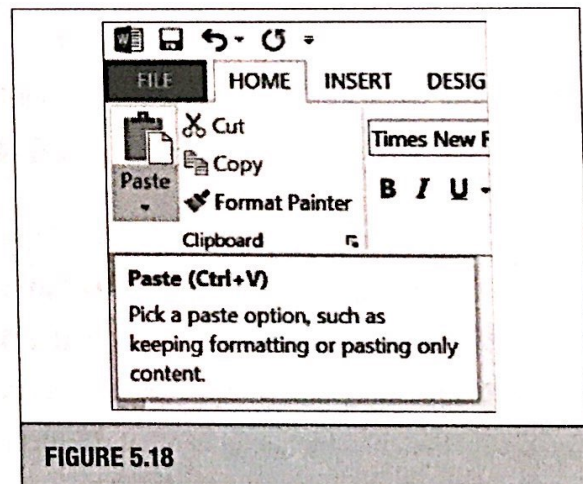
	A	B	C	D	E	F	G	H	I	J	K
1	Generation	Black	White	Red	Green	Blue	Yellow	Orange	Purple	Pink	Brown
2		0	10	10	10	10	10	10	10	10	10
3		1	10	10	8	9	12	10	12	10	9
4		2	7	6	7	7	20	7	25	7	7
5		3	3	5	5	5	22	5	30	12	5
6		4	3	5	3	5	24	3	34	10	3
7		5	0	0	0	2	29	10	38	10	1
8		6	0	0	0	3	30	5	50	2	2
9		7	0	0	0	1	35	2	60	1	0
10											
11	Pearson's r	-0.951	-0.956	-0.97	-0.972	0.99	-0.664	0.985322	-0.683	-0.97	-0.589
12	Turning r into t	7.5	8	10.6	10.1	14.3	2.2	14.1	2.3	10.7	1.8
13	Is t significant?	0.000	0.000	0.000	0.000	0.000	0.073	0.000	0.062	0.000	0.125
14											

FIGURE 5.17

30. Fill in the Columns 4 and 5 in Table 5.1 using your calculated P-value and the information in 14 a through c above.

INSTRUCTIONS—COPY AND PASTING YOUR FIGURE INTO A WORD DOCUMENT

1. Open a new Word document.
2. Click in the white space at the bottom left of the figure.
3. Hit "CTRL C" to copy the figure.
4. Click the "Home" tab.
5. Click the down arrow of the "Paste" icon.
6. Click "Paste Special"
7. Paste a JPEG Picture of your Figure.



INSTRUCTIONS—FIGURE TITLE

1. Click **below** your figure to type in the figure title.
2. In your figure title, answer the questions what, who, where, and when. The figure along with the figure title should stand alone. I should know what the figure is visually representing by its looks and description.
3. Follow the format below.

Figure 1. What who where when.