1. We list for 11 printers, their cost (C, in hundreds of dollars) and their rating (R, on a scale from 0-100). The results are

<table>
<thead>
<tr>
<th>C</th>
<th>4</th>
<th>5</th>
<th>2.5</th>
<th>2</th>
<th>4</th>
<th>2.85</th>
<th>2.1</th>
<th>2.5</th>
<th>2</th>
<th>1.8</th>
<th>1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>90</td>
<td>90</td>
<td>85</td>
<td>83</td>
<td>78</td>
<td>78</td>
<td>75</td>
<td>70</td>
<td>68</td>
<td>63</td>
<td>61</td>
</tr>
</tbody>
</table>

Minitab output is attached. Discuss the scatter plot of the data where C is the dependent variable and R is the independent variable.

a. Interpret the correlation coefficient (r).

r = 0.754, positive moderately strong correlation.

b. Report and Interpret the least square line (regression line) which will be used to predict the Cost of the printer from the Rating.

\[ \hat{C} = -3.738 + 0.085R \]

Each point of rating increases cost by $8.50

c. Test the hypothesis that Cost and Rating are correlated. Explain why this is the same as testing whether the slope of the regression line is not equal to zero.

Ho: Cost and Rating are not correlated.
Ha: Cost and Rating are correlated.

This is the same as testing Ho: the slope = 0, which would eliminate Rating from the model.

Model: Simple Linear Regression ANOVA \( F = \frac{MS_{\text{Regression}}}{MSE_{\text{Error}}} \)

\( P\text{value} = 0.0073 < 0.01 \rightarrow \text{Reject Ho at } \alpha = 0.01 \) (my choice)

Conclusion: Cost and Rating are correlated, higher rated printers cost more.

d. Report and interpret the standard error of C|R (the standard deviation "with respect to" the regression line.)

\[ s_e = \sqrt{0.6165} = 0.785 \text{ This value is the estimate of the standard deviation of the residual error.} \]

e. Report and interpret \( r^2 \), the coefficient of determination.

\[ r^2 = \frac{7.3229}{12.8714} = 0.569 \text{ 56.9\% of the variability of cost can be explained by rating.} \]

f. Analyze the residuals and note any unusual observations.

Observation 4 is most unusual, but still less than 2 standard errors from the regression equation, so there are no apparent outliers.
g. Make Cost predictions for the following printers. Are any of these predictions inappropriate?

- Printer X – Rating =87
- Printer Y – Rating =75
- Printer Z – Rating =21 \(\text{not appropriate due to extrapolation.}\)

h. Find a 95% confidence interval for the expected Cost of Printers X and Y.

   \text{Printer X: ($283, $441)} \quad \text{Printer Y: ($207, $315)}

i. Find a 95% prediction interval for the actual Cost of Printers X and Y.

   \text{Printer X: ($168, $557)} \quad \text{Printer Y: ($75, $447)}
Minitab Output for Question 1

Regression Analysis: Cost versus Rating

The regression equation is
Cost = -3.74 + 0.0846 Rating

Predictor Coef SE Coef T P
Constant -3.738 1.892 -1.98 0.080
Rating 0.08462 0.02455 3.45 0.007

S = 0.785173 R-Sq = 56.9% R-Sq(adj) = 52.1%

Analysis of Variance

Source DF SS MS F P
Regression 1 7.3229 7.3229 11.88 0.007
Residual Error 9 5.5485 0.6165
Total 10 12.8714

Obs Rating Cost Fit SE Fit Residual St Resid
1 90.0 4.000 3.878 0.408 0.122 0.18
2 90.0 5.000 3.878 0.408 1.122 1.67
3 85.0 2.500 3.455 0.316 -0.955 -1.33
4 83.0 2.000 3.286 0.286 -1.286 -1.76
5 78.0 4.000 2.863 0.240 1.137 1.52
6 78.0 2.850 2.863 0.240 -0.013 -0.02
7 75.0 2.100 2.609 0.239 -0.509 -0.68
8 70.0 2.500 2.186 0.285 0.314 0.43
9 68.0 2.000 2.016 0.315 -0.016 -0.02
10 63.0 1.800 1.593 0.406 0.207 0.31
11 61.0 1.300 1.424 0.447 -0.124 -0.19

Predicted Values for New Observations

New Obs Fit SE Fit 95% CI 95% PI
1 3.624 0.351 (2.831, 4.418) (1.679, 5.570)
2 2.609 0.239 (2.067, 3.150) (0.752, 4.466)
3 -1.961 1.382 (-5.087, 1.166) (-5.556, 1.635)

Values of Predictors for New Observations

New Obs Rating
1 87.0
2 75.0
3 21.0
2. The following regression analysis was used to test Poverty (percentage living below the poverty line) as a predictor for Dropout (High School Dropout Percentage). Five items have been blanked out been can be calculated based on other information in the output.

\[
r^2 = \frac{67.45}{283.62} = 0.238
\]

\[
r = \sqrt{0.238} = 0.488
\]

\[
\text{Std. Error} = \sqrt{4.50} = 2.12
\]

\[
F = \frac{67.45}{4.50} = 14.99
\]

\[
Y_{\text{hat}} = 10.577
\]

b. \(Y_{\text{hat}} = 6.212 + 0.291X\)

c. Conduct the Hypothesis Test that Poverty and HSDropout are correlated with \(\alpha = .01\) (Critical Value for F is 7.19 (\(\alpha = .01\), DFnum=1,DF=48)).

\[
\text{Ho: Poverty and HS Dropout are not correlated.}
\]

\[
\text{Ha: Poverty and HS Dropout are correlated.}
\]

\[
\text{Model: Simple Linear Regression ANOVA } F = \frac{\text{MSRegression}}{\text{MSError}} = 14.99 > 7.19 \text{ Reject Ho}
\]

Conclusion: Poverty and HS Dropout are positively correlated.

d. \(Y_{\text{hat}} = 6.212 + 0.291X\)

\[
r^2 = 0.238 \text{ or } 23.8\%
\]
e. North Dakota has a Poverty Rate of 11.9 percent and a HS Dropout Rate of 4.6 percent.

i. Calculate the predicted HS Dropout Rate for North Dakota from the regression equation.

\[ \text{Yhat}_{11.9} = 6.212 + 0.291(11.9) = 9.675 \]

ii. The Standard Error (from part a-iii) is the standard deviation with respect to the regression line. Calculate the Z-score for the actual North Dakota HS Dropout Rate of 4.6 (Subtract the predicted value and divide by the Standard Error). Do you think that the North Dakota HS Dropout Rate is unusual? Explain

\[ Z = (4.6 - 9.675)/2.12 = -2.39 \]

North Dakota's Actual Dropout rate is unusually low.