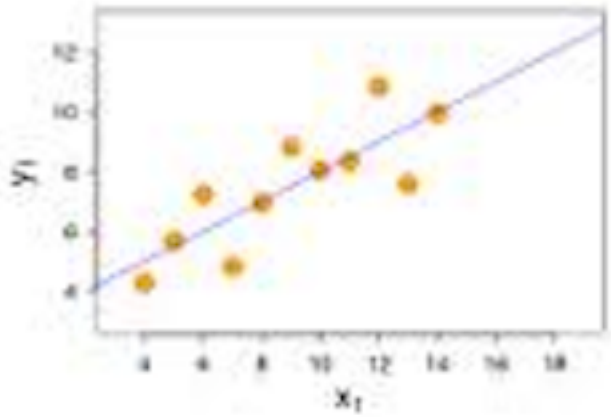
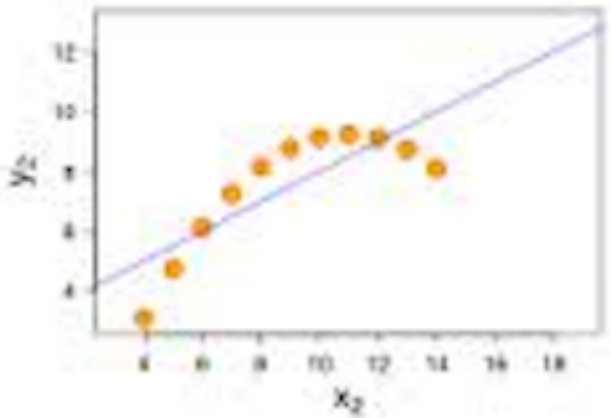
Which is a weak, positive, linear correlation?

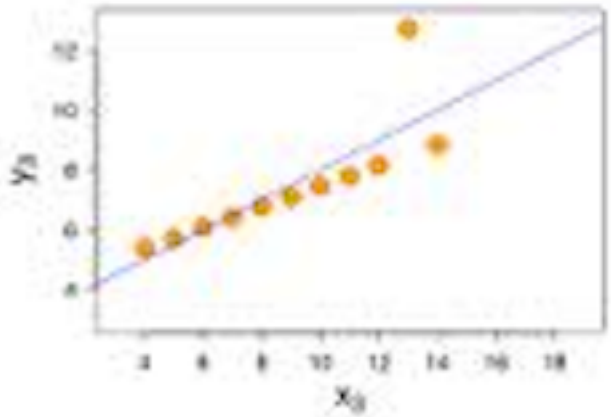
a.



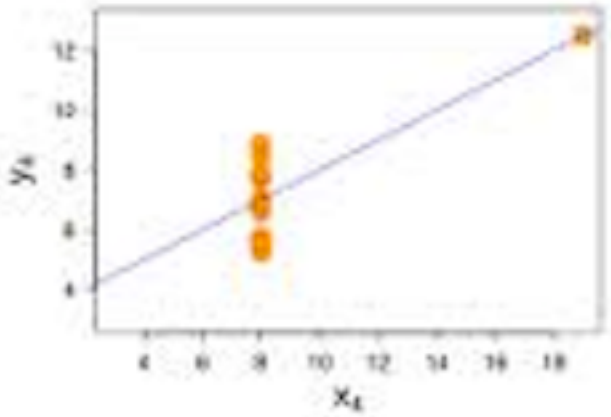
b.



c.

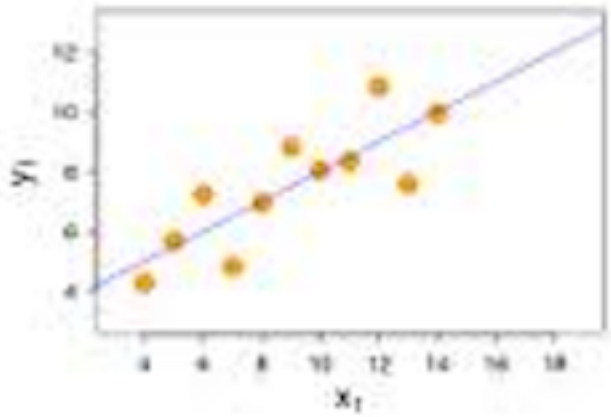


d.

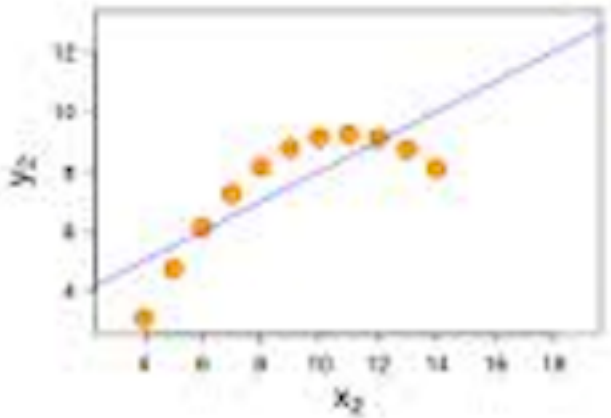


Which has the weakest correlation?

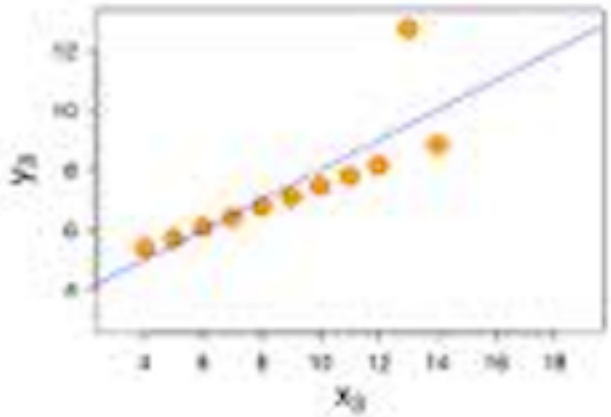
a.



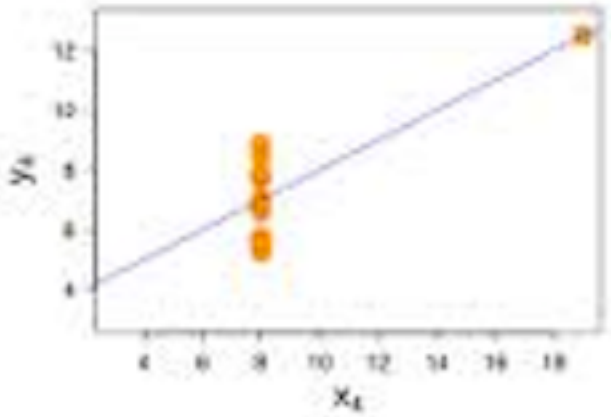
b.



c.



d.



Example: Quiz Grades

You want to minimize the variability in the class (make everyone have close to the same grade).

If you could magically make everyone the same for just ONE factor, which factor would you choose?

A. hours of class/studying
B. taking high school stats
C. years in college

Example: Quiz Grades

Which factor explains the most variability in quiz grades?

A. hours of class/studying
B. taking high school stats
C. years in college

Correlation

Which student would contribute to a **positive correlation**?

A. studied 10 hours **above** M, grade 20 pts **below** M
B. studied 10 hours **above** M, grade is at M
C. studied 10 hours **above** M, grade 20 pts **above** M
D. studied 10 hours **below** M, grade 20 pts **below** M
E. Two of the above

r Values

If in reality there is **no relationship** between two variables, what will the **r value** of a sample be?

A. Exactly zero
B. Close to zero
C. Close to +1
D. Close to -1

Cell Cultures Example

Will the new test with n = 120 be significant?

A. No or unlikely
B. Yes or likely

Cell Cultures Example

The new test of n = 120 for r = 0.2 turns out to be **significant**. Why?

A. In reality, the relationship is now stronger
B. There is more information to confirm the weak relationship
C. The critical value is not as extreme

% of variance

For a quiz, the lowest grade is 22 and the highest is 26 out of 30 points.

Quiz grades and # hours studied have an $r = 0.8$.

How much of the variability in in grades is explained by hours studied?

A. About 1-2 points on the test
B. About 4 points on the test
C. About 11-13 points of the test
D. About 18-21 points of the test

Formula for a line

If you have a **formula for a line**, which can you **NOT** do?

A. predict X from Y
B. predict Y from X
C. know the "exchange rate"
D. know how much Y you have when you have zero X
E. you can do all of these

Fitting a line

You have a bunch of data points with X and Y values.

You want to **fit a line to your data**.

Is there always a line that **goes through all points**?

A. Yes
B. No

Fitting a line

If a line can't go through all points, which is the best option?

A. Go exactly through as many points as possible
B. Get as close to as many points as possibe
C. Don't have any very big errors

Formula for b

We said we wanted a line that minimized large errors (least squares regression line).

If instead we wanted to **go through as many points as possible**, **would the formula for b change**?

A. Yes
B. No

Standard Error of estimate

What does the standard error of the estimate tell us?

A. How much the regression sample data **differs from the population**
B. How much the regression line estimate **differs from actual data**
C. How likely our sample data is **by chance**