**Quiz 6, Version 1**

**Scenario 1**: Since 2006, millions of bats in North American have died due to infection with a fungus that causes a disease called white-nose syndrome. In the laboratory, researchers found that the fungus grows faster in a petri dish if the petri dish is incubated at a relatively warm temperature. Inside caves, bats hibernate in different spots, and temperatures vary amongst hibernation spots. Therefore, researchers hypothesize that fungus growth rates on bats would **increase** with hibernation temperature. The researchers visit 20 caves twice per year: once at the very start of hibernation (November) and once at the very end of hibernation (March). They randomly select 20 little brown bats per cave and mark them with metal forearm bands that have an individual identification number. For each bat, the researchers measure the temperature in the spot where the bat is hibernating and how much fungus is on the bat in November and March. By comparing the amount of fungus between November and March, they can calculate how much the fungus grew on the bat during hibernation.

Q1. Which of these hypotheses are the researchers testing in Scenario 1?

1. H0: βtemperature = 0

Ha: βtemperature > 0

1. H0: βtemperature = 0

Ha: βtemperature ≠ 0

1. H0: βtemperature = 0

Ha: βtemperature < 0

1. H0: βtemperature = 5

Ha: βtemperature ≠ 5

Q2. After they complete their study in Scenario 1, the researchers run a regression analysis, where they find that the intercept is 0 and the slope is 0.3. Which of these is their regression equation?

1. yi = 0 + 0.3\*Temperaturei + εi
2. yi = 0.3 + 0\*Temperaturei + εi
3. yi = 0\*0.3 + Temperaturei + εi
4. yi = 0\*0.3\*Temperaturei + εi

Q3. The regression analysis gives a p value of 0.03 for βtemperature. Using α=0.05, which of these conclusions is **not** correct?

1. Because their p value is small, they fail to reject the null hypothesis.
2. The probability that they could have collected their dataset if the null hypothesis was true is small (the p value is small).
3. They can infer that the fungus grows faster on bats in warm spots.
4. Because their p value is less than 0.05, they should reject the null hypothesis.

Q4. The regression analysis gives a correlation coefficient (r=0.9) and a coefficient of determination (R2 = 0.81). Which of these statements is **not** correct?

1. The slope of the regression line is 0.9.
2. There is a strong correlation between fungus growth rates and the temperatures where bats hibernate.
3. There is a positive correlation between fungus growth rates and the temperatures where bats hibernate.
4. 81% of variation in fungus growth rate can be explained by variation in the temperature where bats hibernate.

Q5. The researchers submit their study for publication in a science journal, and a peer reviewer is worried that the bats in each cave might **not** be **independent** experimental units, which would cause a pseudoreplication issue in the study design. Which of these statements is **false**?

1. The researchers could solve this pseudoreplication issue by re-doing their study in the same 20 caves, but this time increasing the number of replicates in each cave (e.g., n=100 bats in each of 20 caves).
2. If bats within a cave are independent experimental units, the total number of replicates is 400 bats (20 bats x 20 caves).
3. If bats within a cave are not independent experimental units, the total number of replicates is more like 20 bats (1 replicate per cave \* 20 caves), and that sample size is not big enough to meet the Law of Large Numbers requirement for many statistical analyses.
4. Bats within a cave might not be independent experimental units because something that happens inside a cave during hibernation could affect all of the bats at once; for instance, if cavers visit a cave during hibernation, every bat in the cave might become stressed out and more susceptible to disease.

Q6. Why didn’t the researchers use an ANOVA hypothesis test to analyze their data?

1. The researchers could not use an ANOVA because their predictor variable was quantitative.
2. The researchers could not use an ANOVA because their response variable was quantitative.
3. The researchers could not use an ANOVA because their response variable was continuous.
4. The researchers could not use an ANOVA because their predictor variable was categorical.

**Quiz 6, Version 2**

**Scenario 2**: Thousands of people attempt to thru-hike the entire Appalachian Trail each year, and a few hundred of those people make it all the way from Georgia to Maine. Some people finish the whole thru-hike quickly; the speed record is 41 days! Other people hike more slowly; the slowest record is 329 days, which is almost a whole year! The Appalachian Trail Conservancy wants to know if the number of days that it takes to do the whole trail depends on when a hiker starts. For instance, starting in March instead of May might help people to avoid the Georgia heat and hike faster. The Appalachian Trail Conservancy collects data from all thru-hikers on when they started and ended. Then they run an ANOVA analysis to see if the average time to complete the trail depends on whether hikers start in March, April, May, or June.

Q1. Which of these hypotheses is the Appalachian Trail Conservancy testing in Scenario 2?

1. H0: The mean time to complete the trail is the same whether hikers start in March, April, May, or June.

Ha: The mean time to complete the trail is the longer or shorter during at least one month.

1. H0: The mean time to complete the trail varies by month.

Ha: The mean time to complete the trail is the same whether hikers start in March, April, May, or June.

1. H0: The mean time to complete the trail varies by month.

Ha: The mean time to complete the trail increases with month.

1. H0: µtraillength = 3 months

Ha: µtraillength > 3 months

Q2. The ANOVA analysis gives a p value of 0.000001. Using α=0.05, which of these inferences can we make?

1. The average amount of time that it takes to complete the thru-hike depends on which month the hikers start in.
2. It takes hikers longer to complete the thru-hike if they start in April than if they start in June.
3. It takes hikers longer to complete the thru-hike if they start in June than if they start in April.
4. It takes hikers the same amount of time to complete the thru-hike, no matter which month they start in.

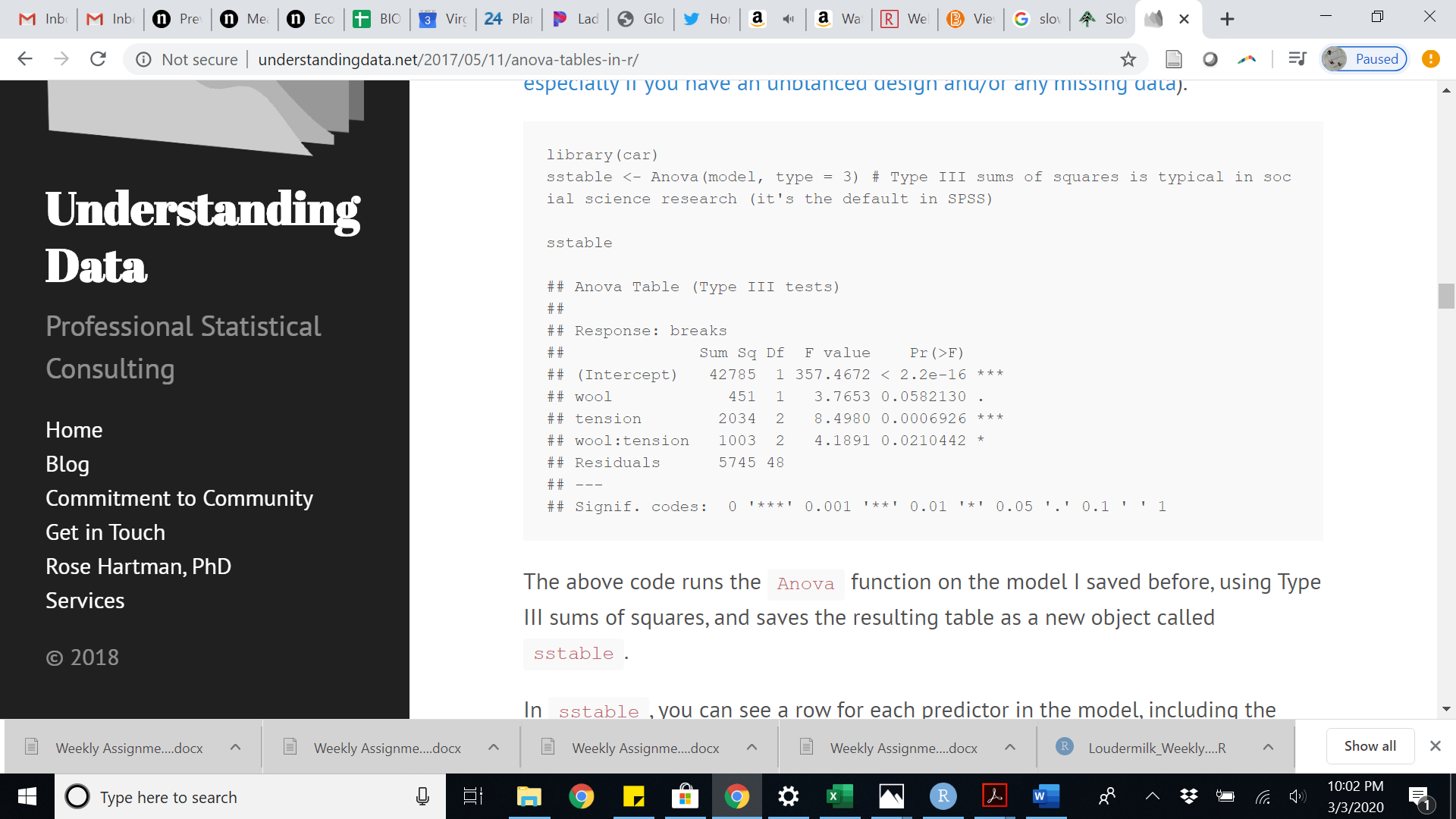
Q3. What type of study and statistics does Scenario 2 represent?

1. This is an observational study and the researchers are using inferential statistics to draw conclusions about a population.
2. This is a manipulative experiment and the researchers are using inferential statistics to draw conclusions about a population.
3. This is a manipulative experiment and the researchers are using descriptive statistics to describe a population.
4. This is an observational study and the researchers are using descriptive statistics to describe a population.

Q4. Which of these statements is **incorrect**?

1. The researchers should have used a classical regression, because their response variable was categorical.
2. The researchers used an ANOVA because their response variable was quantitative.
3. The researchers used an ANOVA because their predictor variable was categorical.
4. If the researchers only compared hikers who started in March or May (comparing just two months), they would have used a t-test instead of an ANOVA.

**Scenario 3. Here is the output from an ANOVA analysis. Use it to answer Question 5 and Question 6.**



Q5. What is the p value for the main effect of wool?

1. 0.058
2. 0.0006926
3. 451
4. 4.1891

6. Which of these conclusions is **correct**?

1. There is no main effect of wool, but there is a significant main effect of tension, and it depends on the interaction between wool and tension.
2. There are no significant effects of wool or tension.
3. Tension decreases the effect of wool.
4. The intercept effect is bigger than the wool effect.