**Breast cancer – post-lab assessment**

***Learning objective 1: Students will be able to construct and interpret frequency histograms.***

Imagine that you are working at a pond at night, studying the species of frogs that call at the pond and where in the pond you find them calling. You are particularly focused on two species. You walk a transect from one edge of the pond directly through the middle to the other side (20 meters away) and you record how many individuals of each species you find calling each ½ meter. The numbers of frogs you found are shown in the figure below. Use the figure to answer questions1–2.

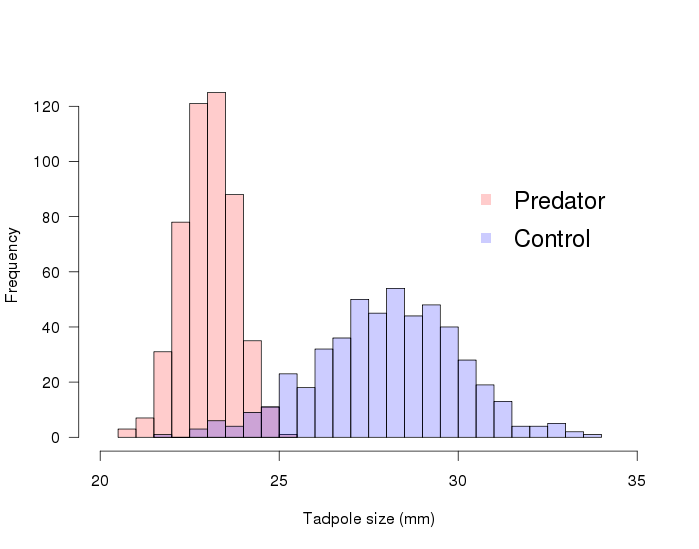
1. Which is NOT a correct statement about the figure?
   1. Species A and Species B do not call in the same parts of the pond
   2. Species A primarily calls from around the edge of the pond
   3. **Species A and Species B call in the same part of the pond**
   4. Species B primarily calls in the center of the pond
   5. I don’t know

**Explanation: This question (the following one) are designed to directly test students on the way that histograms were used in the module. Here, C is the correct answer because it is the only one that is not true. The frogs are occupying different parts of the pond (answer A), Species A is clearly at the two edges (answer B) and Species B is clearly calling in the middle (answer D).**

1. Species A exhibits what kind of distribution?
2. Normal
3. Unimodal
4. **Bimodal**
5. Trimodal
6. I don’t know.

**Explanation: C is correct because Species A clearly shows a bimodal distribution when taking a linear transect through the pond.**

The figure to the right shows frequency histograms of size for two groups of tadpoles, raised either with predators or in a predator-free control environment. Use the figure to answer questions 3–5.

1. Which of the following statements is true?
   1. Tadpoles raised without predators are always larger than tadpoles raised with predators
   2. **Tadpoles raised with predators have less variation in size than tadpoles raised without predators**
   3. Tadpoles raised with predators have more variation in size than tadpoles raised without predators
   4. Tadpoles raised with predators are always larger than tadpoles raised without predators
   5. I don’t know

**Explanation: This question is focused on understanding the variation in a histogram. B is correct because the control treatment clearly has a wider spread of data. A ignores the overlap in the two histograms, C is backwards, and D is backwards and ignores the overlap.**

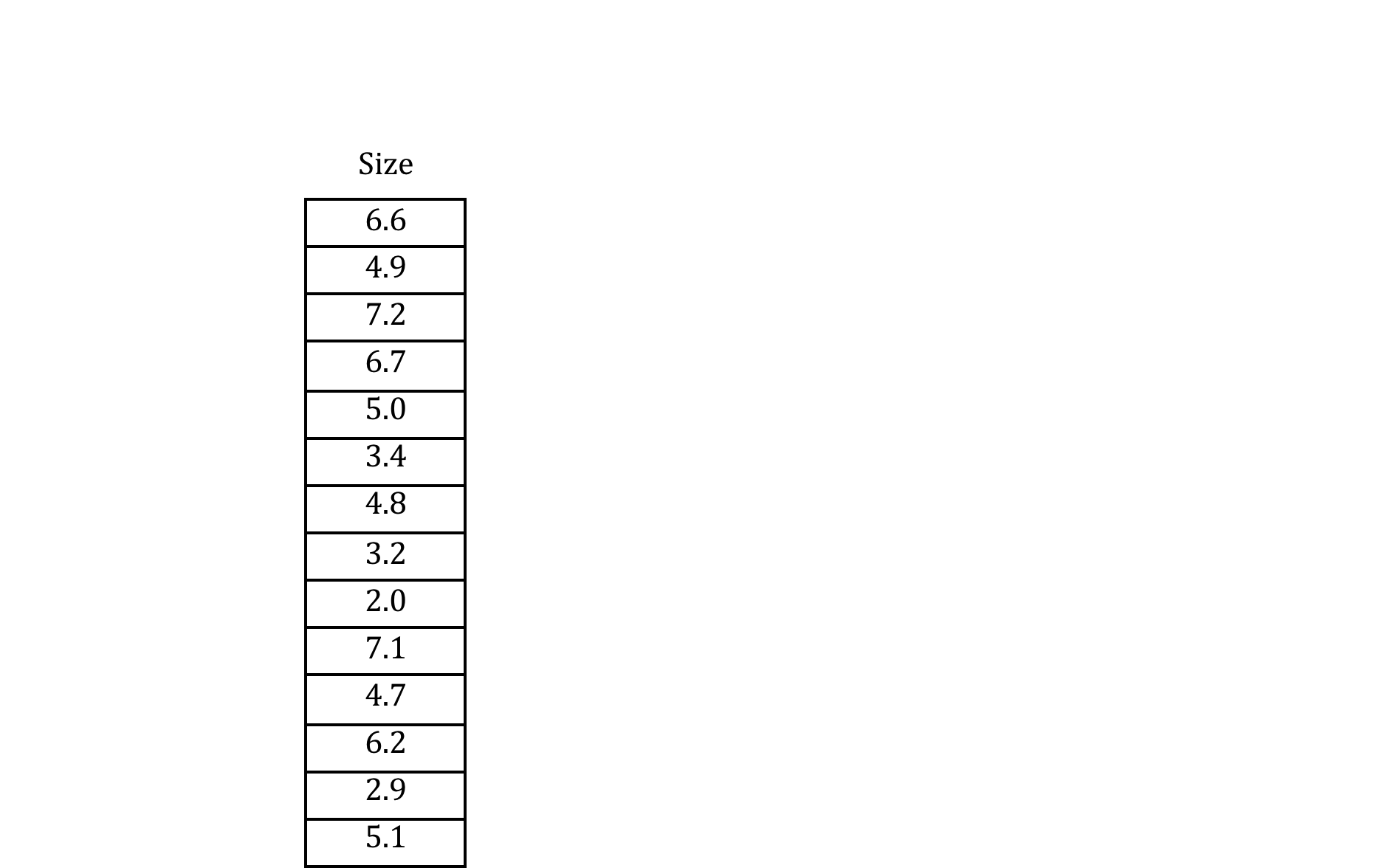
1. Which of the following statements is true?
2. Tadpoles raised without predators are always larger than tadpoles raised with predators
3. **Tadpoles raised with predators are on average smaller than tadpoles raised without predators**
4. Tadpoles raised with predators live longer than tadpoles raised without predators
5. Tadpoles raised with predators are on average larger than tadpoles raised without predators
6. I don’t know

**Explanation: As opposed to question 3, this question is focused on understanding something about the mean from viewing a histogram. B is correct because it states that control tadpoles are on average, not always, larger than tadpoles with predators. A is too absolute, D is backwards and C is not related to the histogram at all.**

1. You randomly selected one tadpole and its size was 24.0 mm. Which of the following statements is true?
2. **The tadpole is most likely from the predator treatment**
3. The tadpole is most likely from the control treatment
4. The tadpole is equally likely from either treatment
5. There is no way to know what treatment it came from
6. I don’t know

**Explanation: A is correct because although 24 mm is within the range of both frequency distributions, the height of the bar for the predator treatment is much higher, indicating a higher probability. B, C and D are just plain wrong.**

1. For the following dataset, which frequency histogram is most accurate?



**Explanation: The correct answer is A. B has the same spread of values in the histogram, but the values are too large to match the table. C has the same mean but too much variation, so the values shown all fall within figure C, but so would many other values and so it is not the most likely. D is far too large.**

***Learning objective 2: Students will be able to use sums of squares to describe differences.***

1. Imagine you are using image analysis to compare two images by comparing the brightness of each pixel in each image. By squaring the difference in brightness between each pair of pixels in each image, and then summing all of those values, you can create a measure of similarity between the two images. Which of the following sums of squares would indicate the two images are most similar?
2. **13**
3. -6
4. 134
5. 10,032
6. I don’t know

**Explanation: The question is designed to provide a clever student with enough information that they could reason the correct answer even if they do not remember anything about sum of squares. However, it does not provide so much information as to give the answer away. The correct answer is A because the sums of squares that indicates the greatest similarity will be the smallest value that is not negative. B is negative. C is slightly larger but not as small as A. D is designed as a red-herring in case students might think that sum of squares should be maximized.**

***Learning objective 3: Students will gain appreciation for the utility of mathematics and computation for biology and medicine***

1. Which of the following is the BESTreason to use a computer program to automate image analysis?
2. Computers are more accurate than humans
3. **It can help save time when you have a lot of images to analyze**
4. Only computers can measure images with precision
5. It is not worth it because humans still have to check the computer’s results by hand
6. I don’t know

**Explanation: The question is designed to make students think about the advantages but also limitations of using computers to assist with the automation of analyses. B is correct because it is the most conservative in its representation of the utility of computer. A is too concrete, C is inaccurate and D is too pessimistic.**

1. What is the centroid of a 2-dimensional shape?
   1. Any point that does not lie on the perimeter of the shape
   2. A point that is on the edge of the shape
   3. The point furthest from the edge of the shape
   4. **The point that is the average distance from the perimeter of the shape**
   5. I don’t know

**Explanation: D is the correct definition of the centroid of a 2D shape. This question is designed to test their retention of the concept of the centroid from the module. A-C are incorrect answers.**

1. Imagine a medical test is said to be “98% sensitive.” What does this mean?
2. 98% of people that test positive have the condition
3. **98% of people with the condition will test positive**
4. 2% of people with the condition will test positive
5. 98% of people with the condition will test negative
6. I don’t know

**Explanation: B is the correct definition of sensitivity, as given in the module. This is a difficult question because the distinction between sensitivity and specificity are nuanced and may be lost on some students. A is incorrect because in the measure of sensitivity, everyone that tests positive is a True Positive, and therefore 100% of people testing positive have the condition. C is the opposite of the true meaning. D is just plain wrong.**

***Learning objective 4: Students will gain appreciation for the interplay between modeling and empirical work***

1. You are studying the development of the brain in a species of ant. You have preserved brains of ants at different stages of development and sliced them exceptionally thin. Using a microscope, you have taken pictures of different parts of the brain and reconstructed them in three dimensions. What might this 3D reconstruction reveal that you could not see from the essentially 2D slices?
2. That the brain has multiple lobes and folds
3. **The volume of different parts of the brain**
4. The function of different parts of the brain
5. How different parts of the brain are connected to one another
6. I don’t know

**Explanation: While modeling can be very informative, there are many things it cannot tell us. Here, B is the correct answer because volume is a 3-dimensional property which cannot be measured from 2-dimensional slices. A would be known from the slides, as would D (at least in the physical sense of proximity). C cannot be known from either slides or the 3D reconstruction.**

1. Which of the following is NOT something that computer modeling could be used for in medicine or other fields of biology?
2. Computers (and digital images) can be used to measure things too small to measure by hand
3. Computers can analyze far more data, and much faster, than can be done by hand
4. **Computers can tell you who will get diseases**
5. Computers can help us to accurately estimate disease risk
6. I don’t know

**Explanation: Again, computer modeling can tell us a lot, but C is wrong because no amount of modeling can tell us who will or will not get diseases. The answer is far too absolute. A, B and D are all things that computers and modeling can and do assist with.**

***General questions to assess student attitude towards the module and understanding of what they have done.***

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 and medical researchers.
8. Strongly disagree
9. Disagree
10. Agree
11. Strongly agree
12. I would have been able to analyze and quantify variation in breast duct morphology 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 breast cancer 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 about breast cancer and the research who discovered the different shapes of tissues.
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 breast cancer image analysis lab.
38. Strongly disagree
39. Disagree
40. Agree
41. Strongly agree