Math 226 **Assignment #3**

Topic: Least squares regression, model assessment, and transformations

Reading Assignment: Read sections 2.5-2.9 in Chapter 2;

Assigned problems:

2.24 Linear correlation

*Note: To solve parts (a) & (c), create tables in Excel (with column format similar to Table in center of page 74). Include tables in assignment.

(a)

	Measured RSF (X)	Calculated RSF (Y)	Zx	Zy	ZxZy
	0.91	0.95	0.05	0.75	0.04
	0.87	0.90	-0.41	0.23	-0.09
	0.83	0.84	-0.88	-0.39	0.35
	0.78	0.77	-1.47	-1.12	1.65
	0.78	0.68	-1.47	-2.06	3.02
	1.00	0.98	1.11	1.06	1.18
	1.01	0.95	1.23	0.75	0.92
	0.98	0.93	0.88	0.54	0.47
	0.94	0.98	0.41	1.06	0.43
	1.00	0.96	1.11	0.85	0.95
	0.94	0.94	0.41	0.64	0.26
	0.93	0.91	0.29	0.33	0.10
	0.97	0.86	0.76	-0.19	-0.14
	0.88	0.80	-0.30	-0.81	0.24
	0.76	0.72	-1.70	-1.64	2.80
Mean	**0.91**	**0.88**		**SUM**	**12.17**
Std Dev	**0.085**	**0.096**			

Correlation r = $\frac{\sum_{i=1}^{n} ZxZy}{(n-1)} = \frac{12.17}{(15-1)} = \mathbf{0.8694}$

(b) The correlation coefficient indicates a strong, positive correlation between the measured RSF and the calculated RSF. If the measured RSF was high, then the calculated RSF also tended to be high.

(c)

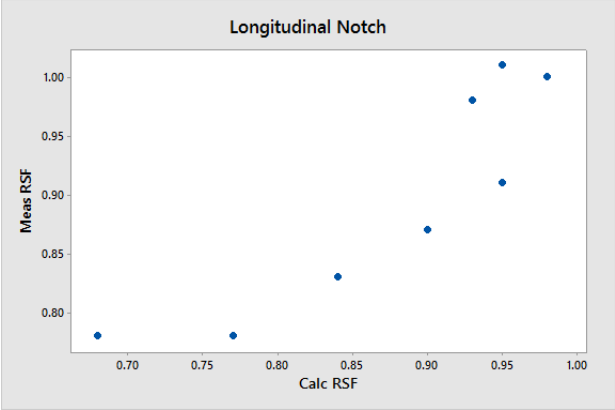
	Longitudinal notch				
	Measured RSF (X)	Calculated RSF (Y)	Zx	Zy	ZxZy
	0.91	0.95	0.16	0.72	0.11
	0.87	0.90	-0.26	0.24	-0.06
	0.83	0.84	-0.69	-0.34	0.23
	0.78	0.77	-1.21	-1.01	1.22
	0.78	0.68	-1.21	-1.87	2.27
	1.00	0.98	1.11	1.01	1.11
	1.01	0.95	1.21	0.72	0.87
	0.98	0.93	0.90	0.53	0.47
Mean	**0.90**	**0.88**		**SUM**	**6.22**
Std Dev	**0.095**	**0.104**			

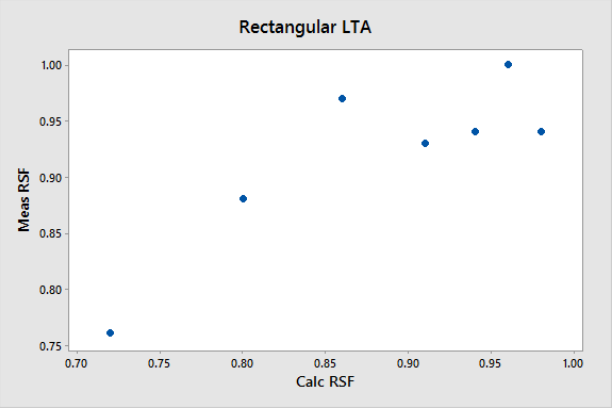
Correlation r = $\frac{\sum_{i=1}^{n} ZxZy}{(n-1)} = \frac{6.223}{(8-1)} = \mathbf{0.8891}$

	Rectangular LTA				
	Measured RSF (X)	Calculated RSF (Y)	Zx	Zy	ZxZy
	0.94	0.98	0.29	1.05	0.30
	1.00	0.96	1.06	0.83	0.88
	0.94	0.94	0.29	0.62	0.18
	0.93	0.91	0.16	0.30	0.05
	0.97	0.86	0.67	-0.23	-0.15
	0.88	0.80	-0.47	-0.86	0.41
	0.76	0.72	-2.00	-1.71	3.43
Mean	**0.92**	**0.88**		**SUM**	**5.10**
Std Dev	**0.078**	**0.094**			

Correlation r $= \frac{\sum_{i=1}^{n} ZxZy}{(n-1)} = \frac{5.102}{(7-1)} = \mathbf{0.8504}$

The strength of the association is comparable for both types of flaws, although the correlation for LN is stronger (more linear) than that for LTA. Looking at the scatterplots, it is noticeable that the LN plots seems to be curved upward, and the LTA plot seems to be curved down, so both could probably be better represented with a different model that is non-linear.

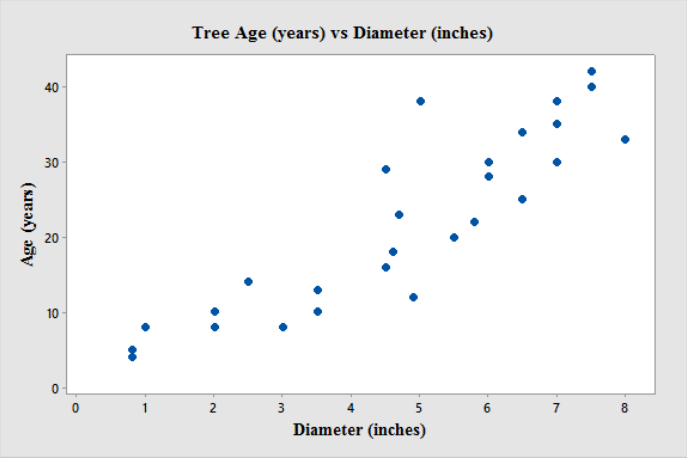




Problem 2

An observational study was completed to analyze the relationship between oak tree diameter (X in inches) and the age of the tree (Y in years).

a. Make a scatter plot and determine the linear correlation coefficient, r. Is there an association between age and diameter?



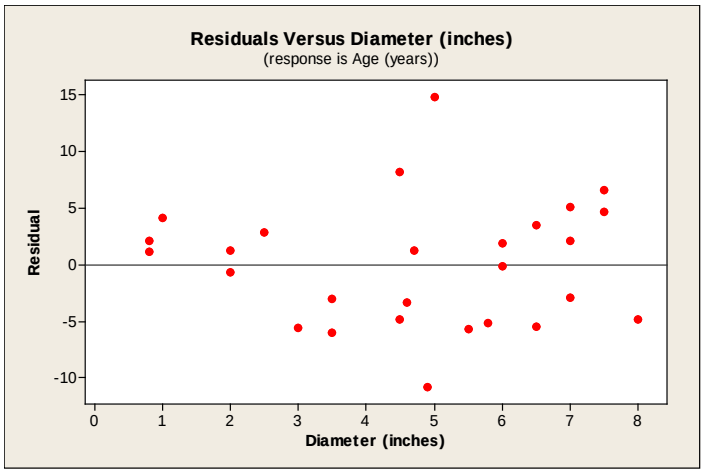
Pearson correlation of Age (years) and Diameter (inches) = 0.888

b. Create a least square regression model using Minitab. Provide the fitted model, R^2, and S.

Age = - 0.95 + 4.85 Diameter *(with Diameter in inches & Age in years)*
S = 5.57194 R-Sq = 78.9% R-Sq(adj) = 78.1%

c. Age = - 0.95 + 4.85(5.5) = 25.725 years
If the diameter of a tree is 5.5 inches, the model predicts that the age of the tree is 25.7 years;

d. Construct a residual plot in Minitab (i.e. Explanatory Variable (x) versus Residuals).



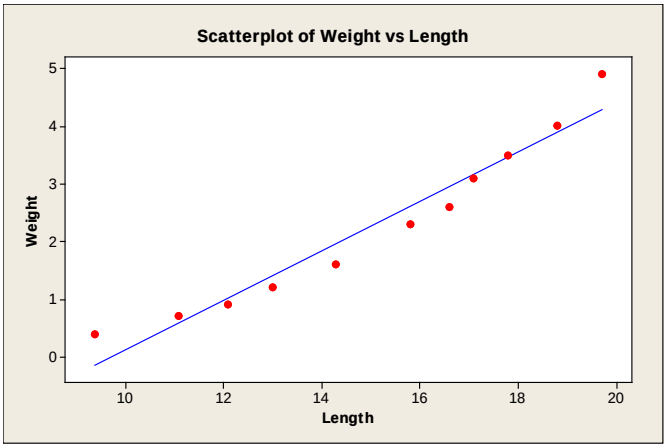
e. Comment on the shape of the residual plot and the interpretation. Based on your assessment, does a linear model seem appropriate for this data

The residual plot displays an arching pattern (or +/-/+ pattern), i.e. there is a potential curvature in the predictor variable. This indicates that the residuals are not randomly distributed, and a linear model, therefore, may not be appropriate for this data,.

Problem 3

a. Response variable = the weight of a largemouth bass (lb);
Explanatory variable = the length of the fish (in);

b. Create (and include) a scatter plot of Y versus X using Minitab. Comment on the appropriateness of a linear model for this data.

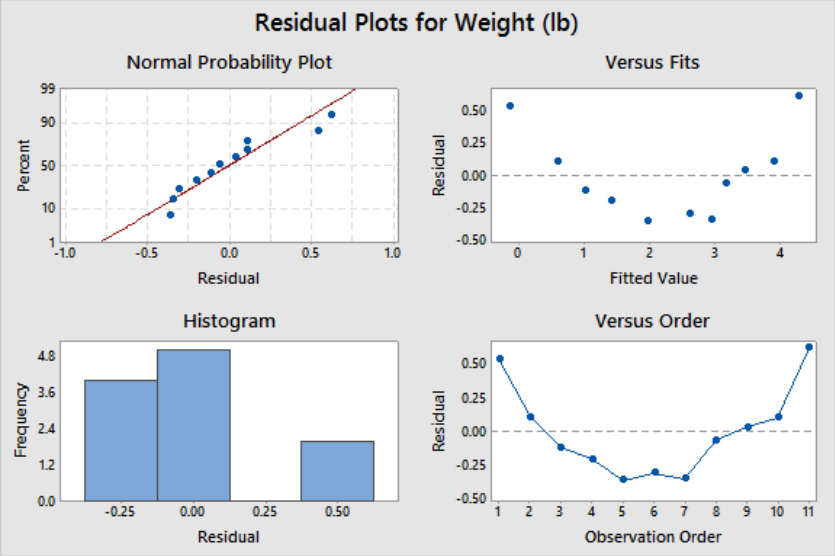


The data do not appear to be linear, but rather are curved (convex upward). A non-linear model may be better for this data.

c. Complete a regression analysis using Minitab to find a least squares linear model for predicting the weight of a largemouth bass from its length. Provide the following results: The linear regression equation, the standard error if the estimates (S), the coefficient of determination (R^2), and the residual plot. Comment on the quality of your model on the basis of S and R^2.

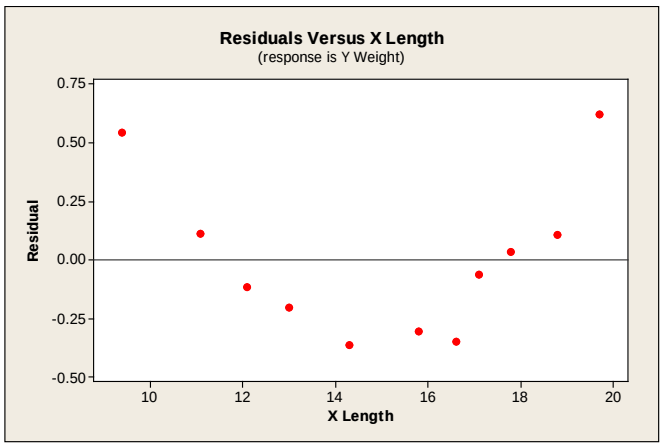
```
Regression Equation
Weight (lb) = -4.177 + 0.4294 Length (in)

S = 0.350974   R-Sq = 94.9%   R-Sq(adj) = 94.3%
```



The R^2 value is fairly high (i.e. close to 1), indicating that about 95% of the variation in Y is described by the linear model. However, the residual plots display a very obvious pattern, and the normal probability plot indicates that the residuals are not normally distributed (i.e. not randomly distributed). The pattern is suggesting a problem with the linear model, and an upward curvature.

d. Create a residual plot (Fish length (x) versus Residuals Errors). Comment on the quality of your model on the basis of the plot.



The plot indicates that there is a very obvious pattern, i.e. curvature in the relationship with X, which should be accounted for in the model.

e. Weight = -4.177 + 0.4294(15) = 2.26358 lb

The weight of a 15 inch long fish is predicted to be 2.26 lb.

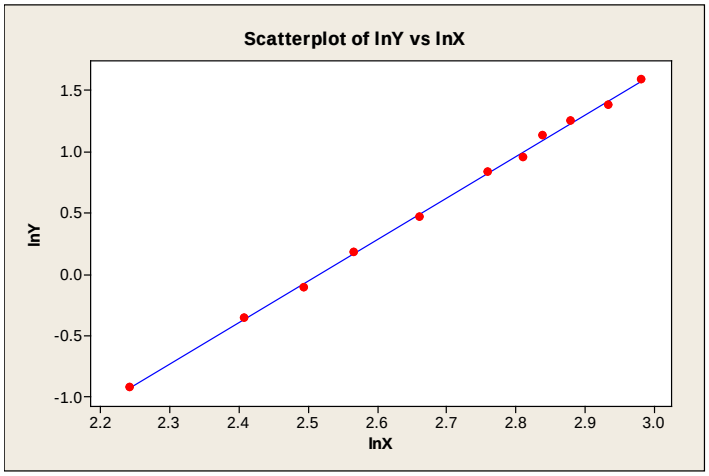
Problem 4

a. Power transformation on the data associated with Weight = f(length^n). Provide a table of transformed data and a scatterplot.

X is length in inches and Y is weight in lb

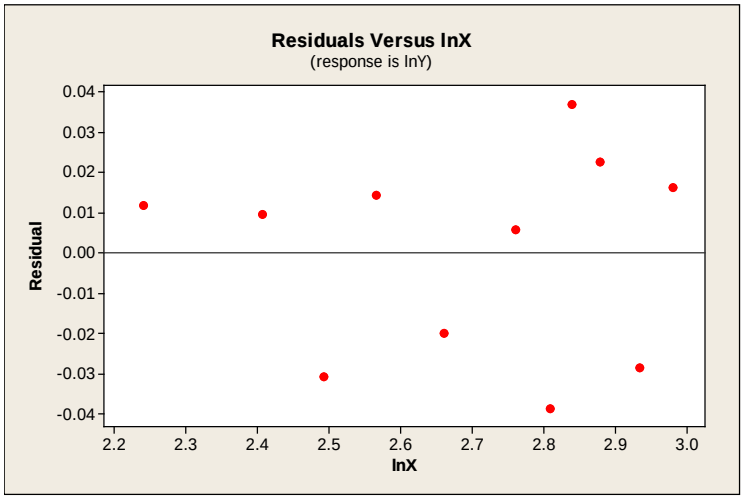
$Y = c(X^3)$ $\quad \ln(Y) = \ln(cX^3)$ $\quad \ln(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(X)$

lnX	lnY
2.24071	-0.91629
2.40695	-0.35667
2.49321	-0.10536
2.56495	0.18232
2.66026	0.47000
2.76001	0.83291
2.80940	0.95551
2.83908	1.13140
2.87920	1.25276
2.93386	1.38629
2.98062	1.58924



b. Perform a regression analysis on your transformed data. Provide the following results: The regression equation, the standard error if the estimates (S), the coefficient of determination (R^2), and the residual plot (Explanatory variable versus residuals).

```
The regression equation is lnY = - 8.50 + 3.38 lnX
S = 0.0263963   R-Sq = 99.9%   R-Sq(adj) = 99.9%
```



c. Comment on the quality of your transformed model on the basis of S, R^2, and residual plot.

This transformed model is a much better fit, with a very high R^2 of 0.99, a much lower S than the non-transformed model, and a residual plot that does not display a pattern.

d. Use your transformed model to predict the weight of a fish whose length is 15 inches.

```
lnY = - 8.50 + 3.38 ln(15) = 0.65164
```

$Y = e^{0.65164} =$ 1.9187 lb