CHAPTER 1

INTRODUCTION TO DATA ANALYSIS

S1.1. Does it make a difference on what day of the week your car was made? Quality personnel for one of the major car manufacturers produced a quality index for each of 46 new cars of the same make and model manufactured over a two week period. The index took into account the number and severity of defects, and the quality of the fit and finish. The larger the index, the higher the quality. Figure 1 shows a plot of the quality index values stratified by the day of the week the car was produced.

a. What should the investigators have checked about the process prior to using this plot?

ANS: *Stationarity.* **(5 points)**

b. Assume the check you described in part a has been done and that the result was satisfactory. Analyze the data using the stratified plot. In particular, address the components of variation in the quality index and tell whether it appears there is a difference in the quality of cars produced on different days.

ANS: *There is substantial between-day variation, mainly attributable to lower quality scores for cars produced on Mondays and Fridays.* **(5 points)** *There is also noticeably greater variation (i.e., less consistency) in the quality index among cars produced on Fridays than on any other day of the week. The within variation for the other days is roughly equal.* **(5 points)** *The message here is that there will be more quality problems with cars produced on Mondays and Fridays. (5 points)*

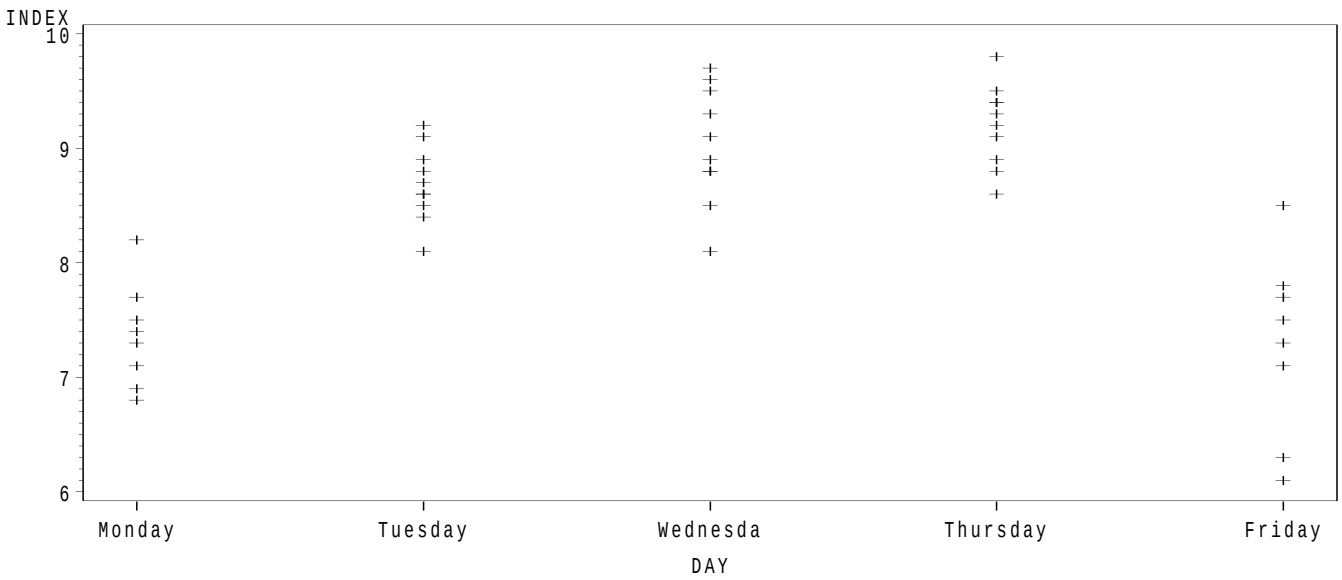

Figure 1: *Stratified plot of quality index by day, car quality data, Problem S1.1.*

S1.2. (20 points) Zeolites are aluminum and silicon crystals which are used as sieves and catalysts in many industries-especially petroleum refining. Figure 2 shows frequency histograms of relative yields, in percent, of 30 samples taken from each of five consecutive production batches of zeolites.

a. Is the process stationary? Why or why not?

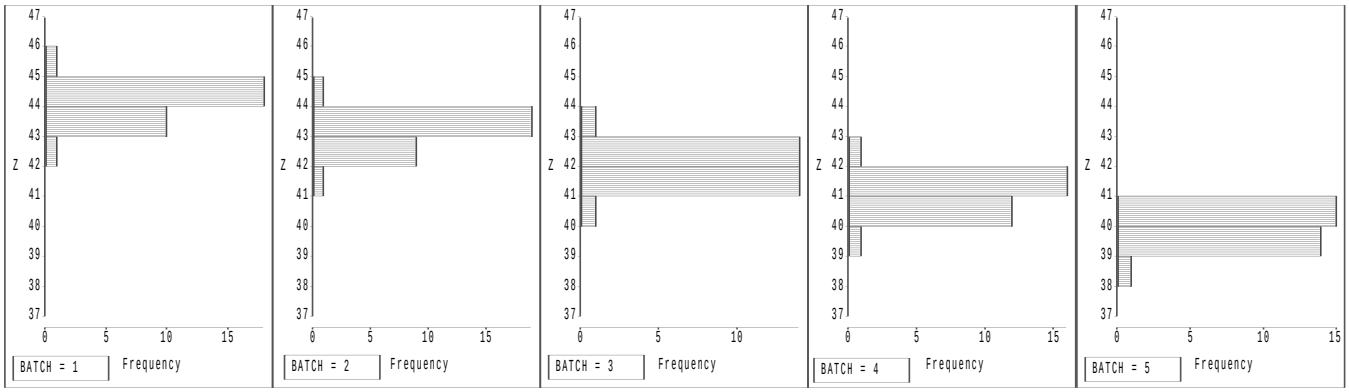

Figure 2: *Frequency histograms of relative yields, in percent, of five consecutive production batches of zeolites.*

ANS: *No. The location of the histograms is dropping, from around 44 in batch 1 to around 39.5 in batch 5.* (10 points)

b. Suppose all five batches' worth of zeolite yields were plotted on a single histogram. Which of the histograms in Figure 3 would result?

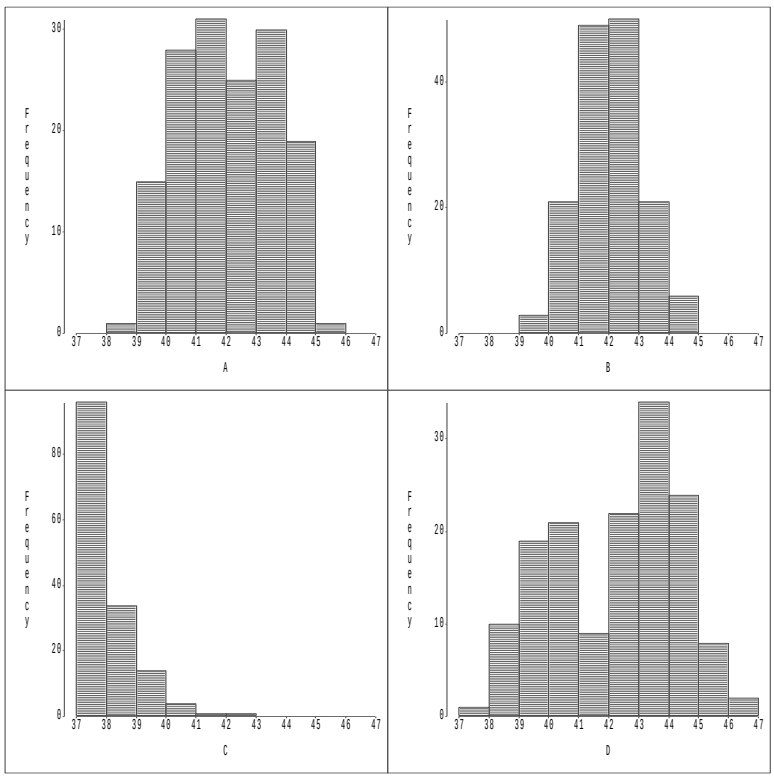

Figure 3: *Four possible Frequency histograms of relative yields, in percent, of production batches of zeolites.*

ANS: *Figure 3a would result.* (10 points)

S1.3. (30 points) Three regulars at a fitness center weigh themselves once per week. The deviations from their ideal

target weights for the last six weeks are:

	Week					
Weighee	1	2	3	4	5	6
Al	−2.9	1.7	−2.2	0.1	1.2	−2.7
W	−1.0	0.7	−1.8	−0.9	1.7	−2.1
Ralph	−0.5	−0.5	0.2	−0.1	0.1	0.2

Assuming the data are from stationary processes, construct a plot to assess the pattern of between and within variation. Analyze the plot.

ANS: *Figure 4 is a stratified plot to assess the pattern of between and within variation. It shows: (i) That Al and W have near-equal within variation, which is substantially larger than Ralph's; (ii) That W and Ralph are on target (i.e., their weight deviations are centered at or near 0), while Al is on the light side, with deviations centered slightly below 0.*

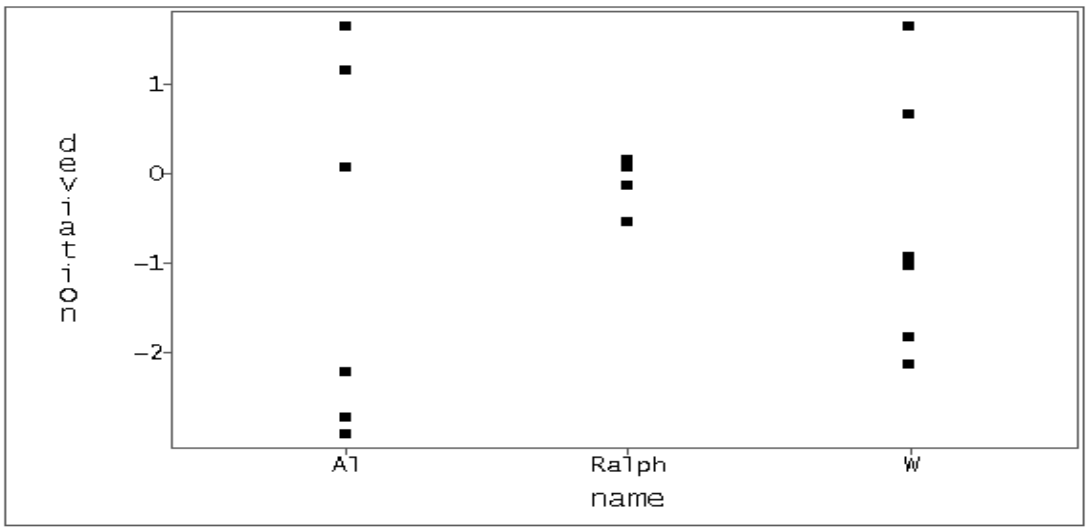

Figure 4: *Stratified plot for test problem 3.*

S1.4. Table 1 lists the postwar olympic winning times in the 100 meter dash.

Year	Time	Year	Time
1948	10.3	1976	10.16
1952	10.4	1980	10.25
1956	10.6	1984	9.99
1960	10.2	1988	9.92
1964	10.0	1992	9.96
1968	9.95	1996	9.84
1972	10.14	2000	9.87

Table 1: *Postwar olympic winning times in the 100 meter dash.*

a. **(15 points)** Draw an appropriate graphical summary of these data. Explain why your summary is appropriate. What conclusion do you draw from your summary?

ANS: *Figure 5 is a line plot of the times versus year. It shows a downward trend, indicating the the process is not stationary. A summary such as a histogram or boxplot would not be appropriate because they assume stationarity.*

Figure 5: *Line plot of modern olympic winning times in the 100 meter dash versus year.*

b. **(10 points)** Jesse claims that the data are well summarized by their mean $\overline{y} = 10.1057$ and standard deviation $s = 0.2205$. Do you agree? Give your reasons.

ANS: *These summaries assume stationarity and are therefore not appropriate because of the nonstationarity of the data.*

S1.5. **(15 points)** A fastener manufacturer wants to conduct a gage R&R study to assess the performance of a gage for measuring the lengths of 3/8" plastic rivets. To do so, they have three operators measure the same rivet ten times in succession using the same gage. The data are:

	Operator		
Order	1	2	3
1	.3730	.3741	.3769
2	.3739	.3752	.3769
3	.3744	.3748	.3766
4	.3747	.3751	.3762
5	.3750	.3744	.3759
6	.3745	.3749	.3756
7	.3755	.3753	.3758
8	.3756	.3749	.3752
9	.3759	.3752	.3749
10	.3760	.3747	.3747

What is the first graph you will produce to examine these data? Produce the graph and draw appropriate conclusions, giving your reasons.

ANS: *A stratified line plot to assess stationarity of the measurements versus order. The plot is shown in Figure 6. It shows that the measurements taken by operators 1 (increasing trend) and 2 (decreasing trend) are not stationary.*

S1.6. **(10 points)** In a gage R&R study, the repeated measurements produced by one operator using a single gage, are in order taken:

2.5, 2.4, 2.7, 2.8, 2.8, 2.9, 3.1, 3.0, 3.3, 3.4, 3.4, 3.5

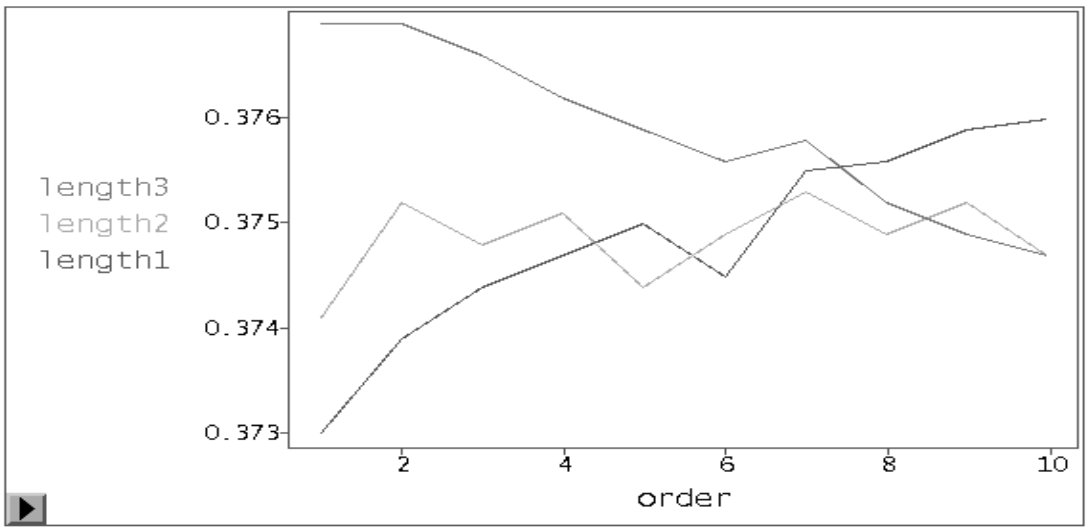

Figure 6: *Stratified line plot of rivet length measurements.*

Is a box and whiskers plot an appropriate summary of these data? Why or why not? If a box and whiskers plot is appropriate, construct and interpret the box and whiskers plot for these data. If a box and whiskers plot is not appropriate, construct and interpret an appropriate plot.

ANS: *A box and whiskers plot is not appropriate since the data are not from a stationary process. This is shown by an appropriate plot, the line plot shown in Figure 7. The increasing trend shows that the measurement process is nonstationary.*

Figure 7: