Quiz Regression

A mail-order catalog business that sells personal computer supplies, software, and hardware maintain a centralized warehouse for the distribution of products ordered. Management is currently examining the process of distribution from the warehouse and is interested in studying the factors that affect warehouse distribution costs. Currently, a small handling fee is added to the order, regardless of the amount of the order. Data have been collected over the past 24 months indicating the warehouse distribution costs and the number of orders received. The results and the scatter plot for the Cost and Orders. Assume the Cost is the dependent variable (Y) and Orders is the independent variable (X). The Costs are in $1000.

The Summary Output Table above is resulted from regression analysis with Orders as independent variable (X) and Costs as dependent variable (Y) to fit a linear trend model to the data set. Use the Table to answer the following questions:

1. Which one(s) of the following is not correct regarding b_1?
    a. The 95% confidence interval is from 0.0131 to 0.0192
    b. The costs will increase $16.10 for each additional order received.
    c. The slope of the regression line and is given by =SLOPE(D4:D27,C4:C27) in Excel@
    d. Because the p-value is 0.0000 and is less than 0.05, thus a non zero b_1 (or $b_1 \neq 0$) is confirmed.

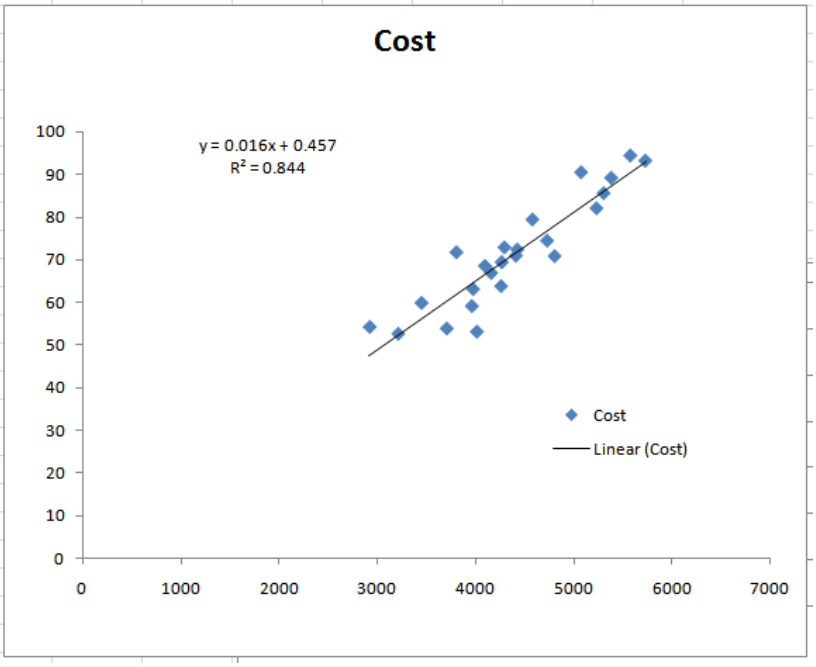
2. Which one(s) of the following is not correct regarding b_0?
    a. The Y intercept of the regression line and is given by =INTERCEPT(D4:D27,C4:C27)
    b. It costs the firm $457.60 to set up the operations (or the fixed cost) without even receiving any order yet
    c. Because the p-value is 0.9451 and is larger than 0.05, thus $\beta_0 = 0$ is confirmed, thus it is not necessary to include the Y intercept term in the regression model
    d. The 95% confidence interval is from -13.1716 to14.0869, it spans zero (or zero is part of the confidence interval), thus $\beta_0 \neq 0$ is confirmed, thus the Y intercept term should be included in the regression model

3. Which of the following Excel@ function or procedure cannot be used to get the estimates for the regression line (b_0 and b_1) or provide predictions:
    a. Use Excel@ =SUMPRODUCT(D4:D27,C4:C27) to compute b_1
    b. Click any data point(s) on the scatter plot of the Costs and Orders, select Add Trendline / Display equations & Display R-Squared value on the chart
    c. Use Excel@ =TREND(D4:D27,C4:C27,XValue) to produce predictions
    d. Select Data/Data Analysis/Regression with Orders and Costs ranges provided to get regression Summary Output

4. Which one(s) of the following is NOT correct regarding the estimated regression function?
    a. The regression line is given by $\hat{Y} = 0.4576 + 0.0161$ X
    b. The regression line is given by $\hat{Y} = 0.0161 + 0.4576$ X
    c. For 3,500 orders, the costs will be around $56,869.10 as given by Excel@ =TREND(D4:D27,C4:C27,3500)
    d. The residual (error) is $12,219.64 for 4015 orders as given in the first row of the data set

5. Which one(s) of the following is NOT correct regarding the R Square (R^2) for the data?
    a. 84.42% of the total variations of the data are explained by the regression line.
    b. 15.58% of the total variations of the data are explained by the regression line.
    c. It is the ratio of the RSS(Regression Sum of Squares) to TSS (Total Sum of Squares) or RSS/TSS.
    d. The multiple R =0.9188 is the square root of the R Square of 0.8442.
6. What is the percentage of the total variations NOT explained by the regression line?
    a. 15.58%
    b. 84.42%
    c. 91.88%
    d. 83.71%
7. The Standard Error of Estimate (S_{YX} *or* S_e) to be used to approximate 95% prediction interval for the costs is:
    a. 5.2183
    b. 6.5719
    c. 0.0015
    d. 27.230
8. Which one(s) of the following is NOT correct regarding the relationship between the standard error of estimate (S_e) and the standard prediction error (S_p)?
    a. S_e is always smaller than S_p
    b. S_e is always greater than S_p
    c. S_e is the RMSE (Root Mean Squared Error) or the square root of MSE
    d. S_p equals to S_e times a value that is greater than one.
9. Which one(s) of the following is NOT correct regarding the 95% prediction interval of the costs when specifically 4000 orders are received?
    a. It is wider than the 95% prediction interval for the mean costs when 4000 orders are received.
    b. It is given by the computed costs ($\hat{Y}$) with the X=4000 orders ($\hat{Y} = 0.4576 + 0.0161 *$ 4000) then plus or minus the margin of prediction error ($t\ S_p$)
    c. It equals to $\hat{Y}_i \pm 2\ S_e$
    d. Non of the others is correct.
10. You may notice that the minimum and maximum numbers of orders for the data set are 2921 and 5735 orders respectively. Which one(s) of the following cannot be done regarding the use of the regression line?
    a. To predict the costs when 3500 orders are received
    b. To predict the costs when 5500 orders are received
    c. To predict the costs when 2500 orders are received
    d. To predict the costs when 4100 orders are received

Month	Sales	Orders	Cost
1	386	4015	52.95
2	446	3806	71.66
3	512	5309	85.58
4	401	4262	63.69
5	457	4296	72.81
6	458	4097	68.44
7	301	3213	52.46
8	484	4809	70.77
9	517	5237	82.03
10	503	4732	74.39
11	535	4413	70.84
12	353	2921	54.08
13	372	3977	62.98
14	328	4428	72.3
15	408	3964	58.99
16	491	4582	79.38
17	527	5582	94.44
18	444	3450	59.74
19	623	5079	90.5
20	596	5735	93.24
21	463	4269	69.33
22	389	3708	53.71
23	547	5387	89.18
24	415	4161	66.8



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9188
R Square	0.8442
Adjusted R Square	0.8371
Standard Error	5.2183
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3246.062	3246.062	119.207	0.000
Residual	22	599.068	27.230		
Total	23	3845.130			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.4576	6.5719	0.0696	0.9451	-13.1716	14.0869
Orders	0.0161	0.0015	10.9182	0.0000	0.0131	0.0192

(Berenson, Levine and Krehbiel, 10th Edition, Q.13.36 on pp. 540-541)