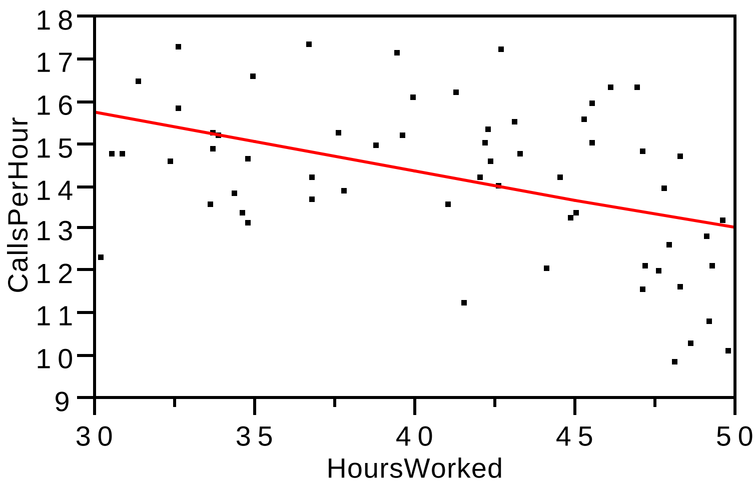
Practice Questions: Simple Regression

A service firm has experienced rapid growth. Because of this growth, some of the employees who handle customer calls have had to work additional hours (overtime). The firm is concerned that over-worked employees are less productive and handle fewer calls per hour than employees who work less demanding schedules. Most employees who work the "conventional" schedule put in 30-40 hours a week, depending upon demand. The firm constructed the regression model shown next relating the number of hours worked (X) to the number of calls serviced per hour (Y) for 60 employees.



Summary of Fit

RSquare	0.19
Root Mean Square Error	1.71

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.8	1.5	13.16	<.0001
HoursWorked	-0.14	0.04	-3.74	0.0004

(1) Does the number of hours worked impact the number of calls handled per hour, or can we explain the declining pattern seen in the plot as a simply a random coincidence?

(2) From the fitted model how many calls on average does an employee who works a 30 hour week process?

(3) For each additional 10 hours of work, how many fewer calls are processed per hour, on average? Might the drop be as large as 2 calls per hour?

(4) How would you interpret the intercept in this fitted model?

(5) When the model was used to predict the number of calls handled by a new employee, the model's prediction of the number of calls per hour was 3 calls per hour *more* than the actual productivity. Can we conclude about the performance of the new employee is consistent with the performance of the employees used to build this model?

(6) This fitted model uses data from 60 employees. If the sample size were increased to 240 (i.e., fit to a sample that is four times larger), then

(a) What would happen to the confidence interval for the slope?

(b) Would the predictive accuracy of the model improve, especially when used to predict the performance within the range of prior experience?

(c) What would happen to the estimates of the slope and intercept?

(d) Would the model's R^2 goodness of fit index increase?

(7) Which of the following is the best characterization of the R^2 statistic for this model?

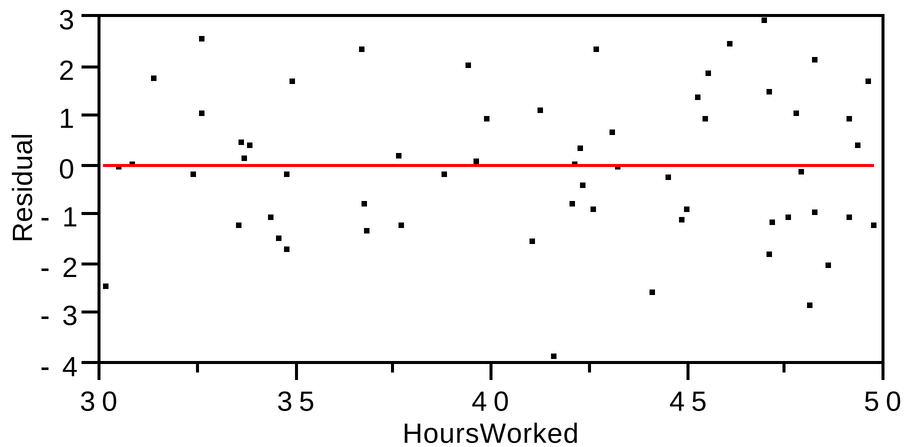
(a) The model explains about 20% of the variation in productivity.

(b) The model predicts about 20% of the employees accurately.

(c) The model predicts about 80% of the employees accurately.

(d) Only 20% of the observations lie within 2 RMSEs of the fitted line.

(8) Does the following residual plot belong to the fitted linear model?



(1) The slope in the fitted model is significantly different from zero, so we cannot dismiss this relationship as a chance occurrence (from either the fact that the confidence interval does not include zero, the t-ratio is larger than 2 in absolute size, or the small p-value which is less than 0.05).

(2) From the fitted equation, we find an estimate of 19.8 – 0.14 (30) = 15.6 calls per hour. You could also get close to this value from looking at the left edge of the plot.

(3) The fitted slope is –0.14 with a confidence interval of

$$[-0.14 - 2(.04), -0.14 + 2(.04)] = [-0.22, -0.06]$$

When multiplied by 10 to get the effect for a 10 hour increase, the associated interval is then [–2.2, –.6], so indeed the drop could be as large as –2 (it's inside the interval).

(4) The intercept is quite far from the range of observation, representing quite an extrapolation. A literal interpretation would be foolish here (i.e., 20 calls handled in no hours).

(5) The RMSE = 1.7 implies that the model's predictions ought to be within about 3.4 calls per hour. This employee is consistent with the performance suggested by the model.

(6) The confidence interval would be about half as long, the RMSE would be about the same, the slope and intercept would be similar (though probably closer to the unknown true values), and the R^2 would be about the same.

(7) A is the correct interpretation.

(8) No, the RMSE is too small and the model does not fit the pair of points on the left of the plot so well as this figure would indicate. The fit is higher than these two in the original figure.