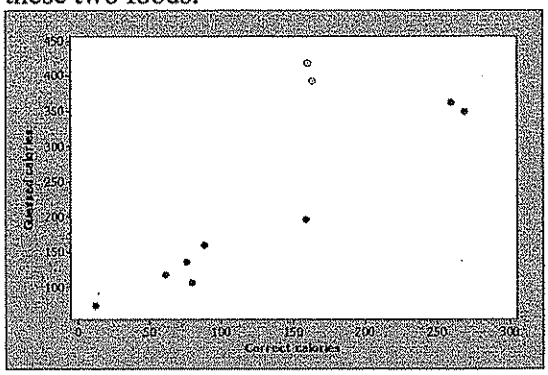
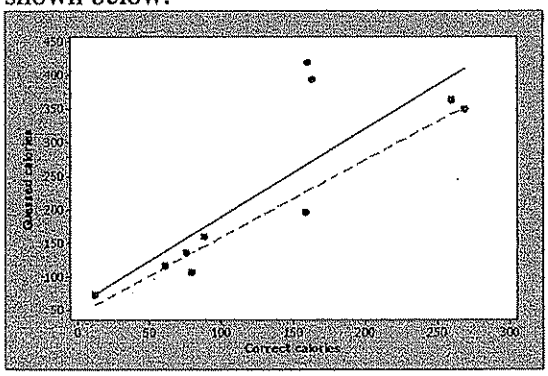
HW p238-239

3.59 (a) The scatterplot is shown below. The guessed values are much higher than expected for these two foods.

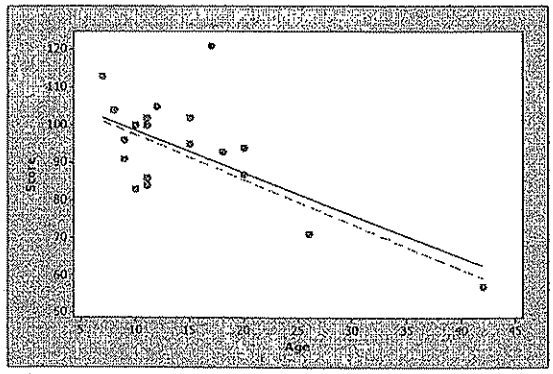


(b) The regression line for predicting y = guessed calories from x = actual calories using all points is $\hat{y} = 58.59 + 1.3036x$ $\left(r^2 = 0.68\right)$. Excluding spaghetti and snack cake: $\hat{y} = 43.88 + 1.14721x$ $\left(r^2 = 0.968\right)$. (c) A scatterplot with the two separate regression lines is shown below.



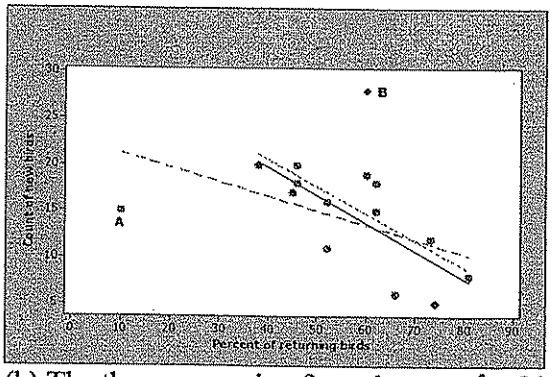
The two removed points could be called influential, in that when they are included, the regression line passes above every other point; after removing them, the new regression line passes through the "middle" of the remaining points.

3.60 (a) Without Child 19, the least-squares regression line for predicting y = score from x = age is $\hat{y} = 109.305 - 1.1933x$.



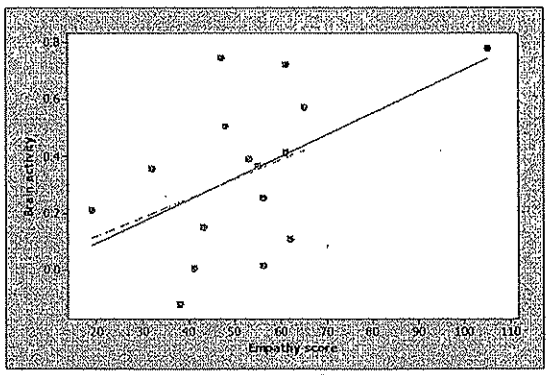
The slope and intercept change slightly when Child 19 is removed, so this point does not appear to be extremely influential. (b) With all children, $r^2 = 0.410$; without Child 19, $r^2 = 0.572$. r^2 increases because more of the variability in the scores is explained by the stronger linear relationship with age. In other words, with Child 19's high Gesell score removed, there is less variability around the regression line.

3.61 (a) A scatterplot with the two new points is shown below. Point A is a horizontal outlier; that is, it has a much smaller x-value than the others. Point B is a vertical outlier; it has a higher y-value than the others.



(b) The three regression formulas are: $\hat{y} = 31.9 - 0.304x$ (the original data); $\hat{y} = 22.8 - 0.156x$ (with Point A); $\hat{y} = 32.3 - 0.293x$ (with Point B). Adding Point B has little impact. Point A is influential; it pulls the line down, and changes how the line looks relative to the original 13 data points.

3.62 (a) *Who?* The individuals are 16 couples in their mid-twenties who were married or had been dating for two years. *What?* The variables are empathy score (a quantitative measure of empathy from a psychological test) and brain activity (a quantitative variable reported as a fraction between -1 and 1). *Why?* The researchers wanted to see how the brain expresses empathy. In particular, they were interested in checking if women with higher empathy scores have a stronger response when their partner has a painful experience. *When, where, how, and by whom?* The researchers zapped the hands of the men and women to measure brain activity, presumably in a lab, doctor's office, or hospital. The results appeared in *Science* in 2004 so the data were probably collected shortly before publication of the article. (b) Subject 16 is influential on the correlation. With all subjects, $r = 0.515$; without Subject 16, $r = 0.331$. (c) Subject 16 is not influential on the least-squares regression line (see the scatterplot below).



The regression lines are: $\hat{y} = -0.0578 + 0.0076x$ (with all subjects) and $\hat{y} = -0.0152 + 0.0067x$ (without Subject 16).