| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 1 [FIGURE]. As a fisheries scientist, you tested whether giving extra food to young salmon would make a significant difference to their growth after a 6-month period. You tried 4 different food types (worms, slugs, larvae, water fleas) and did the experiment twice (using different fish in trial 1 and trial 2). In each trial you had one group of salmon that you fed **normal** amounts of food, and another group of salmon that you fed **extra food**.<br>The figure below shows average (mean) growth after 6 months and a measure of variation around the averages (95% confidence intervals).<br>Which food would you give to salmon in extra amounts to enhance their growth? | a. Worms | 95% confidence intervals do not provide information about statistical significance (averages do) |
| | b. Slugs | 95% confidence intervals do not provide information about statistical significance, AND/OR averages must be similar between trials for results to be significant |
| | c. Larvae | 95% confidence intervals do not provide information about statistical significance, AND/OR overlapping intervals mean the results are "consistently significant" between trials |
| | d. Water Fleas | CORRECT ANSWER – non-overlapping error bars in both trials show significant differences between groups |
| **Question** | **Answers** | **Associated Misconception(s)** |
| 2. You next wanted to know whether living in fish tanks of different depths would affect the growth of young salmon. You compared the average (mean) growth of 100 similar young fish in individual 1 m, 2 m, 3 m and 4 m-deep tanks after they had lived in them for 6 months. You did this experiment twice (using different fish in trial one and trial two), and did **not** find any significant differences in salmon growth.<br><br>What can you say with **confidence**? | a. Growth would have differed if the tank depths were more different from each other | Increasing magnitude of differences between treatment groups will always produce significant differences |
| | b. Growth was affected by depth but differences in the salmon stopped you detecting it | Biological variation in similar individuals will always account for non-significance in results |
| | c. You would have been more likely to see significant differences if you had measured fewer salmon | A smaller sample size should decrease the limits of 95% confidence intervals, AND/OR smaller sample sizes capture more information about a population |
| | d. The growth of young salmon might not be affected by tank depth | CORRECT ANSWER – non-significant differences mean the manipulated factor might not affect the response variable |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 3 [FIGURE]. You went on to test whether water temperature affected the swimming speed (cm/s) of young salmon. This time you tested 70 young fish in each of three temperature treatments (5°C, 10°C, 15°C). After calculating averages (means) and a measure of variation around these averages (95% confidence intervals), you drew the figure below.<br><br>Does water temperature have a statistically significant effect on (mean) swimming speed? | a. Yes because mean swimming speed increases as temperature treatments increase | 95% confidence intervals do not provide information about statistical significance |
| | b. Yes because mean swimming speed is significantly greater in one of the treatments than in the others | CORRECT ANSWER – Non-overlapping 95% confidence intervals between at least two groups suggests the manipulated factor had a significant effect on the response variable |
| | c. No because mean swimming speed is not significantly different between all treatments | 95% confidence intervals must be non-overlapping when comparing ALL treatment groups for there to be a significant effect |
| | d. No because mean swimming speed does not increase by the same amount between treatments | 95% confidence intervals do not provide information about statistical significance AND/OR an effect must be linear to be statistically significant |
| **Question** | **Answers** | **Associated Misconception(s)** |
| 4 [FIGURE]. Another scientist asked if the weight of young salmon was affected by their sex (male/female). He weighed young salmon that had been kept in identical conditions and that were born from parents kept in the same conditions. He followed the same experimental procedure when repeating the study for a second time, <u>but he used a different sample size in the two trials.</u> He calculated averages (means) and a measure of variation around these averages (95% confidence intervals) and drew the figure below. Which of the following statements about sample size is most likely to be true? | a. A very similar number of young salmon were weighed in both trials | The limits of 95% confidence intervals would be expected to vary considerably even if sample sizes are similar between trials |
| | b. A lot more young salmon were weighed in Trial 1 | 95% confidence intervals would be expected to increase with a larger sample size of measurements |
| | c. A lot more young salmon were weighed in Trial 2 | CORRECT ANSWER – 95% confidence intervals would be expected to decrease with a larger sample size |
| | d. More young male salmon were weighed in each trial | 95% confidence intervals would not be expected to vary with sample size, AND/OR averages are expected to increase as sample size increases |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 5 [FIGURE]. You designed an experiment to test whether giving young raccoons extra food would affect their growth after 6 months; raccoons were either placed in a group fed a normal amount of food, or in a group that was fed extra food. You performed the experiment twice (trial 1 and trial 2, using different raccoons in each trial), and used the results to produce the figure below, which shows average (mean) growth and a measure of variation around these averages (95% confidence intervals). In this experiment you chose to formulate one null hypothesis ($H_0$) - the amount of food eaten has no effect on raccoon growth, and one alternate hypothesis ($H_A$) - the amount of food eaten has an effect on raccoon growth. Which of the following is the most accurate interpretation of your results? | a. Fail to reject null hypothesis | A null hypothesis can never be rejected because the alternate cannot be proved |
| | b. Reject null hypothesis | CORRECT ANSWER – statistically significant differences (repeated) mean the null hypothesis should be rejected |
| | c. Prove alternate hypothesis | Statistically significant differences mean the alternate hypothesis has been proved correct |
| | d. Support alternate hypothesis in trial 1 and prove it in trial 2 | Repeating a statistically significant result means the alternate hypothesis has been proved correct, whereas it is initially just supported as being correct |
| **Question** | **Answers** | **Associated Misconception(s)** |
| 6. You were interested in whether squirrels preferred a certain type of food when they had the choice between nuts, seeds and fruit. You found that they chose nuts significantly more often than the other two food types, but chose to perform a second trial (an exact repeat of the experiment, with different squirrels).<br>Complete the following sentence: It is **most** important to do the second trial before concluding that squirrels prefer nuts because you could… | a. Be more confident if they choose them significantly more often again in trial two | CORRECT ANSWER – Statistically significant results must be repeated before you can be confident in the observed relationship |
| | b. Combine results from trial one and two to do a large analysis instead of two small ones | Combining data from separate trials does not introduce bias (experiments carried out under slightly different conditions) |
| | c. Alter your design a little bit to make trial two more accurate | Following different procedures does not invalidate a study's repeatability |
| | d. Use different equipment to be sure that trial one measurements were unbiased | Using different equipment reduces bias, or it proves that there was none in the original trial if similar results are achieved with different equipment in trial two |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 7. [FIGURE] You were interested in whether the amount of time spent sleeping by squirrels and raccoons was affected by temperature. You **predicted** that both species would sleep longer (more hours) in colder temperatures. You placed individuals of each species in a temperature treatment group (4°C, 12°C, 20°C) and counted the number of hours spent sleeping in a 24-hour period. You then calculated averages (means) and a measure of the variation around these averages (95% confidence intervals) and drew the figures below.<br><br>On which species' sleep patterns did temperature have a statistically significant effect? | a. Squirrels only | 95% confidence intervals must be non-overlapping when comparing ALL treatment groups for there to be a significant effect |
| | b. Raccoons only | 95% confidence intervals do not provide information about statistical significance, AND/OR the observed effect must match the prediction to be statistically significant |
| | c. Squirrels and raccoons | CORRECT ANSWER - Non-overlapping 95% confidence intervals between at least two groups for both animals suggests the manipulated factor had a significant effect on the response variable in both |
| | d. Neither squirrels nor raccoons | 95% confidence intervals must be non-overlapping when comparing ALL treatment groups for there to be a significant effect AND the observed effect must match the prediction to be statistically significant |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 8 [TABLE]. You decided to perform a similar sleeping experiment, testing only raccoons. After doing it once (trial one), you decided to do it again (trial two) with different raccoons. This time, however, you used a **much larger** sample size of raccoons.<br><br>Which option in the following table shows how the larger sample size would be expected to affect your results? | a. Answer A | Increasing sample size would be expected to increase the limits of 95% confidence intervals |
| | b. Answer B | Increasing sample size would be expected to increase the limits of 95% confidence intervals AND decrease the range of measurements in a given sample |
| | c. Answer C | CORRECT ANSWER – increasing sample size would be expected to decrease the limits of 95% confidence intervals, but increase the range of measurements in a given sample |
| | d. Answer D | Increasing sample size would be expected to decrease the range of measurements in a given sample |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| 9. You tested whether temperature affected the growth of sunflowers over a 6-month period after germination. You tested 24 plants in each of three treatments (12°C, 13°C, 14°C). You recorded the height (cm) of each plant 6 months after germination, and then calculated averages (means) and a measure of variation around the averages (95% confidence intervals).<br><br>At 12°C, plants grew 114 ± 3 cm, at 13°C they grew 122 ± 4 cm, and at 14°C they grew 120 ± 8 cm. Which of the following is the **most** accurate interpretation of these results? | a. Temperature did affect growth rate because plants grew significantly taller at 13°C than at 12°C | CORRECT ANSWER - Non-overlapping 95% confidence intervals between at least two groups suggests there was a significant effect of the manipulated factor on the measured response variable |
| | b. Temperature did not affect growth rate because plants grew tallest in the middle temperature (13°C) | A significant effect is always unidirectional |
| | c. Temperature did affect growth rate because the average (mean) growth was different in each of the three temperatures | 95% confidence intervals do not provide information about statistical significance; averages do, and as long as they are different, this is significant |
| | d. Temperature did not affect growth rate because plants grown in the coolest (12°C) and hottest (14°C) temperatures did not grow significantly differently | Overlapping 95% confidence intervals for the extreme (most different) treatment groups (e.g. hottest and coldest) means the manipulated factor cannot have a significant effect on the measured response variable |
| **Question** | **Answers** | **Associated Misconception(s)** |
| 10 [FIGURE]. You wanted to know whether using horse manure (natural fertilizer) on sunflowers would make them grow taller (50 plants in one group were given manure, 50 plants in another group were not). All plants were otherwise kept in the same conditions for six months before you measured their height (cm). You then repeated the experiment later in the | a. Yes because plants given manure grew significantly taller than those that were not in both trials | CORRECT ANSWER – Non-overlapping 95% confidence intervals between groups compared in the same trial indicate significant differences |
| | b. Yes because plants given manure grew taller by the same amount than those that were not | 95% confidence intervals do not provide information about statistical significance; averages do, AND these averages must be |

| | | |
|---|---|---|
| season (trial two) with different sunflowers.<br><br>The figure below shows the average (mean) height and a measure of variation (95% confidence intervals) around the average for plants in each group. Did manure help the plants grow taller? | in both trials<br><br>c. No because plants given manure in trial two grew the same amount as those that were not in trial one<br><br>d. No because plants given manure in trial one grew a lot taller than those given manure in trial two | similar in terms of their difference between trials for differences to be significant<br><br>Overlapping 95% confidence intervals between groups compared from different trials indicate non-significant differences<br><br>Dissimilar averages between groups compared from different trials indicate non-significant differences |
| **Question** | **Answers** | **Associated Misconception(s)** |
| 11. [FIGURE] The figure below shows the results of four research teams that tested a different type of manure to see if it could make sunflowers grow taller after a 6-month period; they either grew plants in a normal group (no manure), or in a group that was fertilized with manure. Bars show average (mean) height, lines through bars show a measure of variation around averages (95% confidence intervals).<br><br>One team used a **different number** of plants in their normal group than they did in their manure group. Which team was **most likely** to have done this? | a. Team 1<br><br>b. Team 2<br><br>c. Team 3<br><br>d. Team 4 | CORRECT ANSWER – increasing sample size would be expected to decrease the limits of 95% confidence intervals, so different sample sizes would be expected to cause different limits to 95% confidence intervals when two trials are compared<br><br>Altering sample size would be expected to have no effect on the limits of 95% confidence intervals, AND/OR altering sample size would necessarily be expected to increase the chance of seeing significant differences<br><br>Altering sample size would be expected to have no effect on the limits of 95% confidence intervals, AND/OR altering sample size would necessarily be expected to decrease the chance of seeing significant differences<br><br>Altering sample size would be expected to |

| Question | Answers | Associated Misconception(s) |
|---|---|---|
| | | have no effect on the limits of 95% confidence intervals, AND/OR it would be expected to necessarily provide averages for treatment groups that are more similar than if sample sizes were the same |
| 12 [FIGURE]. You collected seeds from three different strains (types) of sunflowers. Each strain is adapted to different natural conditions but you predicted that plants of all three strains would grow taller when given more water in a controlled setting. Your hypothesis ($H_A$) was that giving different amounts of water to each would affect their height.<br><br>The figures below show averages (means) and a measure of variation (95% confidence intervals) around these averages for height after 16 weeks. Which of the sunflower strains produced results that support your hypothesis ($H_A$)? | a. Strain 1 | 95% confidence intervals must be non-overlapping when comparing ALL treatment groups for there to be a significant effect, AND/OR an effect has to match the prediction to be considered significant |
| | b. Strains 1 and 2 | 95% confidence intervals do not provide information about statistical significance (averages do), AND an effect has to match the prediction to be considered significant |
| | c. Strains 1 and 3 | CORRECT ANSWER - Non-overlapping 95% confidence intervals between at least two treatment groups suggests the manipulated factor had a significant effect on the response variable |
| | d. Strains 1, 2 and 3 | 95% confidence intervals do not provide information about statistical significance (averages do), AND as long as these averages are different, there is a significant effect |