1. Infection with an apparently harmless, newly recognized virus called hepatitis G, seems to interfere with HIV, slowing its progress and prolonging survival of AIDS patients, according to an Associated Press article in the September 6, 2001 edition of the Boston Globe. A study by researchers at the Iowa City Veterans Affairs Medical Center and the University of Iowa looked at 362 HIV-infected patients treated between 1988 and 1999. The researchers found that 144 of these patients were also infected with hepatitis G. Of the 144 who were infected with hepatitis G, 41 died during four years of follow-up compared with 123 of the 218 who were not infected with hepatitis G.

   (a) **(10 points)** What kind of study is this? Be as explicit as possible and justify your answer.

   ANS: *It is either a prospective or retrospective observational study.*

   As described, it seems to be a prospective observational study. HIV patients were classified as being infected with hepatitis G or not, then followed for four years to see if they died or not.

   However, if the patients were first classified into whether they died or did not die during the 4 year follow-up, and then the groups were compared for incidence of hepatitis G, it was a retrospective study.

   At any rate, to get credit, a valid justification must be given.

   (b) **(10 points)** The results look impressive. Can the experimenters conclude that infection with hepatitis G causes a reduction in the HIV mortality rate? Why or why not?

   ANS: *No, they cannot conclude that infection with hepatitis G causes a reduction in the HIV mortality rate, because this is not a controlled experiment*

   (c) **(10 points)** If your answer to (b) was that they cannot conclude that hepatitis G causes a reduction in the HIV mortality rate, tell how you would design a study from which such a conclusion is possible.

   ANS: *You would have to conduct a designed experiment (if this is possible). From among volunteer HIV patients who were uninfected by the hepatitis G virus, you would randomly choose a group to infect with hepatitis G, leaving the other group as a control. You would then follow the patients over a number of years to assess differences in death rates.*

2. The questions below refer to the frequency histogram of a set of 24 data values shown in Figure 1.

(a) **(10 points)** From the histogram, obtain quartiles Q_1, Q_2 and Q_3 of the data. Tell how you got your results. (Hint: Use the book's definitions of quartiles. Do not try to use the computational formulas.)

ANS: *Take Q_1 to be 2 because that is the point that divides the total area of the bars in in the ratio 1/4 to 3/4. Take Q_2 to be 3 because that is the point that divides the total area of the bars in in the ratio 1/2 to 1/2. Take Q_3 to be 5 because that is the point that divides the total area of the bars in in the ratio 3/4 to 1/4.*

(b) **(10 points)** Will the mean be greater than, equal to, or less than the median? Why?

ANS: *Greater than the median, since the histogram is skewed right (balance point is pulled to the right).*

(c) **(5 points)** Calculate the mode, in the manner described in the text.

ANS: *The mode is the center of the modal bar: 2.5.*

(d) **(10 points)** Name one measure of location and one measure of spread that appropriately summarize the pattern of variation in these data. Why did you choose these measures?

ANS: *Either the mode or Q_2 for location, and IQR for spread, because of the skewness of the data.*

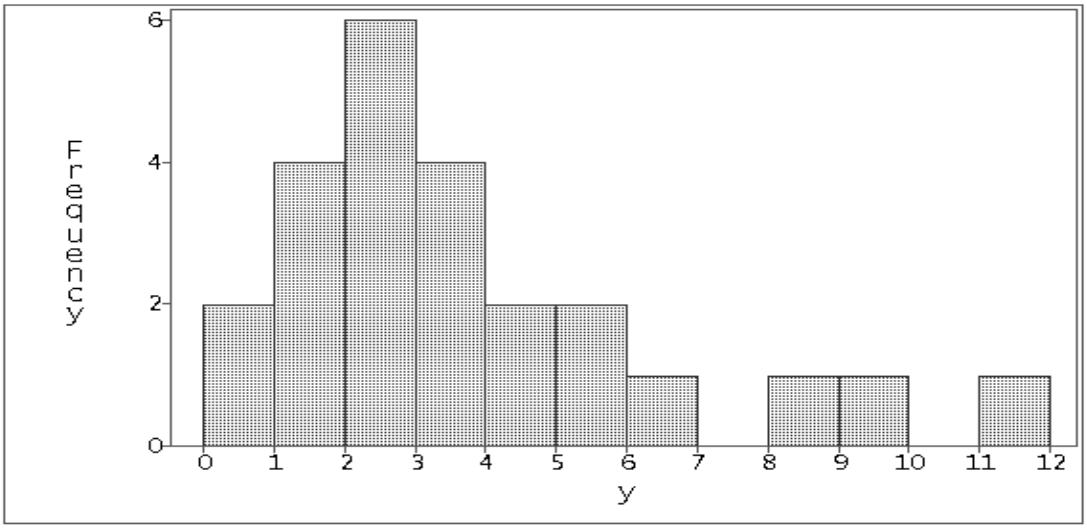

Figure 1: *Frequency histogram of 24 data values.*

3. A manufacturer of roofing shingles wants to compare the performance of shingles with two different types of backings in field tests. To do so, they randomly select 30 communities around the country. In each, they randomly select a single-family house among those volunteered by their owners in response to an ad for a "free roof." They randomly select half the houses to receive one type of shingle and roof the rest with the second type. Various measures of the condition of each roof are obtained over a period of years.

   (a) **(10 points)** What kind of study is this? Be as specific as possible and justify your answer.

   ANS: *It is a controlled experiment, since treatments (roof types) are applied to experimental units (houses) to observe a response (roof condition measures).*

   (b) **(10 points)** Explain how blocking could be used to improve the results. Be sure to tell what the blocks are and why the blocking would be expected to improve the results.

   ANS: *Various answers could be given. Here are two: (1) Put both shingle types on each roof. Then each house is a block. This will eliminate variation due to differences in houses and especially in the locations of communities. (2) In each community, select two houses and randomly assign one to receive one type of shingle and the other to receive the other type of shingle. Then each community is a block. This will eliminate variation due to differences in the locations of communities.*

4. **(15 points)** A gage R & R study was conducted on a weighing process. Three operators weighed a 1 pound weight five times each using the same scale. The data are (deviations, in pounds, from 1 pound):

Operator		
1	2	3
−0.007	−0.004	0.014
−0.008	0.005	0.022
−0.001	0.003	0.007
0.002	0.006	0.020
0.005	−0.001	−0.002

Assuming the data are from a stationary process, analyze the variation in the measurement process in terms of repeatability and reproducibility. Be sure to include an appropriate graph.

ANS: *Figure 2 is a plot of the deviations stratified by operator. It shows that operator 3 has both a repeatability problem (i.e., wide within variation) and a reproducibility problem (i.e., this operator's measurements are higher on average than those of the other two operators). (5 points for the graph)*

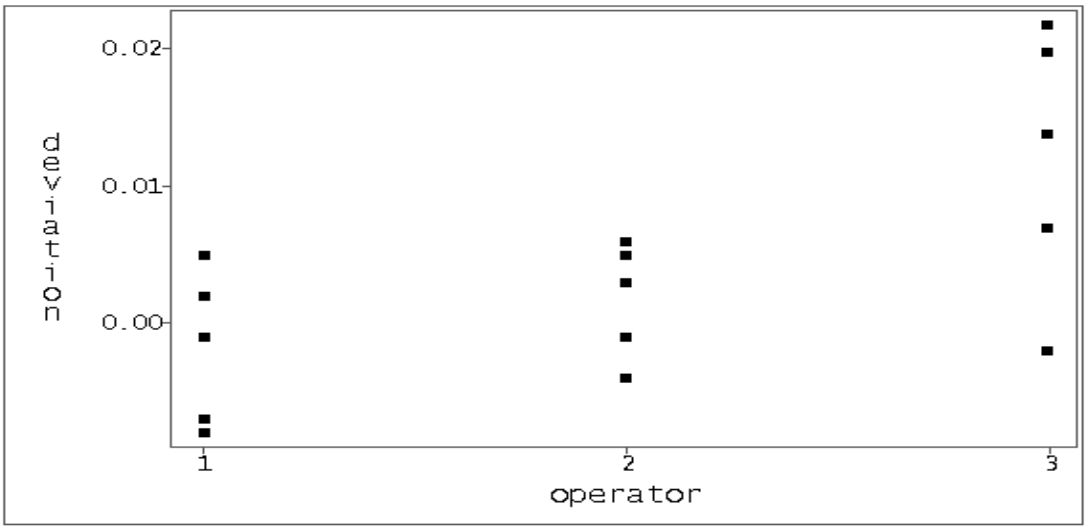

Figure 2: *Stratified plot of deviations in weight from 1 pound by operator.*