U Wroclaw, Fall 2015
Applied Stats
DISCUSSION/LAB 5: SIMPLE LINEAR REGRESSION

We will use data set: slrclass1.MTW.

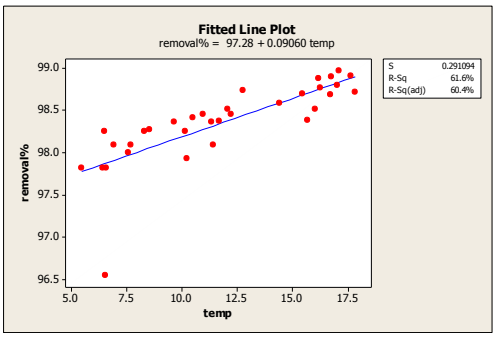
TO DO

In biofiltration of wastewater, air discharged from a treatment facility is passed through a damp porous membrane that causes contaminants to dissolve in water and be transformed into harmless products. The data used for our lab on x=inlet temperature (deg C) and y=removal efficiency (%) was the basis for an article in Water Environment Research, 2001, 426-435. The data is on the class web site …..

1. Plot a scatter plot of the data. Does it suggest a linear relationship between the temperature and removal efficiency?
2. Fit a simple linear regression model to this data.
3. Does the model seem to fit the data?
4. Does removal efficiency seem to increase or decrease with temperature?
5. What is the average difference in the removal efficiency when temperature is raised by 1^oC ?
6. What removal efficiency would you expect at temperature of 15^oC?
7. Find e_8.
8. Estimate σ.
9. Find $\sum e_i$.
10. What is the probability that the removal efficiency will exceed 98% at temperature of 15^oC?
11. Find a 98% PI for the removal efficiency at 15^oC?
12. Find the coefficient of determination for the regression model.
13. Find the Pearson correlation coefficient between removal efficiency and temperature.
14. Test, on 1% significance level, if the slope of the regression line is significantly greater than zero.
15. Perform the analysis of residuals to check if the model we got satisfies the assumptions for the SLR model: residuals should come as iid observations from a normal distribution with mean 0, and constant st. dev. You may use regular residuals or standardized residuals. The standardized residuals should be iid from N(0, 1).

SOLUTIONS- LAB WORK

1. Scatter plot.



When doing regression, save residuals, and fitted values.

Regression Analysis: removal% versus temp

```
The regression equation is
removal% = 97.3 + 0.0906 temp

Predictor     Coef  SE Coef       T      P
Constant   97.2783   0.1603  607.02  0.000
temp       0.09060  0.01284    7.06  0.000

S = 0.291094   R-Sq = 61.6%   R-Sq(adj) = 60.4%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  4.2199  4.2199  49.80  0.000
Residual Error  31  2.6268  0.0847
Total           32  6.8468

Predicted Values for New Observations

New
Obs      Fit  SE Fit        98% CI              98% PI
  1  98.6373  0.0649  (98.4781, 98.7965)  (97.9058, 99.3688)

Values of Predictors for New Observations

New
Obs  temp
  1  15.0
```

Descriptive Statistics: temp, removal%

```
Variable   N  N*    Mean  SE Mean  StDev  Variance  Minimum      Q1  Median
temp      33   0  11.842    0.698  4.008    16.066    5.480   8.000  11.440

Variable      Q3  Maximum
temp      16.100   17.830
```

2. Fitted regression line:

1. Does the model seem to fit the data?

2. Does removal efficiency seem to increase or decrease with temperature?

3. What is the average difference in the removal efficiency when temperature is raised by $1^{o}C$?

4. What removal efficiency would you expect at temperature of $15^{o}C$?

5. Find e_8.

6. Estimate σ.

7. Find $\sum e_i$.

8. What is the probability that the removal efficiency will exceed 98% at temperature of 15°C?

9. Find a 98% PI for the removal efficiency at 15°C?

10. Find the coefficient of determination for the regression model.

11. Find the Pearson correlation coefficient between removal efficiency and temperature.

12. Test, on 1% significance level, if the slope of the regression line is significantly different from zero.

Ho: $\beta_1=0$ vs. Ha: $\beta_1 \neq 0$

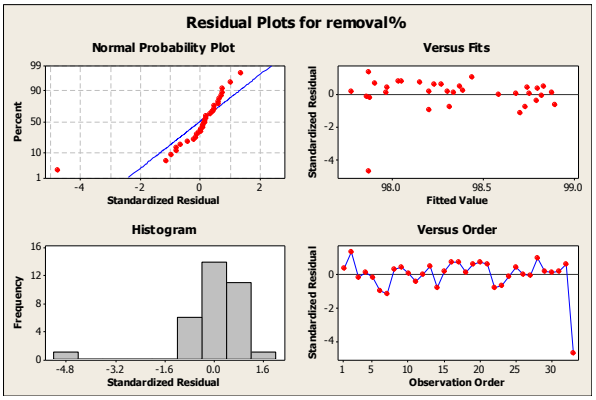
Test statistic:

Predictor	Coef	SE Coef	T	P
Constant	97.2783	0.1603	607.02	0.000
temp	0.09060	0.01284	7.06	0.000

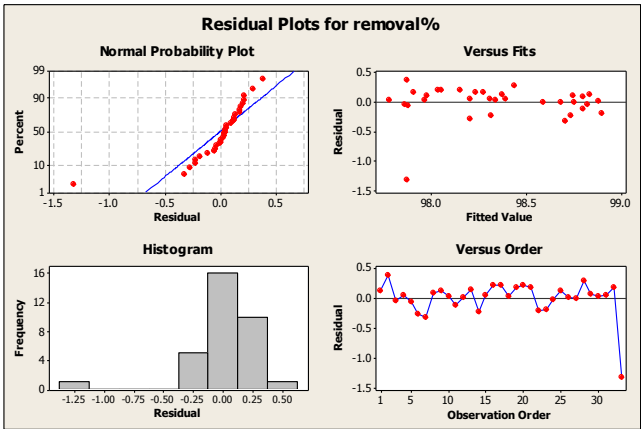
p-value ~0.
Decision: Reject Ho**, the slope of the regression line is significantly different than zero**.

13. Analysis of residuals.

STANDARDIZED RESIDUALS

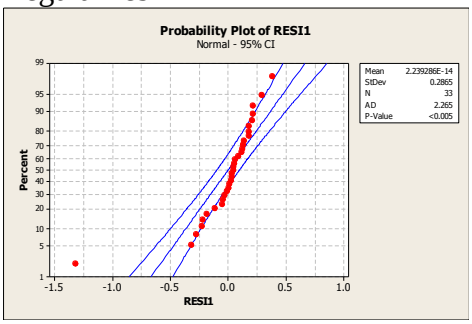


REGULAR RESIDUALS



Probability plots of residuals:

Regular res



Stand res