

Workshop Statistics: Discovery with Data, Second Edition

Topic 9: Correlation Coefficient

Activity 9-6: Properties of Correlation (*cont.*)

(a) A possible scatterplot:



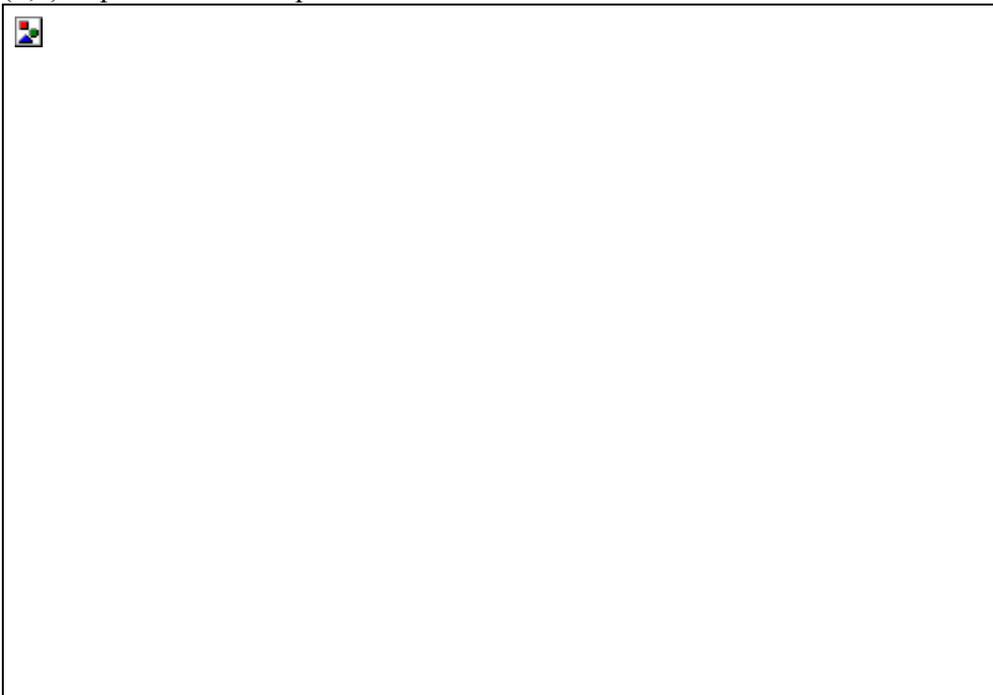
(b) $r = 1$

(c, a) A possible scatterplot:



(c,b) $r = 1$

(d,a) A possible scatterplot:



(d,b) $r = 1$

(e,a) A possible scatterplot:



(e,b) $r = 1$

(f) The degree of the slope evident in a scatterplot does not affect the correlation between the two variables.

Activity 9-7: Properties of Correlation (*cont.*)

- (a) In class A, most of the observations follow a linear pattern, with only one exception.
 - (b) In class B, most of the observations seem to be scattered haphazardly with no apparent pattern.
 - (c) class A: .037; class B: .705; One might expect the more linear pattern of class A to have the higher of the two correlations.
 - (d) It only takes one exception to a linear pattern to throw off what could be a strong correlation. Similarly, a seemingly haphazard pattern could have a fairly strong correlation if a few extreme points line up linearly. Because correlations are easily affected by these outliers, we say that they are not resistant.
 - (e) It appears that if a student score below 50 on the first test, they score below 50 on the second test, although some scores went up and some went down. Similarly, if they scored above 70 on the first test, they still scored above 70 on the second test, but some scores went up and some went down.
 - (f) $r = .954$; Within each group the association was weak, however the groups "lined up" resulting in an inflated overall correlation coefficient. Correlation coefficients must be used in conjunction with scatterplots in order to draw any useful information from them.
- Note: The answers to (e) and (f) above are the solutions for (f) and (g) in the Calculator version.

Activity 9-8: Properties of Correlation (*cont.*)

- (a) Data set A clearly has a positive correlation coefficient because as the variable x increases, so does the variable y .
- (b) Data set C clearly has a negative correlation coefficient because as the variable x increases, the variable y decreases.
- (c) Data set B clearly has a correlation coefficient close to zero because the data as the value of x increase, sometimes the value of y decreases and sometimes the value of y increases. If you sketch a plot of the points, there is no real trend.

Activity 9-9: States' SAT Averages (*cont.*)



- There appears to be a fairly strong negative association between the average SAT score for a state and the percentage of the students of that state who took the test.
- (b) One should not conclude that a cause-and-effect relationship exists between these two variables. There are many other factors that influence average score besides the percentage of students who took the exam, including the quality of the state's educational system, wealth of the state, whether or not the state offers the ACT, etc.

Activity 9-10: College Alumni Donations (*cont.*)

- (a) 1962
- (b) avg gift vs. % give: $r = .553$; avg gift vs. prev avg gift: $r = .390$
- (c)-(d) Answers will vary from student to student.
- (e) avg gift vs. % give: $r = .352$, a decrease; avg gift vs. prev avg gift: $r = .849$, an increase

Activity 9-11: Governors' Salaries (*cont.*)



- (a) There appears to be a weak positive association.
- (b) Answers will vary from student to student.
- (c) .290
- (d) Connecticut, New Jersey, Massachusetts
- (e) Answers will vary from student to student.
- (f) .523
- (g) The correlation went up.

Activity 9-12: Monopoly (*cont.*)

- (a) observational units: properties
- (b) 5



(c)

(d)

	Position	Price	Rent	House
Price	.995			
Rent	.983	.994		
House	.984	.994	.999	
Hotel	.964	.978	.982	.981

(e) Answers will vary from student to student.

(f)

	Position	Price	Rent	House
Price	.997			
Rent	.991	.994		
House	.991	.994	.999	
Hotel	.985	.984	.978	.978

increased: 5; decreased: 2

Note: The answers given for (d)-(f) above pertain to (c)-(e) respectively in the Calculator version.

Activity 9-13: Broadway Shows (*cont.*)

(a)

- A: gross and capacity: .702
- B: top price and capacity: .305
- C: gross and attendance: .966

- D: gross and top price: .574
- E: attendance and top price: .485
- F: attendance and capacity: .846

(b) C F A D E B; This ordering agrees with the ordering based on scatterplots alone in Activity 8-7.

(c) The Lion King

(d)-(f) Answers will vary from student to student.

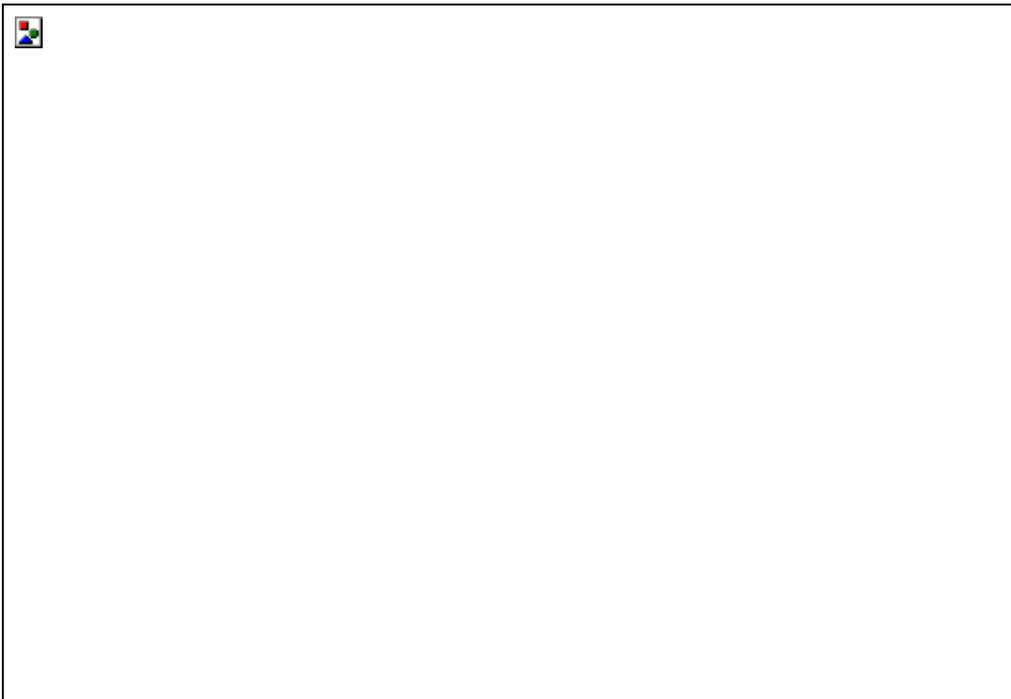
Activity 9-14: Solitaire



$$r = .127$$



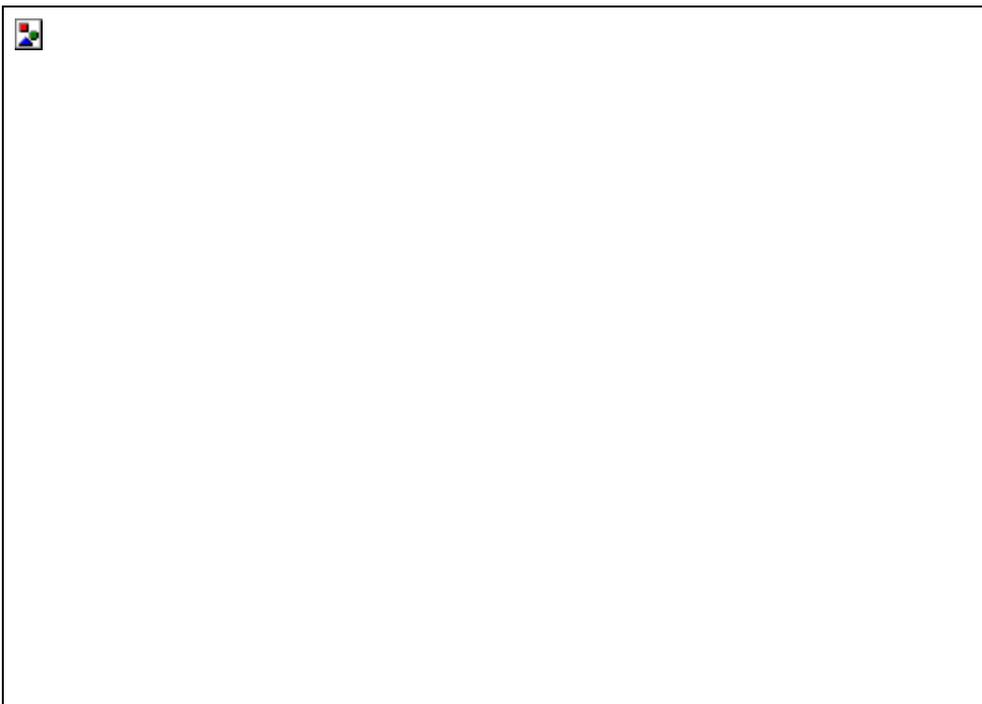
$r = -.053$



$r = .061$



$r = -.493$



$r = .496$



$$r = -.979$$

There is a moderate negative association between win number and time. It may be the case that as the player played more games, he/she got better at the game and tended to be able to win the game in less time. Also, there is a strong negative association between time and points. This indicates that the less time the player took to win the game, the more points he/she was awarded. This is consistent with the rules of the game.

Activity 9-15: Ice Cream, Drownings, and Fire Damage

- (a) No, this does not mean that ice cream causes drowning. More drownings occur during the summer months because more people go swimming. Similarly, more ice cream is purchased during the summer months because it is a cold treat. There is a third variable, time of year or outside temperature, that relates to both variables.
- (b) More fire engines respond to more damaging fires because those fires that are more damaging are larger and more fire engines are needed to extinguish them. This does not mean that the damage would be less extensive if fewer fire engines were dispatched. There is a third variable, severity of the fire, that relates to both variables.

Activity 9-16: Evaluation of Course Effectiveness

The professor's conclusion is not valid. While the high grades in his class did contribute to the high cumulative grade point averages, they were not the only factor. High grades in other classes, the students' study habits, ect., all contributed to the cumulative grade

point average.

Activity 9-17: Climatic Conditions (*cont.*)

(a)

	Jan hi	Jan lo	July hi	July lo	precip	precday	snow	sun
Jan hi	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
Jan lo	.965	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
July hi	.152	-.022	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
July lo	.554	.473	.712	xxxx	xxxx	xxxx	xxxx	xxxx
precip	-.073	.002	.114	.243	xxxx	xxxx	xxxx	xxxx
precday	-.572	-.460	-.130	-.345	.695	xxxx	xxxx	xxxx
snow	-.807	-.825	-.080	-.613	-.157	.444	xxxx	xxxx
sun	.643	.512	.377	.521	-.506	-.826	-.363	xxxx

(b) Jan hi and Jan lo have the strongest association, with a correlation coefficient of .965.

(c) Jan lo and precip have the weakest association, with a correlation coefficient of .002.

(d) Jan lo would be the most useful. July hi would be the least useful.

(e) July lo would be the most useful. Snow would be the least useful.



(f)

There is a weak negative association between annual snowfall and annual precipitation. Perhaps years with more precipitation tend to have less snow and more rain.

Activity 9-18: Space Shuttle O-Ring Failures (*cont.*)

The correlation coefficient between temperature and number of O-ring failures is -.561. When we exclude the flights in which no O-rings failed, the correlation coefficient between temperature and number of O-ring failures becomes -.263. The association is weakened because those flights that were excluded were all on days where the temperature was relatively high. The numbers are no longer present to show that hot days tend to have no O-ring failures.

Activity 9-19: Variables of Personal Interest (*cont.*)

Answers will vary from student to student.