

1. In a recently-published study, medical researchers report findings relating alcohol consumption to the risk of heart disease in people 65 years old or older. Below are three alternative designs for this study. Tell what kind of design each is (be as specific as possible). Be sure to justify your answers.

(a) **(10 points)** The research team interviews a random sample of people 65 years old or older to determine alcohol consumption patterns and history of heart disease.

*ANS: This is a sample survey, since they are taking samples from populations to summarize aspects of the populations.*

(b) **(10 points)** Investigators choose a random sample of patient medicare records. They classify the patients into groups based on alcohol consumption patterns indicated in the records. They then follow the patients for 5 years and compare the incidence of heart disease in the groups.

*ANS: This is a prospective observational study, since the groups are set up at the outset and the outcomes (heart disease or not) in each group are obtained subsequently.*

(c) **(10 points)** Investigators divide a sample of death certificates of medicare patients into two groups: those which list heart disease as a cause of death, and those which do not. They then compare the proportions of non-drinkers, moderate drinkers and heavy drinkers in each group.

*ANS: This is a retrospective observational study, since the groups are based on outcomes (heart disease or not), and the association with a suspected cause for the outcomes is sought (alcohol consumption patterns).*

2. **(10 points)** Suppose in problem 1, that researchers find that the rate of heart disease is significantly lower for moderate drinkers than for either non-drinkers or heavy drinkers. Can they conclude that moderate drinking causes a reduction in heart disease? Justify your answer.

*ANS: No, they cannot conclude cause-effect, since these are observational studies. Only a controlled experiment can establish cause-effect.*

3. Swimco Company's quarterly sales of swimming pools and accessories over a two year period are:

Year	Quarter	Sales (\$1000s)
1	1	12
	2	147
	3	103
	4	28
2	1	14
	2	166
	3	116
	4	33

(a) (10 points) Construct the most appropriate initial plot of these data from among those discussed in class.

ANS: See Figure 1.

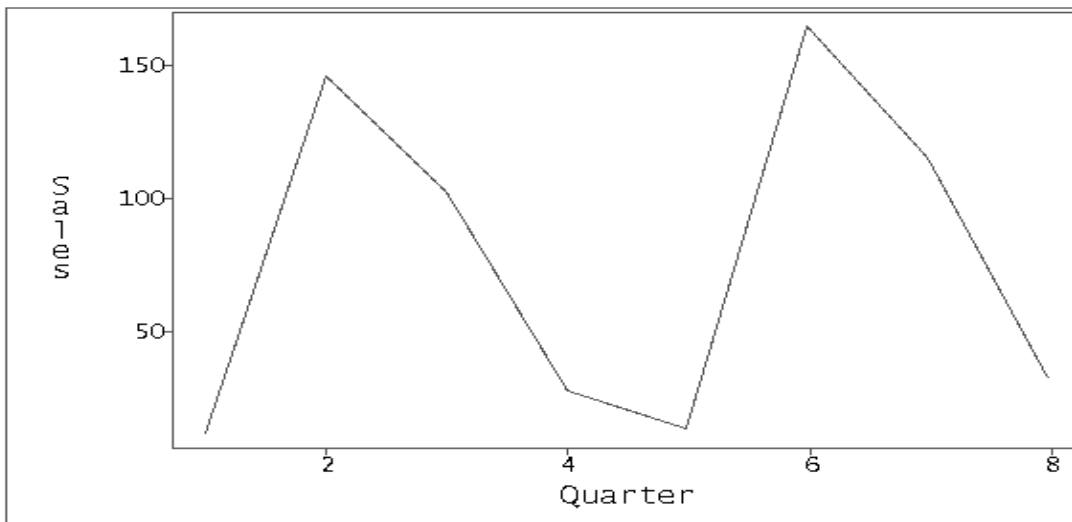


Figure 1: Line plot for question 3.

(b) (10 points) Choose the most appropriate moving average to remove the seasonal variation from these data.

ANS: A 4-term moving average is the most appropriate.

(c) (10 points) Apply the moving average chosen in (b) to these data, and use a plot to demonstrate that the seasonality has been removed.

ANS: The moving averages are:

Year	Quarter	Sales (\$1000s)	Moving average
1	1	12	
	2	147	
	3	103	
	4	28	72.5
2	1	14	73
	2	166	77.75
	3	116	81
	4	33	82.25

Figure 2 shows that the moving average has removed the seasonality.

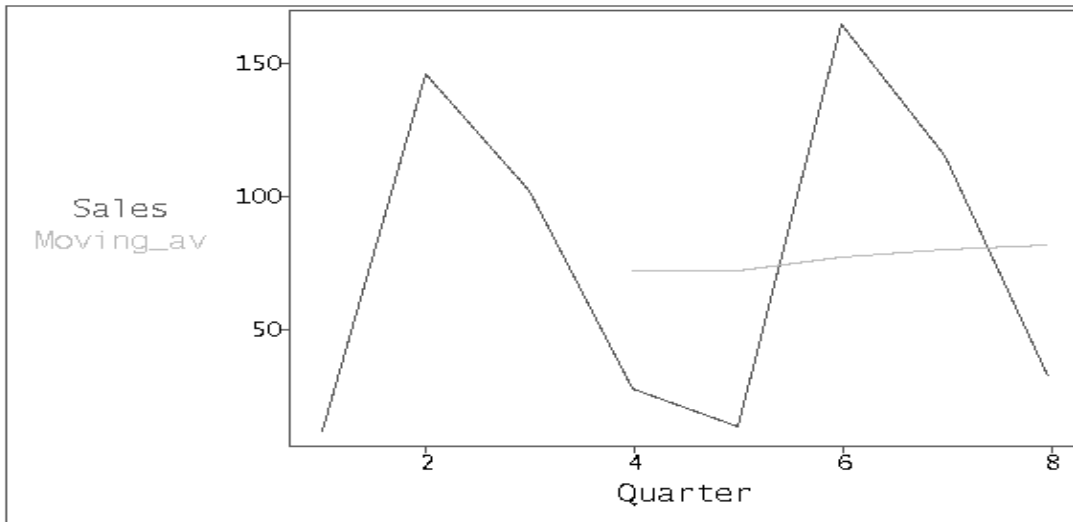


Figure 2: Line plot with moving average for question 3.

4. The ten natural lakes with largest surface areas are listed below.

Lake	Area	Lake	Area
Caspian Sea	143244	Aral Sea	13000
Lake Superior	31700	Lake Tanganyika	12700
Lake Victoria	26828	Lake Baykal	12162
Lake Huron	23000	Great Bear Lake	12096
Lake Michigan	22300	Lake Nyasa	11150

(a) (10 points) Calculate the five number summary of the surface areas.

ANS:  $A_- = 11150$ ,  $Q_1 = 12162$ ,  $Q_2 = 17650$ ,  $Q_3 = 26828$ ,  $A_+ = 31700$

(b) (10 points) Construct a box and whisker plot of the data. Are any values displayed as outliers?

ANS: Figure 3 shows the box and whiskers plot. The value for the Caspian Sea is plotted as an outlier.

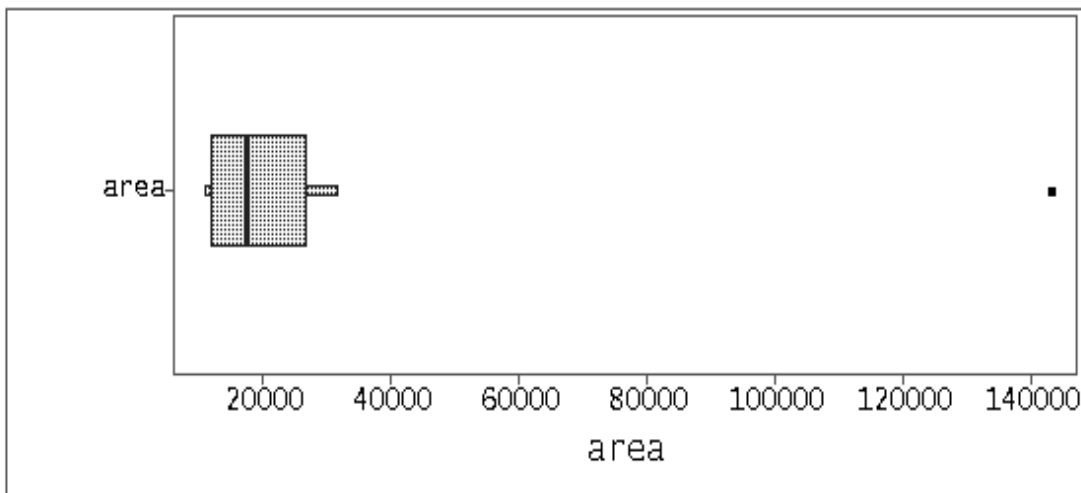


Figure 3: Box and whiskers plot for question 4.

(c) (10 points) Do the data give you any reason to suspect that the box and whiskers plot is not a good summary? Why?

ANS: No, there is no indication of multi-modality or other pattern that would make the plot unsuitable.