

Statistics

Note : For each problem please show all necessary steps!

1. One investor desires to know which of the two stocks is worth investing. The raw data and some summary statistics of the stock prices per share of these two stocks for the last two weeks are given below.

Stock A : 64.5, 65, 65.5, 64.5, 64.5, 62.5, 60, 58.5, 60.5, 60, 60, 62
 Stock B : 198, 195, 194, 195, 195, 190, 189, 186, 189, 185, 184, 185

N Obs	Variable	Range	Mean	Std Dev	Coefficient of Variation
12	STOCKA	7.00	62.45	2.494	3.9941680
	STOCKB	14.00	190.54	5.047	2.6487388

N Obs	Variable	Skewness	Kurtosis
12	STOCKA	-0.28257	-1.63833
	STOCKB	0.00132	-1.74488

- (a) Which stock price is more stable ? Why ? (10 pts.)
 (b) If this investor is a conservative investor, which stock would you suggest to this investor to invest ? (10 pts.)

2. A marketing research firm did two independent surveys of the proportions of viewers in a certain metropolitan area who watched 'MacGyver'. The first survey revealed that 35 of the 100 persons questioned were tuned to the show. In the second survey, 100 of the 350 persons contacted watched the show. Intuitively, what would be a reasonable way to combine the information in the two samples ? The researcher consider the following two estimators for the true proportion of people watching 'MacGyver',

$$(i) \frac{1}{2} \left(\frac{X_1}{n_1} + \frac{X_2}{n_2} \right) \quad (ii) \frac{X_1 + X_2}{n_1 + n_2}$$

where X_i is the persons watching 'Hill Street blues' in the sample $i, i=1,2$ and n_i is the sample size of i th sample.

- (a) Which one of the estimators is better for estimating the true proportion of people watching 'MacGyver'. (Hint : Unbiasedness and minimum variance) (20 pts)
 (b) Find the estimate for the true proportion of people watching 'MacGyver'. (5 pts)

3. The Management Science Department of Hercules Industries provides consulting services to the corporation's 17 divisions. Jeff Bannister, Vice President of Management Science Services, is currently in the process of **deciding whether or not to replace** the department's LTX-30 computer with a new model GTB-6 computer. His analysis indicates that the **change** would be **economically sound only if it would reduce the mean CPU time per job by more than 10 seconds**. To aid in his decision, he orders a sample of 15 jobs to be run on each computer and the CPU time for each run be determined. The following data were collected.

Job	CPU Time (seconds)	
	LTX-30	GTB-6
1	117	84
2	48	32
3	123	97
4	65	43
5	13	11
6	59	42
7	72	49
8	36	26
9	21	19
10	244	207
11	83	65
12	25	20
13	98	59
14	19	23
15	102	78

- (a) Five results from different analyses are

given below respectively.

Which **analyses** do you think are appropriate for answering the above question? **Why?**
Please explain your reasons? (10 pts)

- (b) Use the analysis you have chosen and $\alpha = .05$ to make conclusions. (Don't forget write down the null and alternative hypotheses. Do you need any assumptions?) (15 pts)

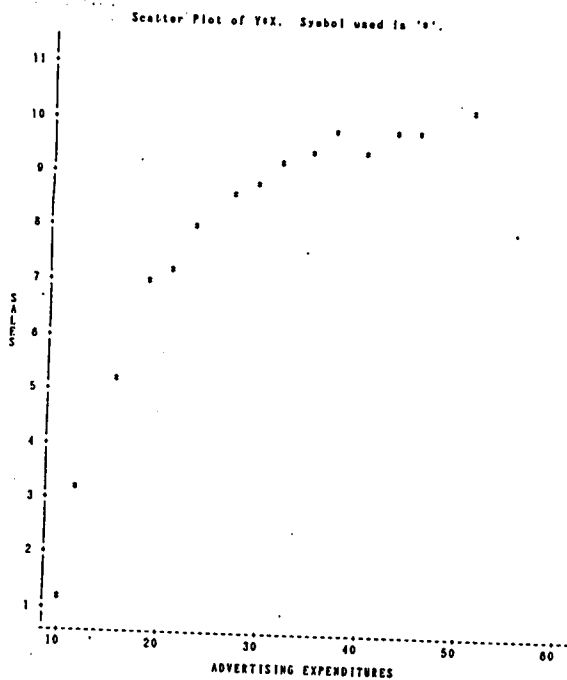
Analysis Methods	Test Statistics	degree of freedom	P-value (Two Sided)
A : Two sample independent T test (for unequal variances)	T = 0.4037	28	0.6895
B : Two sample independent T test (for equal variances)	T = 0.4037	27	0.6896
C : Paired T test	T = 2.3654	14	0.033
D : One way ANOVA	F = 0.16	1	0.6896
F : ANOVA from block design	F = 5.6	1	0.033

4. Mr. Bentley owns several furniture stores in a large metropolitan. He is interested in the relationship between his monthly advertising expenditure (X) and the corresponding monthly sales (Y). He suspects that sales will increase but after a certain point will slow down (that is, continue to increase but at a slower rate). In other words, after spending a certain amount on advertising, he will reach a point where there will be little gain in sales, even though he spends a much larger amount on advertising.

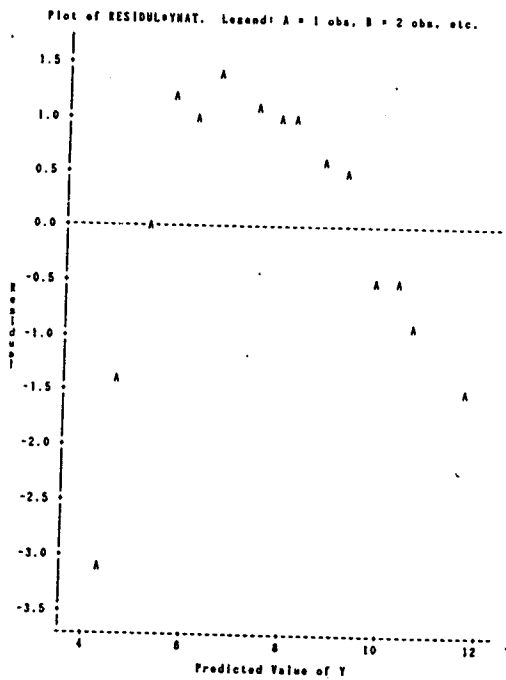
Data were gathered from the company records covering 15 nonconsecutive months of sales (in tens of thousands of dollars) and advertising expenditure (in hundreds of dollars).

- (a) From the **scatter plot**, do you think simple linear regression ($Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$) is appropriate for this data set to estimate the relationship between monthly advertising expenditure (X) and the corresponding monthly sales (Y)? **Why or Why not?** (5 pts)
- (b) What message can you tell from the **residual plot** obtained from $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$? (10 pts)
- (c) What message can you tell from the **stem-and-leaf plot** for residuals obtained from $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$? (7 pts)
- (d) Can you suggest a better model for this data set? (8 pts)

Scatter Plot



Residual Plot



Stem-and-Leaf Plot for residuals

Stem	Leaf	#
1	000124	6
0	056	3
-0	955	3
-1	54	2
-2		
-3	1	1

Month	Y (sales)	X (advertising expenditures)
1	6.9	18.1
2	8.5	27.3
3	1.2	10.1
4	9.4	34.8
5	3.2	11.8
6	5.2	15.0
7	8.0	22.9
8	7.2	20.4
9	8.8	29.2
10	10.2	50.5
11	9.8	37.3
12	9.3	40.2
13	9.8	43.1
14	9.8	45.0
15	9.2	31.5