Chapter 9

21. [4 points] Interest rates.
a) $r = 0.88$. The correlation between rate and year is +0.88, since the scatterplot shows a positive association.
b) According to the model, interest rates during this period increased at about 0.25% per year, starting from an interest rate of about 0.64% in 1950.
c) The linear regression equation predicting interest rate from year is Rate = 0 640282 + 0 247637 (Year -1950) = 13 022

According to the model, the interest rate is predicted to be about 13% in the year 2000.
d) This prediction is not likely to have been a good one. Extrapolating 20 years beyond the final year in the data would be risky, and unlikely to be accurate

25. [4 points] Gestation.
a) The association would be stronger if humans were removed. The point on the scatterplot representing human gestation and life expectancy is an outlier from the overall pattern and detracts from the association. Humans also represent an influential point. Removing the humans would cause the slope of the linear regression model to increase, following the pattern of the non-human animals much more closely.
b) The study could be restricted to non-human animals. This appears justifiable, since one could point to a number of environmental factors that could influence human life expectancy and gestation period, making them incomparable to those of animals.

c) The correlation is moderately strong. The model explains 72.2% of the variability in gestation period of non-human animals.

d) For every year increase in life expectancy, the model predicts an increase of approximately 15.5 days in gestation period.

$Ges\hat{\ }t \cdot$ **=**- 39.5172 + 15.4980(*LifEx)*

$Ge\hat{\ }st = 270.4428$

According to the linear model, monkeys with a life expectancy of 20 years are expected to have gestation periods of about 270.4 days. Care should be taken when assessing the accuracy of this prediction. First of all, the residuals plot has not been examined, so the appropriateness of the model is questionable. Second, it is unknown whether or not monkeys were included in the original 17 nonhuman species studied. Since monkeys and humans are both primates, the monkeys may depart from the overall pattern as well.

27. [4 points] Elephants and hippos.
a) Hippos are more of a departure from the pattern. Removing that point would make the association appear to be stronger.

b) The slope of the regression line would increase, pivoting away from the hippos point.

c) Anytime data points are removed, there must be a justifiable reason for doing so, and saying, "I removed the point because the correlation was higher without it" is not a justifiable reason.

d) Elephants are an influential point. With the elephants included, the slope of the linear model is 15.4980 days gestation per year of life expectancy. When they are removed, the slope is 11.6 days per year. The decrease is significant.

Chapter 10

5. [5 points] Models.
a) $y = 1.2 + 0.8\ x$
$y = 1.2 + 0.8\ (-2)$
$y = -0.04$

b) $\log y = 1.2 + 0.8\ x$
$\log y = 1.2 + 0.8\ (-2)$
$\log y = -0.04$

c)) $\operatorname{sqrt}(y) = 1.2 + 0.8\ x$
$= -0.04$
$y = (-0.04)^2$

d) $1/y = 1.2 + 0.8\ x$
$1/y = 1.2 + 0.8\ (-2) = -1.04$
$y = -1/0.04$

e) $y = 1.2\ x^{0.8}$
$y = 1.2\ (-2)^{0.8}$

13. [2 points] GDP.
a) Although more than 97% of the variation in GDP can be accounted for by the model, the residuals plot should be examined to determine whether or not the model is appropriate.
b) This is not a good model for these data. The residuals plot shows curvature.

19. [4 points] Baseball salaries.

a) The scatterplot of year versus highest salary is moderately strong, positive and possibly curved, yet straight enough to try fitting a linear model. The highest salary has generally increased over the years.

b) The linear model, $\widehat{Sal\,ary} = 1802.94 - 0.9097(Year)$, is not an appropriate model. The residuals plot shows a strong bend.

c) Re-expression using the logarithm of the salaries straightens the plot significantly.

d) $\log(\widehat{Sal\,ary}) = -98.509 + 0.04985(Year)$

is a good model for predicting the highest salary from the year. The residuals plot is scattered and $R^2 = 87.9\%$, indicating that 87.9% of the variability in highest salary can be explained by the model.