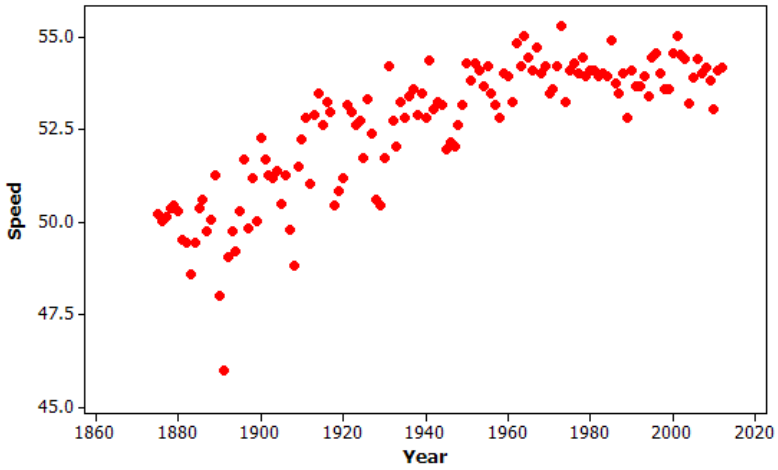
Name _____ Date _____

1. The Kentucky Derby is a horse race held each year. The following scatter plot shows the speed of the winning horse at the Kentucky Derby each year between 1875 and 2012.



a. Is the association between *speed* and *year* positive or negative? Give a possible explanation in the context of this problem for why the association behaves this way considering the variables involved.

b. Comment on whether the association between *speed* and *year* is approximately linear and then explain in the context of this problem why the form of the association (linear or not) makes sense considering the variables involved.

c. Circle an outlier in this scatter plot and explain, in context, how and why the observation is unusual.

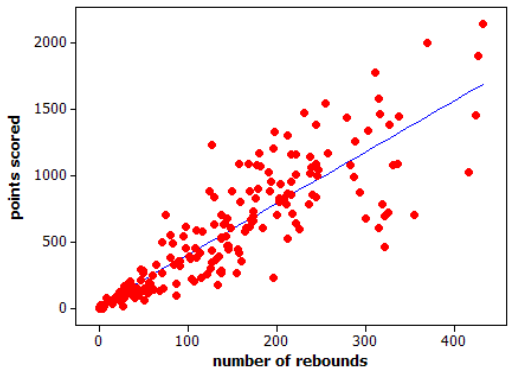
2. Students were asked to report their gender and how many times a day they typically wash their hands. Of the 738 males, 66 said they wash their hands at most once a day, 583 said two to seven times per day, and 89 said eight or more times per day. Of the 204 females, 2 said they wash their hands at most once a day, 160 said two to seven times per day, and 42 said eight or more times per day.

   a. Summarize these data in a two-way table with rows corresponding to the three different frequency-of-hand-washing categories and columns corresponding to gender.

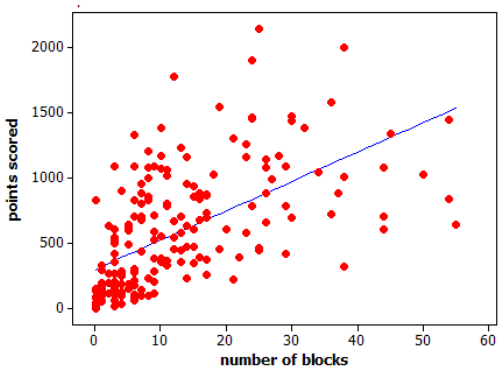
   b. Do these data suggest an association between *gender* and *frequency of hand washing*? Support your answer with appropriate calculations.

3. Basketball players who score a lot of points also tend to be strong in other areas of the game such as number of rebounds, number of blocks, number of steals, and number of assists. Below are scatter plots and linear models for professional NBA (National Basketball Association) players last season.

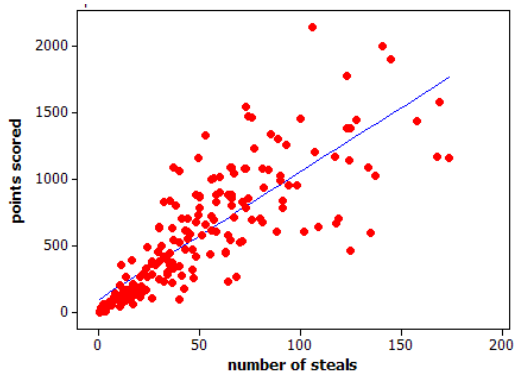
$y = 21.54 + 3.833x$



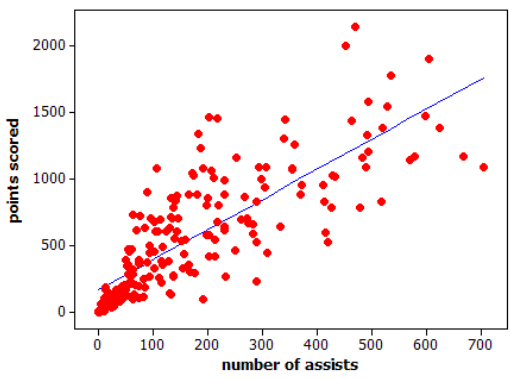
$y = 294.9 + 22.45x$



$y = 98.03 + 9.558x$



$y = 166.2 + 2.256x$



a. The line that models the association between points scored and number of rebounds is $y = 21.54 + 3.833x$, where $y =$ points scored and $x =$ number of rebounds. Give an interpretation, in context, of the slope of this line.

b. The equations above all show $y =$ number of points as a function of the other variables. An increase in which of the variables (rebounds, blocks, steals, and assists) tends to have the largest impact on the predicted points scored by an NBA player?

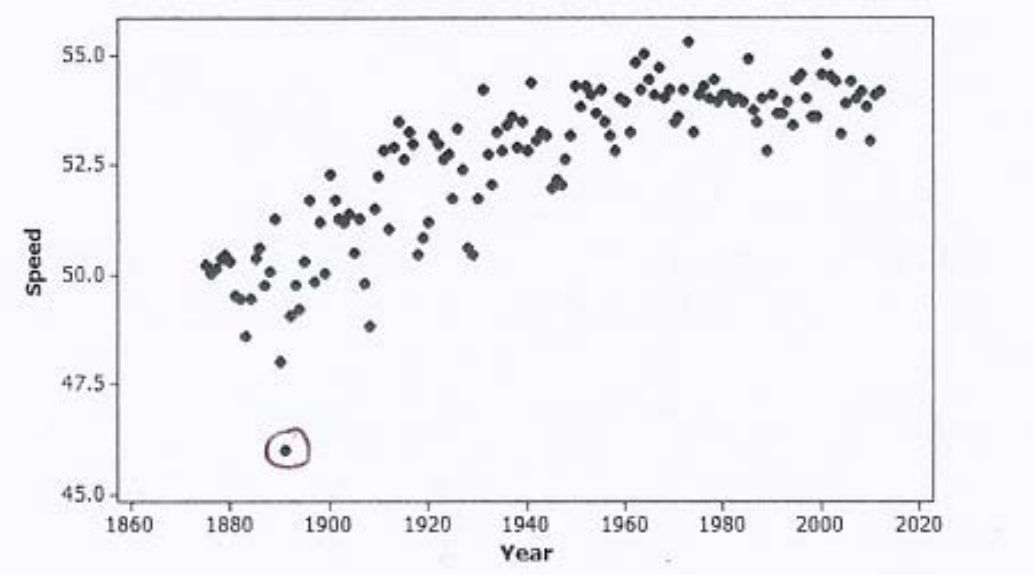
c. Which of the four linear models shown in the scatter plots above has the worst fit to the data? Explain how you know using the data above.

A Progression Toward Mastery					
Assessment Task Item		**STEP 1** Missing or incorrect answer and little evidence of reasoning or application of mathematics to solve the problem.	**STEP 2** Missing or incorrect answer but evidence of some reasoning or application of mathematics to solve the problem.	**STEP 3** A correct answer with some evidence of reasoning or application of mathematics to solve the problem OR an incorrect answer with substantial evidence of solid reasoning or application of mathematics to solve the problem.	**STEP 4** A correct answer supported by substantial evidence of solid reasoning or application of mathematics to solve the problem.
1	a 8.SP.A.1	Student does not use the data in the scatter plot or context to answer question.	Student discusses horses getting faster with newer training methods but does not discuss the data in the scatter plot.	Student discusses the overall increase of speeds but does not discuss how that data implies horses getting faster over time.	Student discusses the overall increase of speeds and how that data implies horses getting faster over time.
	b 8.F.B.5	Student does not use the data in the scatter plot or context to answer question.	Student does not recognize the nonlinear nature of the data.	Student discusses the nonlinear nature of the data but does not relate to the context.	Student discusses the curvature in the data, which indicates that speeds should level off.
	c 8.SP.A.2	Student does not use the data in the scatter plot or context to answer question.	Student picks the year with the fastest or lowest speed of the winning horse and does not explain choice.	Student picks the year with the lowest speed of the winning horse but does not interpret the negative residual.	Student picks the year with the lowest speed of the winning horse and states that the speed is much lower than is expected for that year.
2	a 8.SP.A.4	Student does not use the data given in the stem.	Student gives tallies of the two distributions separately without looking at the cross-tabulation.	Student constructs the table but uses gender as the row variable.	Student constructs a 3×2 two-way table, including appropriate labels.

	b **8.SP.A.4**	Student answer is based only on context without references to data.	Student gives some information about the association but does not back it up numerically. OR Student says the results cannot be compared because the numbers of males and females are not equal.	Student attempts to calculate the six conditional proportions but compares them inappropriately. OR Student does not correctly complete all the calculations.	Student calculates the six conditional proportions, compares them, and draws an appropriate comparison (e.g., 20% of females wash eight or more times compared to 12% of males).
3	**a** **8.F.B.4**	Student cannot identify the slope from the stem.	Student interprets slope incorrectly.	Student interprets slope correctly but not in context or not in terms of model estimation.	Student interprets slope correctly and predicts that on average, for each additional rebound, an increase of 3.833 points is scored.
	b **8.SP.A.3**	Student does not relate to the functions provided above the scatter plots.	Student focuses on the strength of the association in terms of how close the dots fall to the regression line.	Student appears to relate the question to the slope of the equation but cannot make a clear choice of which variable has the largest impact or does not provide a complete justification.	Student relates the question to the slope and identifies number of blocks as the variable with largest impact.
	c **8.SP.A.2**	Student does not use the data in the scatter plots to answer the question.	Student focuses only on the slope of the line or on one or two values that are not well predicted.	Student focuses on the vertical distances of the dots from the line but is not able to make a clear choice.	Student selects number of blocks based on the additional spread of the dots about the regression line in that scatter plot compared to the other variables.

Name _____ Date _____

1. The Kentucky Derby is a horse race held each year. The following scatter plot shows the speed of the winning horse at the Kentucky Derby each year between 1875 and 2012.



a. Is the association between *speed* and *year* positive or negative? Give a possible explanation in the context of this problem for why the association behaves this way considering the variables involved.

The association is positive overall, as horses have been getting faster over time. This is perhaps due to improved training methods.

b. Comment on whether the association between *speed* and *year* is approximately linear and then explain in the context of this problem why the form of the association (linear or not) makes sense considering the variables involved.

The association is not linear. There is probably a physical limit to how fast horses can go that we are approaching.

c. Circle an outlier in this scatter plot and explain, in context, how and why the observation is unusual.

The winner that year was much slower than we would have predicted.

2. Students were asked to report their gender and how many times a day they typically wash their hands. Of the 738 males, 66 said they wash their hands at most once a day, 583 said two to seven times per day, and 89 said eight or more times per day. Of the 204 females, 2 said they wash their hands at most once a day, 160 said two to seven times per day, and 42 said eight or more times per day.

a. Summarize these data in a two-way table with rows corresponding to the three different frequency-of-hand-washing categories and columns corresponding to gender.

	males	females
≤ 1	66	2
2–7	583	160
≥ 8	89	42
	738	204

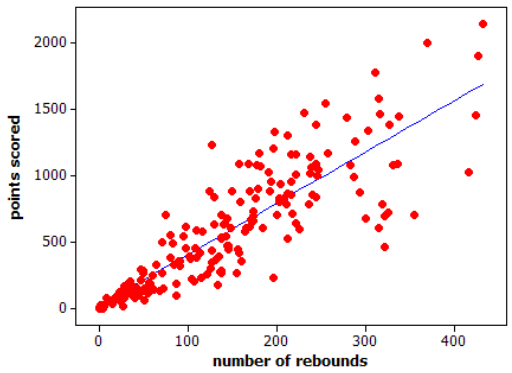
b. Do these data suggest an association between *gender* and *frequency of hand washing*? Support your answer with appropriate calculations.

	males	females
≤ 1	.0894	.0098
2–7	.7900	.7843
≥ 8	.1206	.2059

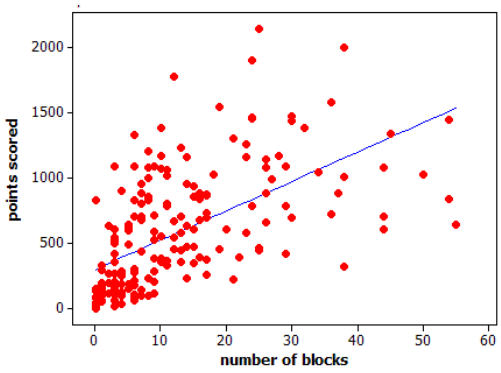
Males are more likely than females to wash hands at most once per day. Females are more likely to wash 8 or more times per day.

3. Basketball players who score a lot of points also tend to be strong in other areas of the game such as number of rebounds, number of blocks, number of steals, and number of assists. Below are scatter plots and linear models for professional NBA (National Basketball Association) players last season.

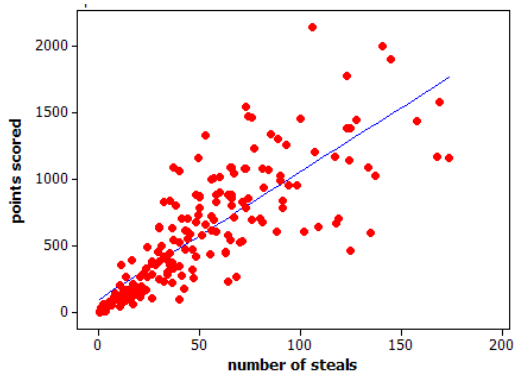
$y = 21.54 + 3.833x$



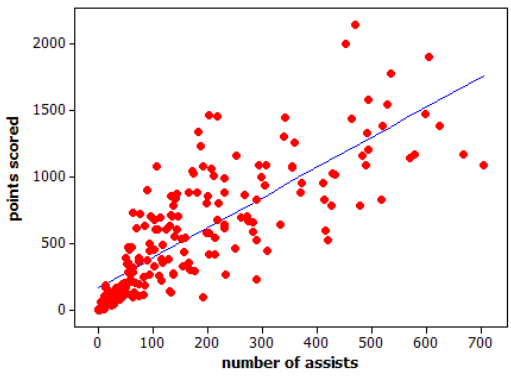
$y = 294.9 + 22.45x$



$y = 98.03 + 9.558x$



$y = 166.2 + 2.256x$



a. The line that models the association between points scored and number of rebounds is $y = 21.54 + 3.833x$, where $y =$ points scored and $x =$ number of rebounds. Give an interpretation, in context, of the slope of this line.

If the number of rebounds increases by one, we predict the number of points increases by 3.833.

b. The equations above all show $y =$ number of points as a function of the other variables. An increase in which of the variables (rebounds, blocks, steals, and assists) tends to have the largest impact on the predicted points scored by an NBA player?

Each additional block corresponds to 22.45 more points, the largest slope or rate of increase.

c. Which of the four linear models shown in the scatter plots above has the worst fit to the data? Explain how you know using the data above..

Probably number of blocks because the association is weaker. There is more scatter of the points away from the line.