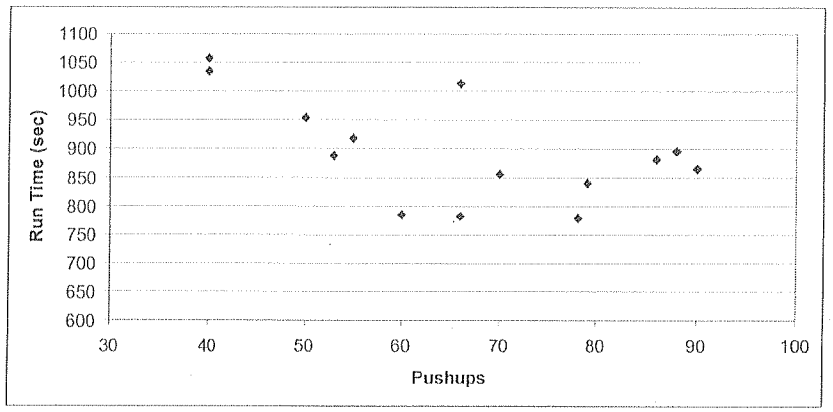
SOLUTION

MA206 Suggested Problems
Lesson 3. Minimizing Sum of Squares of a Line

1. Given the following sample of pushups and run times from the APFT, develop a linear regression model that will predict a soldier's run time (in seconds) from the number of pushups that he did.

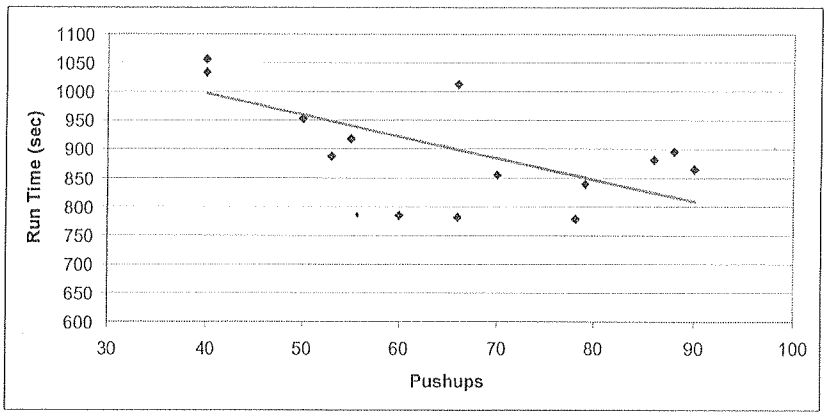
Pushups	40	40	53	55	50	66	79	78	70	86	90	66	60	88
Run Time (sec)	1056	1034	887	919	954	782	840	778	855	881	865	1013	785	896

a. Create a scatter plot of the data. Does it seem sensible to model this data with a linear equation?



It appears that there may be a useful linear relationship between Pushups and Run Time.

b. Develop a linear model by finding values for m and b that minimize the SSE. Add the linear model to your graph. Comment on the adequacy of your linear model.



$\hat{y} = -3.089x + 1099.27$

SSE = 72,036

SOLUTION

The model seems to capture the overall trend of the data set reasonably well, but the very large SSE value indicates that a linear model is not the best choice for this data set.

c. Using your model, predict the run times for soldiers who completed 62 and 99 pushups. Are these predictions valid? Why or why not?

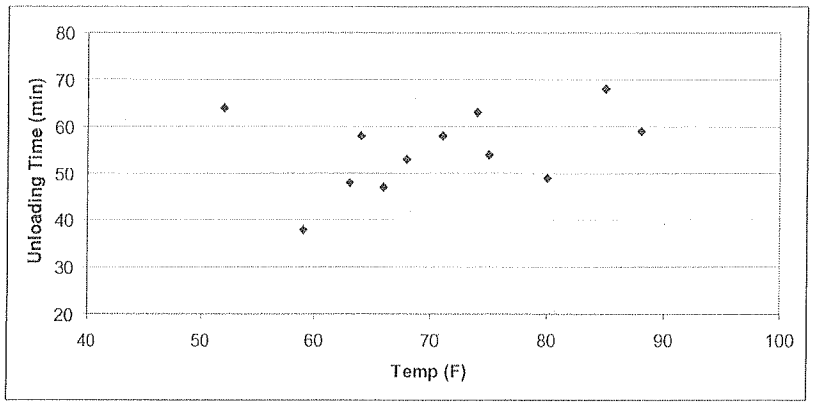
When $x = 62$, $\hat{y} = -3.089(62) + 1099.27 = 907.76$ sec. This prediction is mathematically correct according to our model, but cannot be completely relied upon because a linear model is not a great choice for this data set.

When $x = 99$, $\hat{y} = -3.089(99) + 1099.27 = 793.48$ sec. This prediction is not valid because $x = 99$ is outside the range of the x-values in our original data set. The danger of **extrapolation** is that even if we had a very strong linear relationship here (which we don't), the linear relationship may not be valid for such x-values.

2. A warehouse manager is interested in possible improvements to labor efficiency if air conditioning is installed in his warehouse. He has provided you with the following sample data:

Temp (F)	52	68	64	88	80	75	59	63	85	74	71	66
Unloading Time (min)	64	53	58	59	49	54	38	48	68	63	58	47

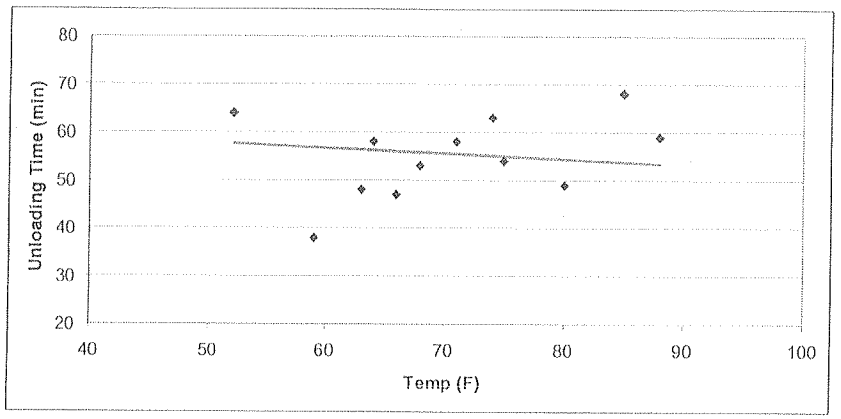
a. Create a scatter plot of the data. Does it seem sensible to model this data with a linear equation?



It appears that there might be a useful linear relationship between Temp and Unloading Time.

b. Develop a linear model by finding values for m and b that minimize the SSE. Add the linear model to your graph. Comment on the adequacy of your linear model.

SOLUTION



$\hat{y} = 0.2659x + 36.19$

SSE = 703.33, which indicates that a linear model is a poor fit for this data set.

c. Provide your recommendation to the warehouse manager.

Since there does not seem to be a strong linear relationship between Temp and Unloading Time, the manger should not install air conditioning in the warehouse.