

FAQ for Simple Linear Regression

A baseball analyst would like to develop a model to predict the number of wins during the 2000 baseball season based on ERA. The analyst collected several variables: Wins, ERA, RunsScored, HitsAllowed, WalksAllowed, Errors, and Saves from 30 professional baseball teams. The [data](#) are in this linked file.

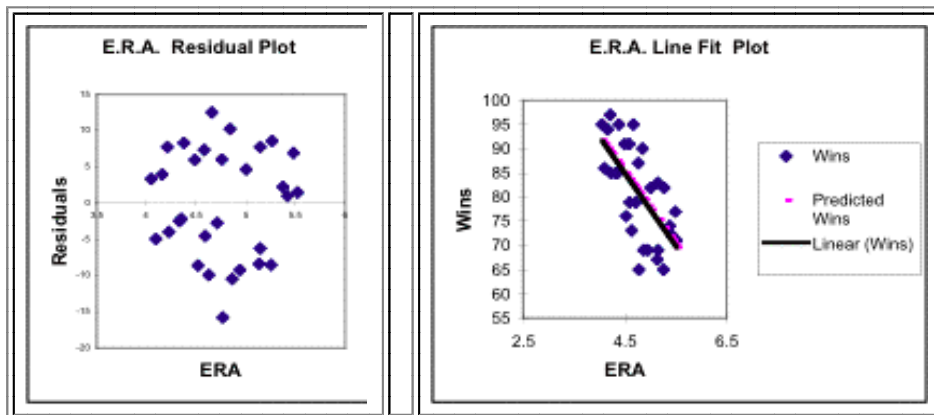
Summary Output

Regression Statistics	
Multiple R	0.6598
R Square	0.4354
Adj. R Square	0.4152
Standard Error	7.6363
Observations	30

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1259.11	1259.11	21.5924	7.282E-05
Residual	28	1632.75	58.3127		
Total	29	2891.86			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	152.8097	15.5308	9.8392	1.375E-10	120.9964	184.6230
ERA	-15.0927	3.2480	-4.6468	7.282E-05	-21.7459	-8.4394



Part A: Scatterplot between Wins and ERA

- Identify which variable is the X or explanatory variable. [\[go to answer\]](#)
- Identify which variable is the Y or response variable. [\[go to answer\]](#)
- Describe the pattern of points as they appear on the graph. [\[go to answer\]](#)
- What kind of a relationship do you see? [\[go to answer\]](#)
- Are there any "outliers"? [\[go to answer\]](#)

Part B: Correlation between *Wins* and *ERA*

1. Describe the strength and direction of the correlation. [\[go to answer\]](#)
2. Compare this relationship with the pattern of points on the scatter diagram between the two variables. [\[go to answer\]](#)

Part C: Simple Linear Regression output for these two variables.

1. Write the specific estimated regression equation for this problem. [\[go to answer\]](#)
2. Using the estimated regression equation predict the average *Wins* for a team with an *ERA* of 4.50. **Show all your calculations.** [\[go to answer\]](#)
3. Is the previous prediction extrapolation? Why or why not? [\[go to answer\]](#)
4. Interpret the slope estimate; that is, explain what it means in terms of this problem. [\[go to answer\]](#)
5. How much variation in *Wins* is accounted for by this regression model? Express your answer as a **percent**. What measure did you use to answer this question? [\[go to answer\]](#)
6. What is the standard error of the estimated regression line? Include the **unit of measurement** in your answer. [\[go to answer\]](#)
7. Using a 5% level of significance, is there evidence of a linear relationship between *Wins* and *ERA*? Be sure to state the hypothesis, test statistic, *p*-value, and conclusion. [\[go to answer\]](#)
8. Give a 95% confidence interval for the true (i.e., population) slope. [\[go to answer\]](#)
9. If the original correlation between these two variables were not known, how could it be calculated using the statistics in the regression output? **Show formula and calculations.** How do you determine the sign of the correlation? [\[go to answer\]](#)

Part D: Residual Plot

1. Describe what you see on the residual plot. [\[go to answer\]](#)
2. Is the linear model a good fit for the data? **Explain** your answer. [\[go to answer\]](#)
3. For this data set, look and see what the data values are for the first pair of observations (*ERA*, *Wins*). Then **calculate the residual**, i.e. calculate the difference between the observed *Wins* and the predicted *Wins* for that team. **Show all the formulas and calculations.** [\[go to answer\]](#)

Part E: Fitted Line Plot

1. Describe the appearance of the Fitted Line Plot. [\[go to answer\]](#)
2. Is the model a good fit for this data? Be sure to state your decision and give the reasons that support your decision. [\[go to answer\]](#)
3. How useful do you think this regression model is for predicting *Wins*? Explain. [\[go to answer\]](#)
4. What other independent variables might be considered for inclusion in the model [\[go to answer\]](#)

ANSWERS:

answers are obtained from your instructor