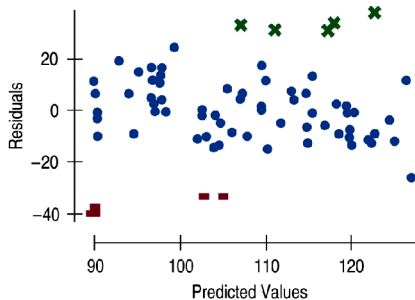
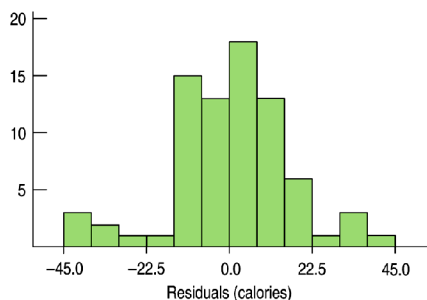


Chapter 9: Regression Wisdom

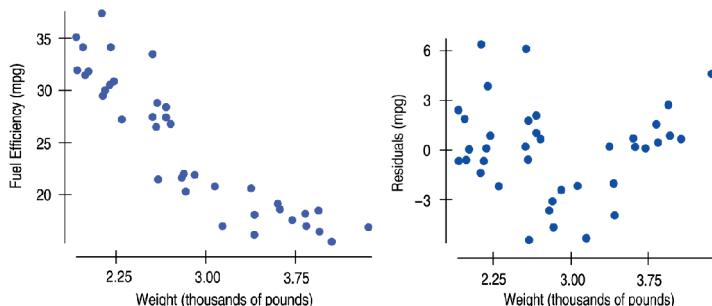
Residuals

To determine whether a linear model is appropriate, we examine the _____. It is a good idea to look at both a _____ of the residuals and a scatterplot of the _____ versus the _____. If the histogram of the residuals has multiple _____, that may indicate that there are subgroups within the set of data. If a linear model is appropriate, the histogram should look approximately _____ and the scatterplot of residuals should show _____.



If we see a curved relationship in the residual plot, the linear model is _____.

Another type of residual plot shows the residuals versus the explanatory variable. Note that the scale on the horizontal axis is exactly the same as that of the original scatterplot of x and y .



When using a model to predict values of the response variable, two types of predictions can be made: _____ or _____. Making predictions within the given domain of x -values is called _____. Making predictions outside the given domain of x -values is called _____. Only _____ is reliable. _____ is unreliable, because patterns can change abruptly. In the above example, we would be extrapolating if we use a model to predict the fuel efficiency of cars weighing more than _____ thousand pounds or less than _____ thousand pounds.

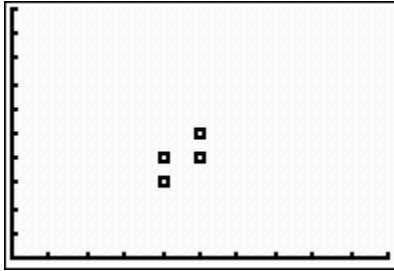
Even if a linear model is appropriate, remember that association does not imply _____. There may be _____ variables that account for the association. Also, if we're plotting _____ values (like the mean) instead of individual values, the association may appear stronger than it really is.

Outliers, Leverage, and Influential Points

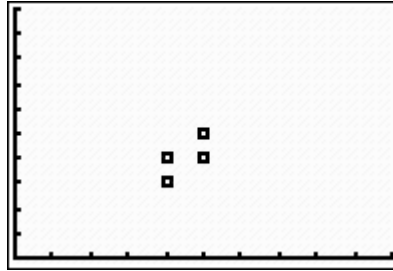
When examining a scatterplot it is important to look for any unusual points. There are three ways a point can be considered unusual.

1. A point is unusual if it stands away from the others. This is called _____
2. A point is unusual if its x -value is far from the mean of all the x -values. This is called _____
3. A point is unusual if, when omitted, it significantly changes the linear model. This is called _____

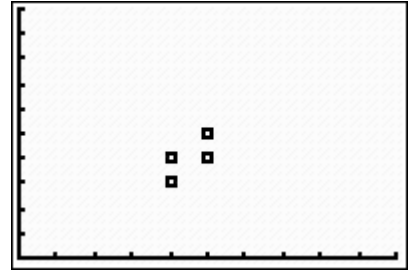
For each scatterplot, add a point with the given characteristics (if possible):



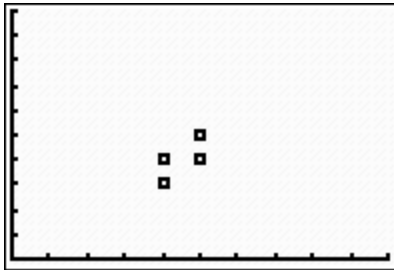
Outlier, influential,
high leverage



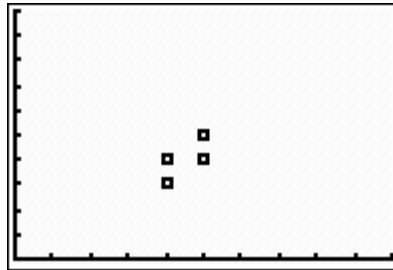
Outlier, influential
low leverage



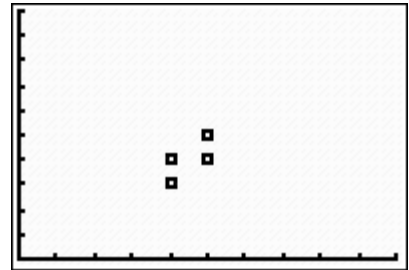
Outlier, not influential
high leverage



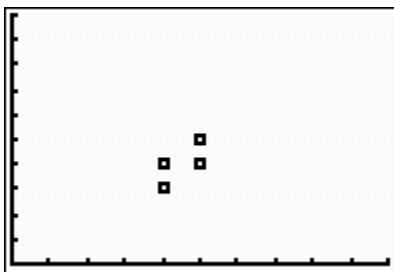
Outlier, not influential,
low leverage



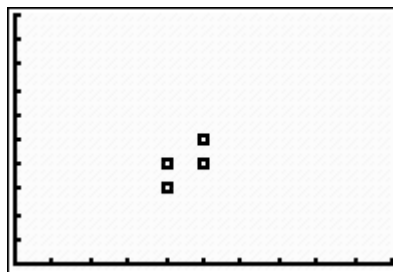
Not an outlier, influential
high leverage



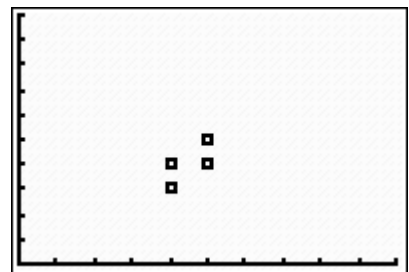
Not an outlier, influential
low leverage



Not an outlier, not influential,
high leverage



Not an outlier, not influential
low leverage



Outlier, no leverage

If there is an unusual point in your scatterplot, it must be mentioned. It is then wise to examine two models, one with the point included, and one with the point omitted.