

STAT 2509A
Test#1

(October 4th, 2012)

All solutions must be shown clearly.

1. Suppose that the following data were collected on emphysema patients: the number of years the patient smoked and inhaled and a physician's subjective evaluation of the extent of lung damage. The latter variable is measured on scale of 0 to 100. Measurements taken on 10 patients are as follows:

Lung Damage	55	60	50	30	75	70	70	55	30	35
Yrs of Smoking	25	36	22	15	48	39	42	31	28	33

$$\sum y_i = 530$$

$$\sum x_i = 319$$

$$\sum y_i^2 = 30\,600$$

$$\sum x_i^2 = 11\,053$$

$$\sum x_i y_i = 18\,055$$

- [1] (a) The response variable, y , is: _____
- [1] (b) The explanatory variable, x , is: _____
- [6] (c) State a SLR model making sure you give all assumptions necessary for statistical inference.
- [5] (d) Find the least squares estimates of β_0 and β_1 . Find the least squares fitted regression line.

Assuming no violations of the assumptions, answer the following questions:

- [6] (e) Find s^2 , an estimate of σ^2 .
- [6] (f) Use the t-test to test whether there is a significant linear relationship between the years of smoking and the lung damage. Use $\alpha = 0.10$.
- [4] (g) Find a 90% confidence interval for the true population slope, β_1 .
- [23] (h) Complete the following ANOVA table and hence test whether there is a significant linear relationship between the years of smoking and the lung damage. Use $\alpha = 0.10$.

Source	d.f.	SS	MS	F
Regression				
Error				
Total		2510		

- [5] (i) Find the values of the coefficient of correlation, r , and coefficient of determination, r^2 , and interpret their meanings in this problem. What is your conclusion about the model?
- [5] (j) Find a 95% confidence Interval of the average lung damage of people who smoked for 40 years.
- [5] (k) Find a 95% Prediction Interval of the lung damage of an individual who smoked for 40 years.

2. An investigator was interested in examining the effect of different doses of a new drug on the pulse rates of human subjects. Four doses of the drug were used in the experiment (1.5, 2.0, 2.5 and 3.0 ml/kg of body weight). Three persons were randomly assigned to each of the four drug doses. After a pre-study pulse rate was recorded for each individual, subjects were injected with the appropriate drug dose. One hour later, pulse rates were again recorded. The changes in pulse rates are listed in the accompanying table

Dose x_i	Pulse rates y_{ij}	n_i	\bar{y}_i	$\sum_j (y_{ij} - \bar{y}_i)^2$
1.5	20, 21, 19	3	20	2
2.0	16, 17, 17	3	16.666667	0.6666666
2.5	15, 13, 14	3	14	2
3.0	8, 10, 8	3	8.6666667	2.6666667

- [5] (a) Decompose SSE into the sum of squares due to the pure error, SSPE, and sum of squares due to the lack of fit, SSLF.

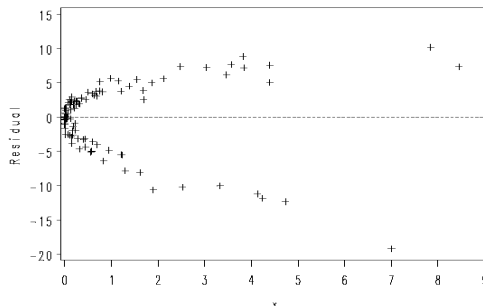
Hint: $SSPE = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 = 7.3333333$

$$\sum x_i = 27 \quad \sum x_i^2 = 64.5 \quad \sum y_i = 178 \quad \sum y_i^2 = 2854 \quad \sum x_i y_i = 373$$

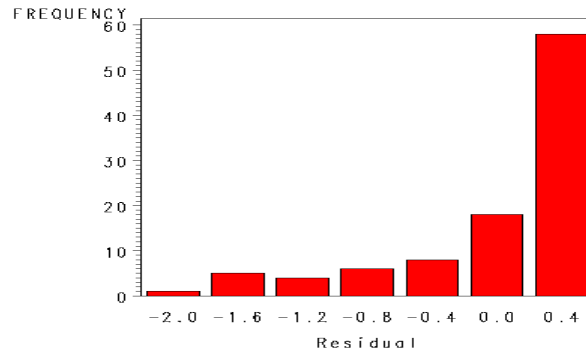
- [6] (b) Test whether the linear model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate. Use $\alpha = 0.05$.

3. State which violations of the SLR model (if any) are indicated by each of the following residual plots. Give reasons for your answer.

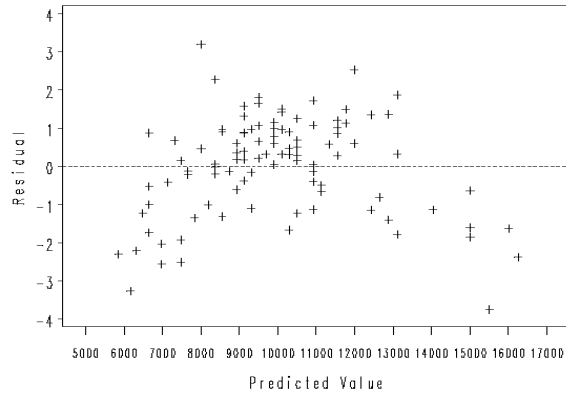
- [3] (a)



[3] (b)



[3] (c)



[3] (d)

