

Friday, September 21, 2018

- Warm-up
 - Describe the following association

positive

non linear

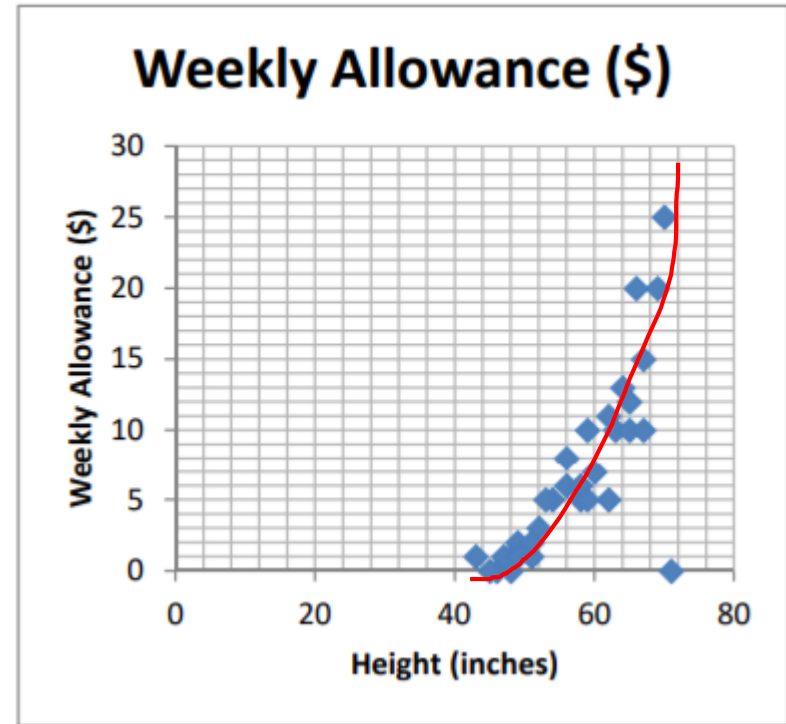
one potential outlier

cluster between 40-60, 60-80

Strong

- Check Homework

- Correlation



Correlation Coefficient (r)

- The **correlation coefficient (r)** gives us a numerical measurement of the strength and direction of the linear relationship between the explanatory and response variables.

$$r = \frac{\sum z_x z_y}{n - 1}$$

- For now calculate with technology

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

ESSENTIAL QUESTION: What is the difference between association and correlation and how is it reflected in a coefficient or scatterplot?

CORRELATION CONDITIONS

- **Correlation** measures the strength of the *linear* association between two *quantitative* variables.
- Before you use correlation, you must check several conditions:
 - **Quantitative Variables Condition**
 - **Straight Enough Condition**
 - **Outlier Condition**

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

ESSENTIAL QUESTION: What is the difference between association and correlation and how is it reflected in a coefficient or scatterplot?

Quantitative Variables Condition:

- Correlation applies only to quantitative variables.
- Don't apply correlation to categorical data masquerading as quantitative.
- Check that you know the variables' units and what they measure.

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

ESSENTIAL QUESTION: What is the difference between association and correlation and how is it reflected in a coefficient or scatterplot?

Straight Enough Condition:

- You can *calculate* a correlation coefficient for any pair of variables.
- But correlation measures the strength only of the *linear* association, and will be misleading if the relationship is not linear.

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

ESSENTIAL QUESTION: What is the difference between association and correlation and how is it reflected in a coefficient or scatterplot?

Outlier Condition

1 time w/outlier remove it 2nd time w/out it

- **Outliers can distort the correlation dramatically.**
- **An outlier can make an otherwise small correlation look big or hide a large correlation.**
- **It can even give an otherwise positive association a negative correlation coefficient (and vice versa).**
- **When you see an outlier, it's often a good idea to report the correlations with and without the point.**



Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

ESSENTIAL QUESTION: What is the difference between association and correlation and how is it reflected in a coefficient or scatterplot?

CORRELATION BASIC FACTS

- Correlation treats x and y symmetrically:
 - The correlation of x with y is the same as the correlation of y with x .

$$r_{\text{handspan} - \text{toutsie pops}} = 0.51$$
$$r_{\text{toutsie pops} - \text{handspan}} = 0.51$$

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

CORRELATION BASIC FACTS

- Correlation has no units.

0.72 ~~in~~

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

CORRELATION BASIC FACTS

- Correlation is not affected by changes in the center or scale of either variable.

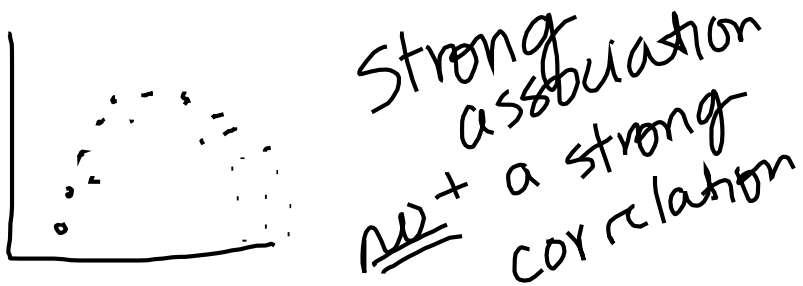
height (in) \uparrow shoe size = 0.7
height (cm) \uparrow shoe size = 0.7 \leftarrow stays the same no matter the units

Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

CORRELATION BASIC FACTS

- Correlation measures the strength of the *linear* association between the two variables.
- Variables can have a strong association but still have a small correlation if the association isn't linear.

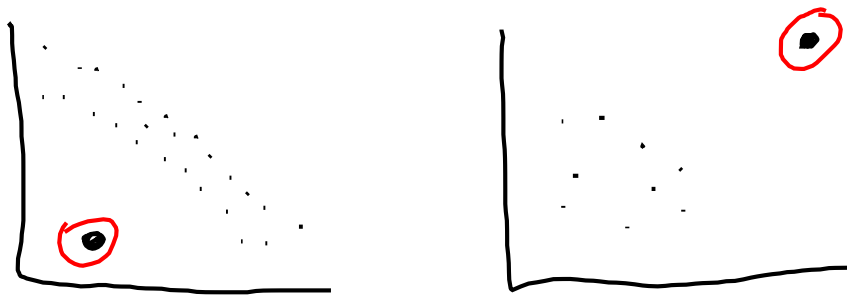


Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

CORRELATION BASIC FACTS

- Correlation is sensitive to outliers. A single outlying value can make a small correlation large or make a large one small.



Content/Language Objective: I will be able to correctly discuss the meaning of association and correlation using a scatterplot and correlation coefficient both in writing and with my group/teacher/class.

Social Objective: I will participate respectfully so that my classmates and I can be part of the lesson.

Check for Understanding...

Temperature and air pollution are known to be correlated. We collect data from two laboratories, in Boston and Montreal. Boston makes their measurements of temperature in Fahrenheit, and Montreal in degrees centigrade. Boston measures pollution in particles per cubic yard of air; Montreal uses cubic meters. Both report a correlation of exactly 0.58 between temperature and pollution. Which of the following is true:

- A. Boston really has the higher correlation, because Fahrenheit temperatures are higher than Centigrade.
- B. Montreal really has the higher correlation, because cubic meters are bigger than cubic yards.
- C. Both cities have the same correlation, because correlation is independent of the units of measurement.
- D. We do not know which city has the really higher correlation.

We measure heights and weights of 100 twenty-year old male college students. Which will have the higher correlation:

- A. $\text{corr}(\text{height}, \text{weight})$ will be much greater than $\text{corr}(\text{weight}, \text{height})$
- B. $\text{corr}(\text{weight}, \text{height})$ will be much greater than $\text{corr}(\text{height}, \text{weight})$
- C. Both will have the same correlation.
- D. Both will be about the same, but $\text{corr}(\text{weight}, \text{height})$ will be a little higher.
- E. Both will be about the same, but $\text{corr}(\text{height}, \text{weight})$ will be a little higher.

The correlation between the ages of the husbands and wives in the United States was which of the following?

A. + 1.0

B. + 0.85

C. zero

D. - 0.85

E. -1.0

Suppose men always married women who were 10 percent shorter than they were. The correlation coefficient of the heights of married couples would be:

- A. 0.10 if the correlation were computed with `corr(male.height, female.height)`
- B. -0.10 if the correlation were computed with `corr(female.height, male.height)`
- C. 0.10 no matter which way the correlation were computed.
- D. 1.0 since the height of the man is always predictable from the height of the woman



*Assignment : p 165
#11, 12, 15, 16*