Correlation

Correlation

- We are in sections 12.2, 12.3 of the textbook
- Measure two factors (variables) from each individual. When a change in one variable is associated with a change in a second variable we call this *correlation*.
  - positive correlation - increase in one is associated with increase in the other
  - negative correlation - increase in one is associated with decrease in the other
- Caution! Correlation does not imply causation. Look for underlying hidden variables that might explain the relationship. Or the two may be completely unrelated.
- The *correlation coefficient*, r measures the strength of a correlation.
  - r close to 1, the association is close to a straight line, the correlation is positive
  - r close to -1, the association is close to a straight line, the correlation is negative
  - r close to 0, no linear correlation, data does not resemble a line
- Caution! r close to 0 means no **linear** correlation. The data may be associated in another shape.

1. For each of the following scenarios, identify a reasonable possibility for a hidden variable.

   (a) In a particular city, they find a strong correlation between the amount of ice cream sold on a given day and the number of speeding tickets given on that day.

   (b) In a particular study, they find that teenagers who eat dinner with both parents at least 5 times per week are less likely to become addicted to drugs.

2. Data on precipitation and temperature from the Huntington Tri-state Airport, WV, Station USW00003860, from 1961 to 2019 is available from NOAA.

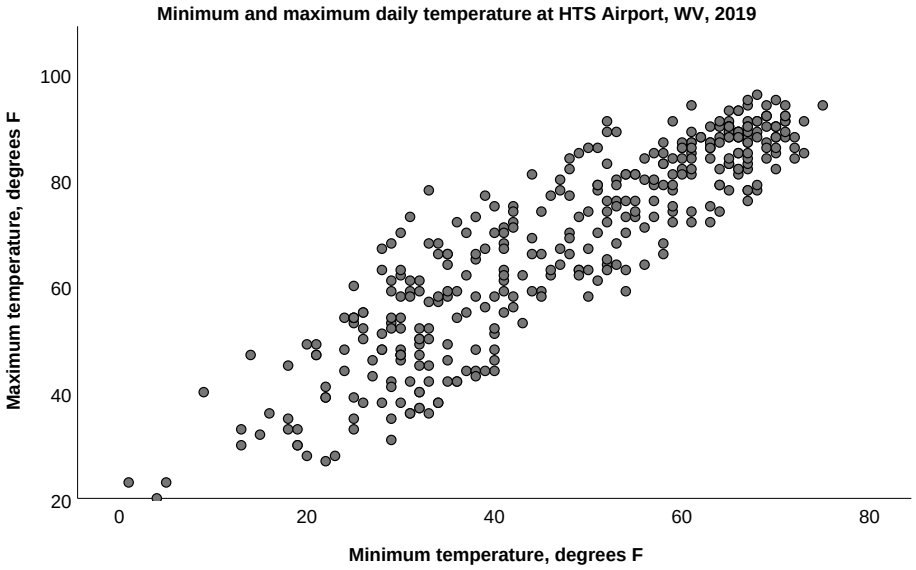
   For both pairs of variables,

   - Do you suspect these variables to be correlated?
   - If you suspect the variables are correlated, describe what you suspect to be the correlation.
   - Do you suspect that the change in one variable causes a change in the other (is there a causal relationship)?
   - If you suspect that one does not cause the change in the other, identify a reasonable possibility for a hidden variable.

   (a) Consider the variables minimum temperature and maximum temperature.

   (b) Consider the variables minimum temperature and precipitation.

3. We're going to focus on year 2019, only. A scatter plot for minimum and maximum temperature is shown.

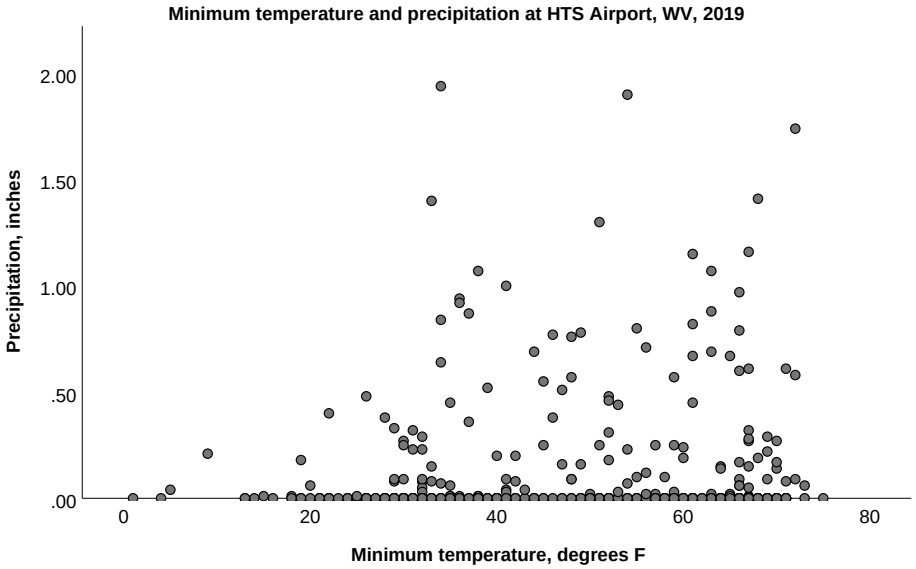


(a) Describe the relationship between the two variables, if any. (What pattern is shown in the scatterplot? Are the variables correlated?)

(b) Does the relationship between the variables match your expectation? Why or why not?

(c) The correlation coefficient between minimum and maximum temperature is $r = 0.899$. Explain what this mean about the two variables and how it matches what you can see in the scatterplot.

4. We're going to focus on year 2019, only. A scatter plot for minimum temperature and precipitation is shown.



(a) Describe the relationship between the two variables, if any. (What pattern is shown in the scatterplot? Are the variables correlated?)

(b) Does the relationship between the variables match your expectation? Why or why not?

(c) The correlation coefficient between minimum temperature and precipitation is $r = 0.098$. Explain what this mean about the two variables and how it matches what you can see in the scatterplot.