**Leaf cutter ant foraging – post-lab assessment**

***Learning objective 1: Students will be able to interpret linear regression analyses (including R2, p-values, hypothesis testing, slopes/y-intercepts)***

The two figures below represent data from an experiment where fish were prodded to swim for different amounts of time and the investigators measured how far they swam during that time. The two figures represent two different species of fish. Examine the figures and answer questions 1–5.

1. Which figure would you expect to have a higher R2?
   1. The one on the right
   2. **The one on the left**
   3. They should have the same R2
   4. Neither figure has an R2
   5. I don’t know

**Explanation: B is correct because the figure on the right quite clearly has much more variation in the data around the regression then does the figure on the left. Thus, the figure on the left will have a higher R2.**

1. Which statement most accurately describes the figure on the left based on the information you have?
   1. **As fish swim for longer, they also swim farther**
   2. When fish are prodded to swim, they swim as far as they can and then stop
   3. Some fish swim very far even when prodded to swim for a very short time
   4. There is no relationship between the length of time a fish swam and the distance it swam
   5. I don’t know

**Explanation: A is the most correct description of the data in the figure on the right. A clear positive relationship can be seen between the independent variable (time a fish spent swimming) and the dependent variable (how far a fish swam). B is possibly correct, but it is not based on any information students have been given. C more accurately describes the figure on the right. D is clearly incorrect.**

1. What is the independent variable in these figures?
   1. **Amount of time a fish swam**
   2. Number of times a fish was prodded
   3. Distance that a fish swam
   4. Number of fish used in the experiment
   5. I don’t know

**Explanation: A is the independent variable as visible in the figure and also based on the introductory information given. B and D are data that are not represented in this figure. C is the dependent variable.**

1. If you were told that the p-value for the figure on the right was 0.102, what would that mean?
   1. There was a significant relationship between the amount of time a fish swam and the distance it swam
   2. **There was not a significant relationship between the amount of time a fish swam and the distance it swam**
   3. For every second that a fish swims, it travels 0.102 cm greater distance
   4. The regression line explains 10.2 % of the variance in the data
   5. I don’t know

**Explanation: B is correct because a p-value greater than 0.05 indicates a lack of a significant relationship between the two variables. A is the opposite. C is a confusion of the slope and D is a confusion of R2.**

1. If the slope of the regression in the left figure is 0.5 and the intercept is 2.5, about how far would you expect a fish to swim if it swam for 7 seconds?
   1. 4.5 cm
   2. 2.5 cm
   3. **6 cm**
   4. 10 cm
   5. I don’t know

**Explanation: C is correct and requires students to do some very basic math. The equation for the regression is given as y = 0.5\*x + 2.5. Students should be able to calculate that 7 \* 0.5 = 3.5. Added to the intercept, this means a fish would swim 6 cm. A and B are too small, and D is far too large.**

The figure to the right represents data on the relationship between the length of time tadpoles spent in the water and their size at metamorphosis (when they grow arms and crawl on to land). Answer questions 6–8 about the figure.

1. Which of the following is a likely slope for the regression line shown in the figure?
   1. 1.0
   2. 6.4
   3. 0.0004
   4. **0.1**
   5. I don’t know

**Explanation: This seems difficult but is not too hard. Students should be able to see that the regression line is from 20–40 on x-axis (20 days total) and moves 2 g on the y-axis, from 1 g to 3 g. A rise-over-run of 2 over 20 gives a slope of 0.1. All other values are clearly incorrect.**

1. Which is the dependent variable?
   1. Length of the larval period
   2. Number of tadpoles in the experiment
   3. **Size at metamorphosis**
   4. Slope of the regression
   5. I don’t know

**Explanation: C is clearly the dependent variable, or the y-axis. A is the independent variable. B is not related and D is clearly incorrect.**

1. What would it mean if the R2 for the regression was 0.32?
   1. The regression line passes through 32 different data points
   2. **The regression explains 32% of the variation in the data**
   3. The regression explains 0.32% of the variation in the data
   4. For every day that a tadpoles spends in the water, it grows 0.32 grams
   5. I don’t know

**Explanation: B is the correct interpretation of what R2 means. A, C and D are red herrings. C in particular tests if students understand the difference between proportions and percentages. D is an interpretation of the slope.**

Read the following scenario and answer questions 9–11.

“A scientist comes to you looking for an volunteer to help analyze some data. She conducted research in the Amazon studying the number of piranhas that were swimming at different times of day, trying to understand patterns of activity in these predatory fish. The data were collected by taking still photographs and video underwater of piranhas swimming at three locations: 1) at the base of a dock near a village, 2) in the middle of the river, and 3) where a tributary joins the main river. The photos and videos were taken every 2 hours throughout the day.”

***Learning objective 2: Students will use image analysis software to generate data from an image set.***

1. Which of the following are NOT data you could collect from the images or videos?
   1. Average size of piranhas swimming in different locations
   2. Average number of piranhas active at different times of day
   3. Average of size of piranhas swimming in different locations at different times of day
   4. **Average level of hunger of piranhas at different times of day**
   5. I don’t know

**Explanation: D is the correct as it is the only type of data that you could not retrieve from still images or videos. A, B and C could all be obtained easily.**

1. You are using the program ImageJ to measure average length of piranhas seen at different times. A 20 cm bar was placed in the view of the camera for a scale. Which of the following steps is necessary for every picture you analyze.
   1. Adjust the contrast so you can more clearly see every fish
   2. Zooming in and out to see a particular fish better
   3. **Measuring the 20 cm bar for scale**
   4. Exporting the measurements to MS Excel to analyze the data
   5. I don’t know

**Explanation: C is the only step that would have to be done in every single image. A, B and D are useful but are not necessary for every image analyzed.**

***Learning objective 3: Students will be able to generate meaningful scientific hypotheses given a context for study.***

1. Imagine that you have worked hard to analyze the still images and have found that the average number of fish active in the morning is greatest by the tributary but is greatest in the afternoon in the middle of the river. Which of the following is the most plausible hypothesis to explain these results?
2. **Piranhas sleep in the tributary, safe from their larger predators, but forage in the main river during the day**
3. Humans fish from boats in the afternoon and the piranhas are trying to eat the humans
4. Piranhas spend the afternoon in the middle of the river because it is easier for swimming
5. Piranhas like to spend the morning in the tributary because there are trees that have fruits that monkeys like to eat and occasionally a monkey drops a fruit into the river that the piranhas can eat
6. I don’t know

**Explanation: A is the most plausible hypothesis. B is incorrect because it is not plausible that piranhas would be hunting humans in boats. C and D are incorrect because neither provides an explanation for the migration from one site to another.**

The figure to the right shows the results of your measurements of piranha size at the three different sampling locations. Use the figure to answer questions 12–15.

1. Which of the following hypotheses is the most plausible explanation of these data?
2. There is less food in the middle of the river than near the tributary
3. Piranhas in the tributary and the middle of the river are picked on by the piranhas from the dock and so are smaller
4. Piranhas near the dock are not actually bigger, but the measurements taken from images were in error
5. **Piranhas get extra food from the village by staying near the dock and so grow larger**
6. I don’t know

**Explanation: D is correct because the fact that piranhas are largest by the dock is best explained by a proximity to the human village. A is not related to the larger size of fish by the dock, B is nonsensical, and for C there is no reason to suspect more measurement error for one site only.**

***Advanced learning objective: ANOVA, analysis of categorical variables.***

1. Which of the following p-values would be most likely for the data in the above figure?
2. 1.3
3. **0.03**
4. 0.64
5. 30
6. I don’t know

**Explanation: B is correct the figure clearly shows a significant difference in size between the Dock site and two other sites, and 0.03 is the only p-value that is below 0.05.**

1. Which of the following is a correct conclusion from these data?
2. The size of piranhas differs in all three different sampling locations
3. Piranhas in the tributary are the largest fish in the river
4. We cannot determine anything conclusive about the sizes of fish from these data
5. **Piranhas that live in the middle of the river and the tributary are not significantly different in size**
6. I don’t know

**Explanation: D is correct. Even if students have not learned about ANOVA, they should be able to tell that sizes of the fish in the middle and tributary sites do not differ, but they both differ from fish at the dock site. A implies that they all differ from one another, which is wrong. B is clearly wrong and C is too conservative.**

1. In general, image analysis refers to:
2. Determining the meaning of images
3. Using software like Photoshop to alter images
4. Producing quality graphs and figures that describe analysis of scientific information
5. Obtaining data from images
6. I don’t know
7. It is clear to me that mathematics and statistics are a critical part of the scientific process for biologists.
8. Strongly disagree
9. Disagree
10. Agree
11. Strongly agree
12. I would have been able to analyze and quantify variation in ant morphology or movement without mathematics and statistics.
13. Strongly disagree
14. Disagree
15. Agree
16. Strongly agree
17. Learning about statistics was more interesting as part of the leafcutter ant lab than it would have been otherwise.
18. Strongly disagree
19. Disagree
20. Agree
21. Strongly agree
22. Learning about statistics was more interesting because of the background information and videos about ants.
23. Strongly disagree
24. Disagree
25. Agree
26. Strongly agree
27. In general, I am interested to learn to use mathematics and statistics in biology.
28. Strongly disagree
29. Disagree
30. Agree
31. Strongly agree
32. I feel confident that I can learn mathematics and statistics.
33. Strongly disagree
34. Disagree
35. Agree
36. Strongly agree
37. I am more interested to learn to use mathematics and statistics in biology after completing the leafcutter ant image analysis lab.
38. Strongly disagree
39. Disagree
40. Agree
41. Strongly agree