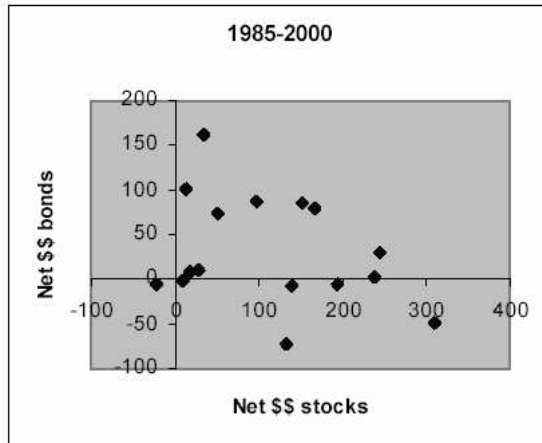


HW 2

2.12, 2.14, 2.18 (ex02-18.xls), 2.26, 2.28 (ex02_18.xls), 2.48 (ta02_001.xls), 2.68 (ex02_013.xls), 2.86, 2.88, 2.90

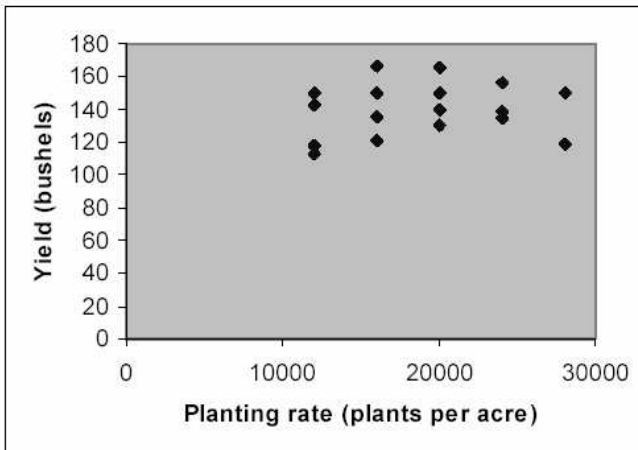
2.12



This scatterplot shows a negative trend with a weak linear relationship.

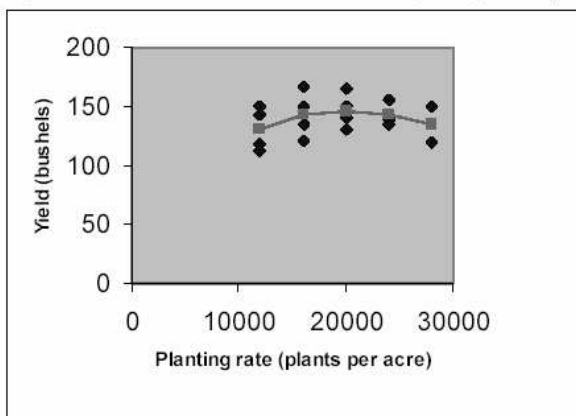
2.14 a) *Planting rate* is the explanatory variable.

b)



c) This scatterplot shows a slight nonlinear pattern. The yield increases from 12,000 plants per acres to 20,000 plants per acre and then starts to decrease beyond that level.

d) I would recommend a rate of 20,000 plants per acre. This resulted in the highest average yield.



2.18 a)



b) California stands out on the scatterplot because it has an unusually large number of Target stores compared to Wal-Mart stores. It appears that in most states the number of Wal-Mart stores is greater than Target stores. California does not follow this trend. c) This relationship is a weak positive relationship. It has a slight linear appearance.

2.26 a)



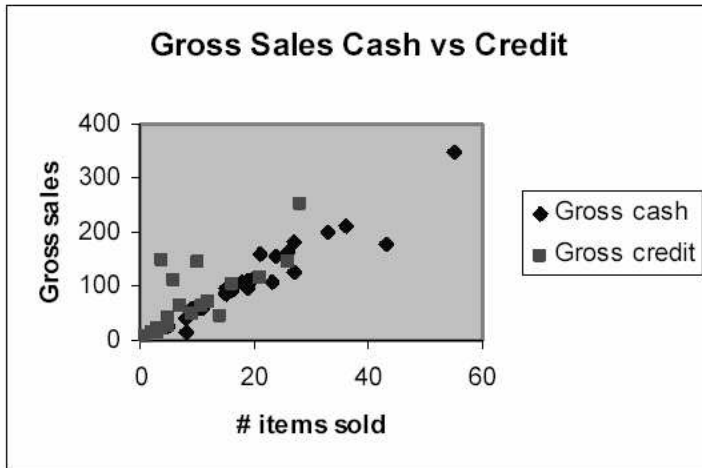
b) $r = 0.825$. This value of r makes sense based on the above scatterplot. The points show a fairly strong relationship but the two foods, spaghetti and snack cake; do not lie close to the other points. The fact that every guess was higher than the correct number of calories does not influence the correlation. If every guess were exactly 100 calories higher than the correct one then the correlation would be 1.0.

d) The correlation coefficient is now 0.984. The correlation increased because after removing the spaghetti and snack cake, the remaining eight points fall very close to a straight line.

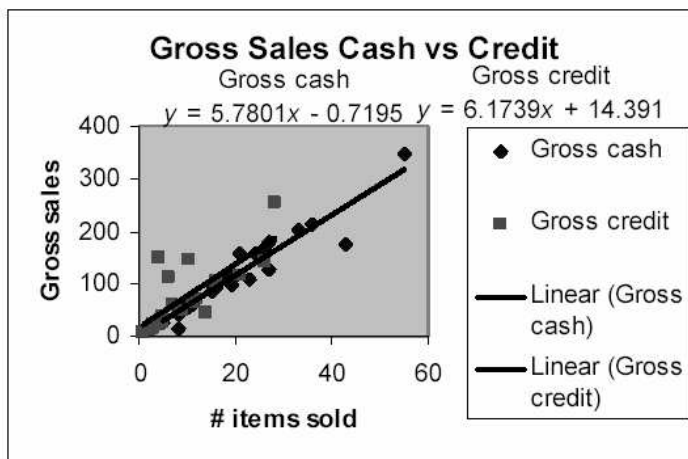


2.28 a) $r = 0.659$. b) The correlation should increase. California is an outlier and does not fall on the linear pattern that the rest of the data do. $r = 0.746$. The correlation increased for the reason stated. c) The correlation for all fifty states will most likely decrease if Texas is removed. Texas helps define the linear pattern observed on the scatterplot. Without this data point, the linear relationship does not appear as strong. $r = 0.584$.

2.48 a)

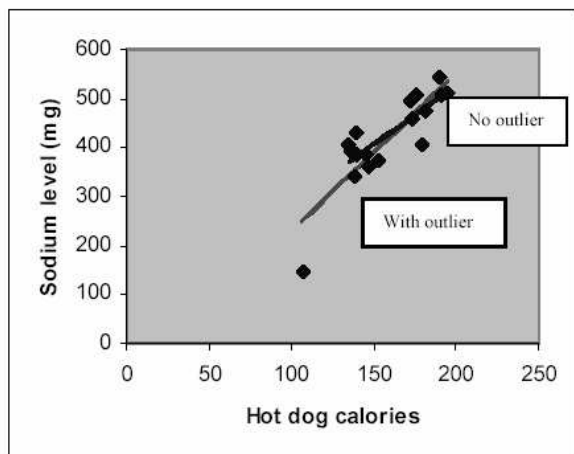


b)



The slope for the gross cash sales regression line is: 5.7801. The slope for the gross credit sales regression line is: 6.1739. The larger slope for gross credit sales means that on average when people purchase with credit cards they tend to purchase more expensive items.

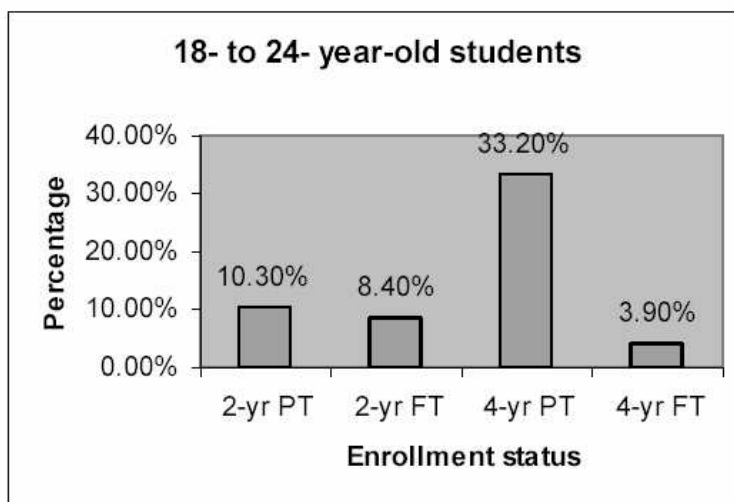
2.68 The regression line with the outlier is: $y = -91.2 + 3.212x$. The regression line without the outlier is: $y = 46.9 + 2.4x$. The scatterplot below shows the two lines. The outlier was not extremely influential. In fact the correlation coefficient was slightly higher with the outlier than without the outlier because the outlier point was part of the overall linear pattern.



b) The sodium level would be between 390 and 406 mg.

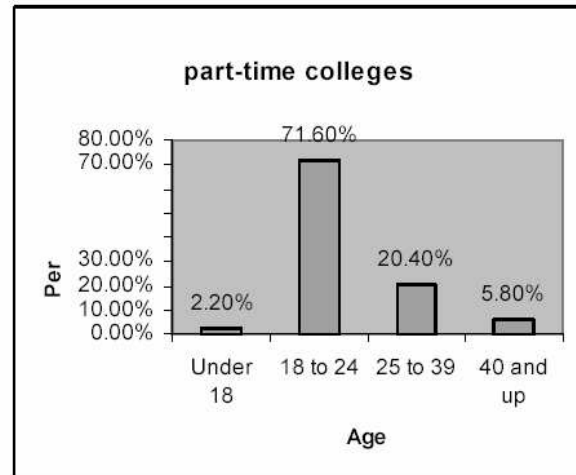
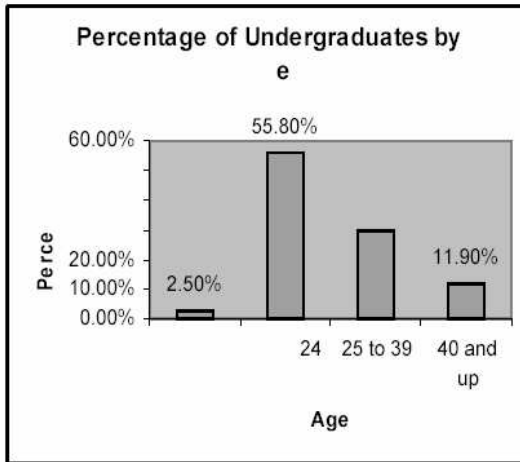
2.86 a) 14,340,000. b) 55.8%.

c)



d) This age group makes up over half of the students enrolled in a college, and over 30% of the undergraduates enrolled as 4-year part-time students come from this age group.

2.88 a) By age the counts of undergraduates are: Under 18–353,000, 18 to 24–8,001,000, 25 to 39–4,272,000, 40 and up–1,716,000.



c) The main difference in the two age distributions is that while the majority of students are in the age group 18 to 24, almost 72% of 2-year part-time students are in this age group while only 56% of all undergraduates are in this group. **d)** The reason the total in this column is different than the actual sum is most likely due to round off error.

2.90 a) 55.8%. **b)** 48.5%. **c)** 74.5%.