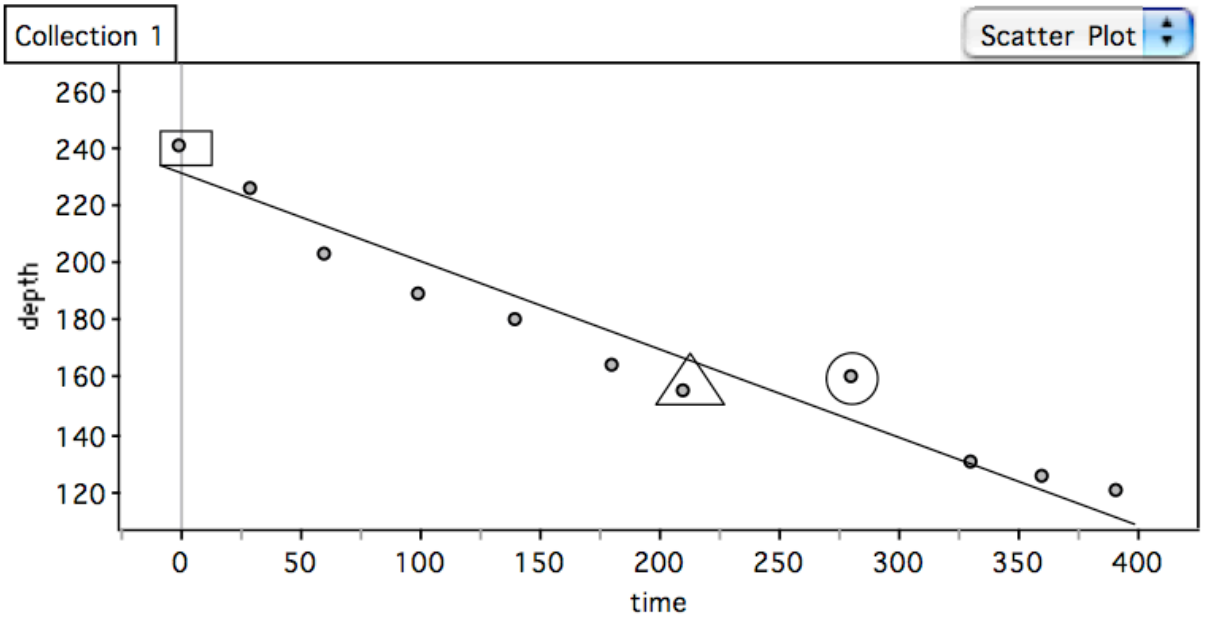
A.P. Statistics Problem for Linear Regression

A diver is investigating a wreck under the water and has to come up to the surface slowly. Following is a chart detailing his depth from the time he starts ascending.

time (min)	0 sec	30 sec.	1 min.	1min. 40 sec	2 min 20 sec	3 min	3 min 30 sec	4 min 40 sec	5 min 30 sec	6 min	6 min 30 sec
depth (ft)	240	225	203	189	180	164	155	160	130	125	120

1. Draw a scatterplot which is suitable for predicting depth based on time. There is one point that lies outside the pattern of the plot. Circle it.



2. Find the equation of the least squares line of the data. Comment on the strength of the association.

$\hat{depth} = -.285\,time + 225.676$

$r = -.973$ - very strong negative association

3. Eliminate the point you circled and recalculate the equation of the least squares line. Sketch the line on the scatterplot above. Comment on the strength of the association compared to your answer with that circled point included.

$r = -.980$ - association stronger

4. Using the model above (point removed), explain the meaning of the slope of the line.

for every second, the diver's depth decreases (therefore he ascends) by .292 feet
note: you can also calculate in terms or minutes ... the diver descends by 17.52 feet per minute

5. Describe the relationship between time and depth using r^2 to make your description more precise.

96% of the variation in depth can be explained by the LSRL

6. Using the model above (point removed), predict the depth of the diver at the following times. Comment on the confidence of your prediction.

a) 2 minute, 50 seconds b) 5 minutes c) 7 minutes, 10 seconds

176.02 ft 138.03 ft 100.05 ft

Since the association is strong, interpolation (a and b) is fairly safe. Extrapolation (c) is riskier but since the 7 minutes and 10 seconds is fairly close to the given data, the confidence is still fairly high.

7. Find the difference between the observed depth at 1 min, 40 seconds, and the predicted depth.

$189 - 196.48 = -7.48$

8. Explain why a linear model might not be the best to describe this data.

The residual plot shows a pattern which suggests that the model is curved rather than linear.

9. In the diagram above, place a square around the point which has the largest residual. Place a triangle around the point which has the second largest residual.

10. Using the linear model, how long would your predict before the diver reaches the surface. In reality, do you believe it is a longer or shorter duration? Explain.

772.38 seconds or 12.88 minutes. If the data is truly curved, it may take more time. But the pressure is less when the depth is smaller so the it may take less time. This shows that extrapolation is dangerous because the we really have no idea what is happening as time is larger.

11. What is the exact sum of the residuals?

Zero. The sum of the residuals is always zero. Don't be fooled by the calculator.

12. One more piece of data is added. It is found that the diver is at 40 feet, 7 minutes into his ascent. Explain the role of this new point and how it will affect the slope of the LSRL without actually calculating it.

The new point is highly influential will dramatically decrease the slope of the LSRL.

13. New data is added to change the mean time to mean time to 6 minutes and 10 seconds with standard deviation 2 minutes and 5 seconds. The mean depth is now 115.5 feet with standard deviation 31.6 feet. If the association is still negative and $r^2 = .68$, find the slope of the regression line of time versus depth.

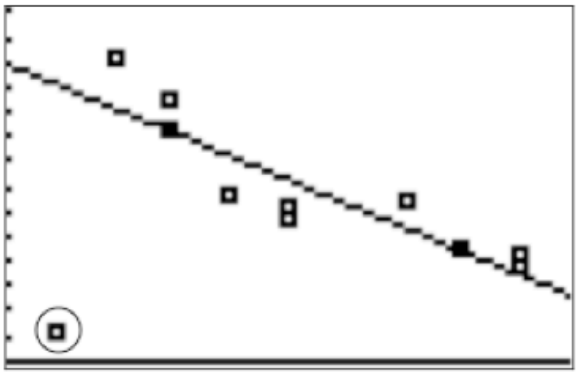
$$b = r\left(\frac{s_y}{s_x}\right) = -\sqrt{.68}\left(\frac{31.6}{125}\right) = -.208\,\frac{\text{ft}}{\text{sec}} = -\sqrt{.68}\left(\frac{31.6}{2.083}\right) = -12.51\,\frac{\text{ft}}{\text{min}}$$

A.P. Statistics – Linear Regression Worksheet - Solutions

The busiest season for Walmart is the Christmas holiday and weekends see a tremendous number of customers. Last year, Walmart conducted a study as to the amount of waiting in time in checkout lanes its customers had to wait. On Saturdays and Sundays of its holiday season, it opened a different number of checkout lanes for customers between 1 PM and 4 PM, its busiest times. The measurement was the average wait time for a customer to go through the lane and complete the transaction. A different number of lanes was opened each day. The data is below.

Date	11/22	11/23	11/29	11/30	12/6	12/7	12/14	12/15	12/21	12/22	12/29	12/30
Weather	Sun	Cloudy	Sun	Rain	Sun	Sun	Cloudy	Cloudy	Cloudy	Rain	Snow	Sun
Lanes	5	12	11	7	12	8	6	10	8	6	4	8
Avg Wait Time (Min)	12.2	4.2	4.4	6.75	3.8	5.75	10.4	6.5	6.25	9.2	1.1	5.6

1. What is the explanatory variable? Lanes What is the response variable? Avg wait.
2. For answer # 1, make a scatterplot on your calculator and draw it below.



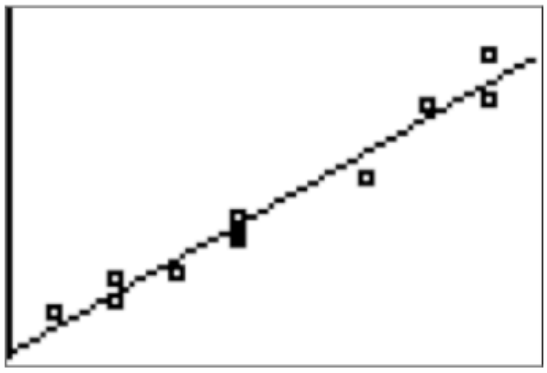
x = lanes, y = avg wait

3. There is a clear outlier on your scatterplot. Circle it.
4. Give a reason that would justify eliminating the outlier. It was snowing. Few customers.
5. Generate the least squares regression line that describes the data (with the outlier eliminated). (2 dec places). $\hat{y} = -0.96x + 14.96$
6. Give the meaning of the slope in the LSRL. For every extra lane open, .96 minute less of a wait.
7. What specific point is the LSRL guaranteed to go through? $(\bar{x}, \bar{y}) = (8.45, 6.82)$
8. Describe the data in terms of form, direction, and strength. Fairly strong negative linear correlation
9. Find the value of r. $r = -.88$. Interpret it. Fairly strong negative linear correlation
10. Find the value of $r^2 = .78$. Interpret it.
    78% of the variation in waiting time can be explained by the LSRL.
11. What is the predicted average waiting time if Walmart opens 9 lanes? 6.32 minutes
12. What is the predicted average waiting time if Walmart opens 7 lanes? 8.24 minutes
13. What is the difference between the actual waiting time and predicted waiting time for 7 lanes? –1.5 min
14. If Walmart decides that it wants its customers to wait no longer than 4.5 minutes, how many lanes should it open? 11 lanes.
15. If Walmart decides that it wants its customers to wait no longer than 10 minutes, how many lanes should it open? 6 lanes
16. What is the predicted average waiting time if Walmart opens only one lane? 14 min.
17. What is the predicted average waiting time if Walmart opens 20 lanes? -4.24 min
18. Why do the last two problems and answers make little sense for this problem?
    Extrapolation is always dangerous. The minimum waiting time must be positive.

On the same days of the previous problem, Walmart will need to hire extra staff to man the checkout lanes. They will have to pay out more money for the workers. The chart below describes how much money in pay it paid out during those hours on these specific days.

Date	11/22	11/23	11/29	11/30	12/6	12/7	12/14	12/15	12/21	12/22	12/29	12/30
Weather	Sun	Cloudy	Sun	Rain	Sun	Sun	Cloudy	Cloudy	Cloudy	Rain	Snow	Sun
Lanes	5	12	11	7	12	8	6	10	8	6	4	8
Avg Wait Time (Min)	12.2	4.2	4.4	6.75	3.8	5.75	10.4	6.5	6.25	9.2	1.1	5.6
Salary for extra staff	$213	$451	$444	$256	$498	$301	$229	$364	$291	$253	xxxx	$320

19. What is the explanatory variable? Lanes What is the response variable Salary.
20. For answer # 19, make a scatterplot on your calculator and draw it below.



x = lanes, y = salary

21. Generate the least squares regression line that describes the data and draw it on the graph above. 2 dec places) $\hat{y} = 38.72x + 1.71$
22. Give the meaning of the slope in the LSRL.
    For every extra lane opened up, Walmart pays $38.72 in extra salary
23. What specific point is the LSRL guaranteed to go through? $(\bar{x}, \bar{y}) = (8.45, 329.09)$
24. Describe the data in terms of form, direction, and strength. Strong positive linear association.

25. Find the value of r. $r = .98$. Interpret it. Strong positive linear association.
26. Find the value of r^2. $r^2 = .96$. Interpret it.
    96% of the variation in Salary can be explained by the LSRL.

27. What is the predicted extra salary if Walmart opens 9 lanes? $350.19.
28. What is the predicted extra salary if Walmart opens 7 lanes? $272.75
29. What is the difference between the actual extra salary and predicted extra salary for 7 lanes? –$16.75
30. If Walmart will pay no more than $500 in extra salary, find the maximum lanes it can open. 12
31. If Walmart will pay no more than $250 in extra salary, how the maximum lanes it can open. 6
32. What is the predicted extra salary if Walmart opens no lanes? $1.71
33. What is the predicted extra salary if Walmart opens 50 lanes? $19.37.71
34. Why do the last two problems and answers make little sense for this problem?
    Extrapolation is dangeous. It shouldn't cost anything extra to open no lanes and no Walmart has 50 lanes.
35. A Walmart manager needs to decide how many lanes to open. Based on the data of the and the previous page, explain the dilemma he has in making a decision.
    Open too few lanes, they don't pay much extra money but people have a long waiting time.
    Open too many lane, people have a very short waiting period, but they pay a lot more money.