CLT1 Recitation Worksheet Answer Key

CLT Week 1: Why is there a "Climate Debate?" How to interpret data.

1. Correlation always implies causation. True/false? Discuss with your group and explain why by giving a few NEW examples.

False

Correlation (from the lecture notes):

- how closely two sets of information or data are related
- a third variable that correlates with the two sets exits

Causation (from the lecture notes):

- the act which produces cause and effect
- Once correlation is established, further studies using scientific method are needed to make a causal claim

Two data sets can be correlated but does not mean that one is causing the other.

Examples (given in the lecture):

- Ice cream consumption vs. drowning death -- positively (and strongly) correlated but no causation
- More such examples from here: http://www.fastcodesign.com/3030529/infographic-of-the-day/hilarious-graphs-prove-that-correlation-isnt-causation

2. The number of certain penguin species has been monitored for the last 30 years to study the effects of climate change. During this time, the temperature anomalies were recorded in the area. Data below shows the change in surface temperature and recorded number of penguins living in the habitat.

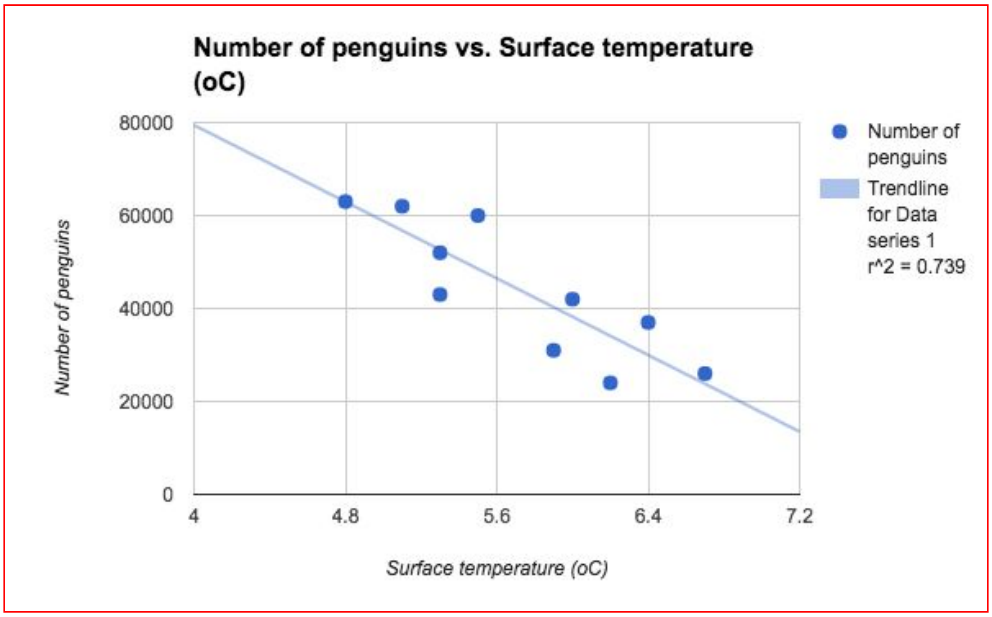
Time (years)	Surface temperature (°C)	Number of penguins
1	4.8	63000
4	5.1	62000
8	5.3	52000
12	5.3	43000
15	5.5	60000
19	6.0	42000
21	5.9	31000
24	6.2	24000
28	6.4	37000
30	6.7	26000

a. The number 0.071 written in scientific notation is 7.1E-2. Now you write down the number of penguins at time 21 years in scientific notation. **Recall that when entering numbers on SUCourse quizzes, this notation will always be true;** you must use it especially for entering very large or very small numbers**.** ANS: 3.1E4

b. Go to your gmail account (=sabanciuniv email account). Click on the sign  that appears in the top right corner of your account, and select "Sheets." Start a Blank new spreadsheet.

Enter the data above to the spreadsheet; plot the number of penguins as a function of surface temperature by selecting the relevant data followed by clicking **Insert→Chart, Chart type→Scatter**. Look at the data and predict if the correlation is negative or positive. Explain why you think so.

Negative because as one parameter increases, the other decreases.



c. Now click on the plot to open chart editor, and select **Customize** (or Customization). Go to **Series** section, then select **trendline**. Fit a linear line to your data. Then, select use equation as the label. What kind of information does that equation gives you?

The equation written on the plot is the mathematical formulation of the linear fit line, and from this equation we can predict the number of penguins at any given temperature with the assumption that the data follows this linear trend at any temperature range.

d. Predict the number of penguins at a surface temperature of 7°C. 17615

e. Is there a causation between the surface temperature and the number of penguins? Why or why not? Write down at least two possible scenarios leading to the observed correlation.

It indicates that as the temperature of the surface rises more penguins die and that there two parameters are *correlated*. Some possible scenarios follow:

- global warming increases surface temperature → penguins are less fit to live in high temperatures so they decrease in numbers (direct causation).
- global warming increases surface temperature AND sea temperature → penguins feed on fish that are less fit to live in high temperatures so the penguins cannot find enough food and decrease in numbers. (indirect causation)
- global warming increases surface temperature → penguin eggs which are laid on land are less fit to survive higher temperatures there so they decrease in numbers. (no causation)
- (this one is ridiculous, but people come up with all sorts of these ideas, so here we go:) penguins are cold blooded; so they may be killed by some other reason, but now because there are less of them, the surface temperature may be higher; i.e. the causality may be in the reverse direction.

All of the above generate testable hypotheses, so the observation of high correlation does not tell us the causality right away, but provides us a starting point. (completely reverse causation)

If you'd like to save the spreadsheet with a particular name, click on "*Untitled spreadsheet*" on the top left corner of your sheet and type the name of the file before you close the sheet. Everything is automatically saved to your google drive.

3. You are a scientist studying the effect of climate change on ticks (parasites that carry the Lyme disease bacteria) by recording the geographic habitat range of ticks in a certain region in Europe throughout 10 years. During this time, you observed that the average temperature in this region increased ~1.5°C. The data below shows the average air temperature (summer) and the range of tick habitat that you measured. You would like to find temperature dependence of tick habitat range.

Temperature (°C)	**Range of Tick Habitat (km^2)**
24.0	250.0
24.2	253.5
24.3	254.2
24.6	256.0
24.6	258.5
24.9	261.8
25.3	263.5
25.1	267.1
25.3	270.4
25.5	272.6

a. Just by looking at the data table, can you tell whether or not the range of tick habitat is correlated with the temperature? Explain why or why not.

b. Using Google Sheets (or Excel), plot the tick habitat range as a function of temperature, and guess a functional form for the correlation (linear, polynomial, exponential or logistic).

c. Now, use 'trendline' to find the equation of the line/curve that fits the data best.

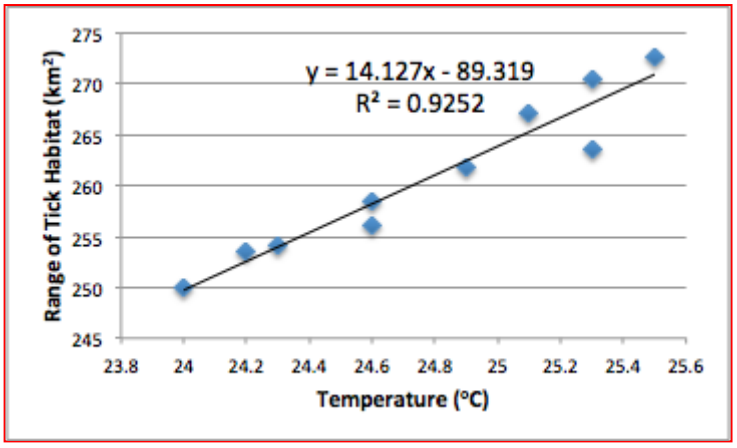
d. Is there causation between the two parameters? Explain.

e. Does this correlation directly mean that the incidence of Lyme disease in this region also depends on the temperature?

Sample Answers:

a. Yes, we can tell that the two parameters are correlated because as one increases and the other also increases overall. However, without plotting and quantifying somehow the correlation, we cannot say much about the degree of the correlation.

b. Linear function probably is sufficient to describe the correlation. Polynomial or exponential function may also fit the data, but definitely not logistic.



c. If linear function is chosen: (Range of habitat) = 14.1T − 89.3, where T is temperature in °C

d. Just from this dataset we cannot say anything for certain but there can be causation between these two parameters in the scenarios such as:

- warmer weather encourages ticks to migrate (move) further
- more ticks may be able to survive warmer winter that they can spread in wider area or once they spread they have higher survival rate
- etc…

e. The habitat range-temperature correlation cannot directly mean that the Lyme disease incidence is also correlated with temperature, but it could suggest the correlation because when the tick habitat area increases, the overall tick population increases and probably the number of ticks that carry the disease increases. Discussion of this sort is expected here.