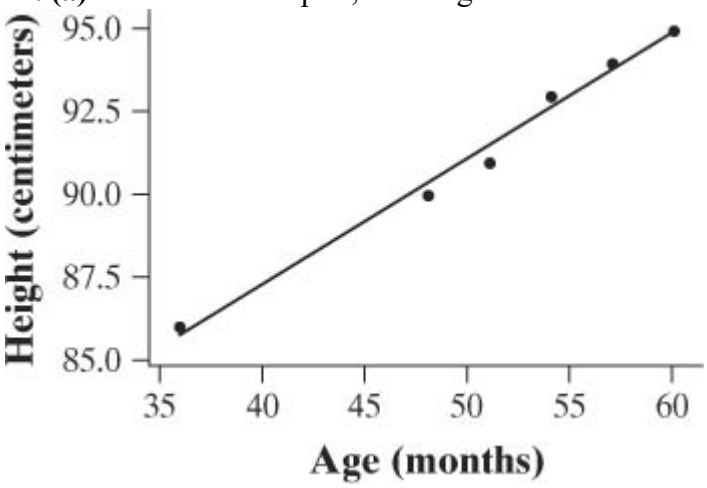
Key AP Stats Midterm Review Chapters 3, 4, & 5

Chapter 3

1. D 2. E 3. C 4. A 5. A 6. C 7. B 8. E 9. B 10. C

11. (a) Here is a scatterplot, with regression line added.



(b) The regression line for predicting y = height from x = age is $\hat{y} = 71.95 + 0.3833x$ **(c)** At age 480 months, we would predict Sarah's height to be $\hat{y} = 71\ 95 + 0.3833(480) = 255.934$ cm. Herheight in inches would be $\frac{255.934}{2.54} = 100.76$ in. **(d)** This height is impossibly large (about 8 feet, 4 inches) because we used extrapolation. Obviously, the linear trend does not continue all the way out to 40 years. Our data were based only on the first 5 years of life.

12. (a) The unusual point is the one in the upper-right corner with isotope value about −19.3 and silicon value about 345. This point is unusual in that it has such a high silicon value for the given isotope value. **(b)** (i) If the point were removed, the correlation would increase because this point does not follow the linear pattern of the other points. (ii) And since this point has a higher silicon value, if it were removed, the slope of the regression line would decrease and the y intercept would increase.

13. (a) The regression equation is $\hat{y} = 92.29 - 0.05762x$ The variable y represents the percent of the grass burned, and x represents the number of wildebeest. **(b)** The slope of the regression line suggests that for every increase of 1000 wildebeest (this is a 1 unit increase in x since x is measured in terms of 1000s of wildebeest), we predict that the percent of grass area burned will decrease by about 0.058. **(c)** $r = -0.804$. There is a moderately strong, negative linear relationship between wildebeest abundance and percent of grass area burned. **(d)** The linear model is appropriate for describing the relationship between wildebeest abundance and percent of grass area burned. The residual plot shows a fairly "random" scatter of points around the "residual = 0" line. There is one large positive residual at 1249 thousand wildebeest. Since $r^2 = 0.646$, 64.6% of the variation in percent of grass area burned is explained by the least-squares regression of percent of grass area burned on wildebeest abundance. That leaves 35.4% of the variation in percent of grass area burned unexplained by the linear relationship.

Chapter 4

1. C 2. E 3. D 4. C 5. B 6. B 7. D 8. D 9. D 10. B 11. D

12. **(a)** The experimental units are the acacia trees. The treatments are placing either active beehives, empty beehives, or nothing in the trees. The response variable is the damage caused by elephants to the trees. **(b)** Randomly assign 24 of the acacia trees to have active beehives placed in them, 24 randomly to have empty beehives placed in them, and the remaining 24 to remain empty. To do this, assign the trees numbers from 01 to 72 and use a random number table to pick 24 two-digit numbers in this range. Those trees will get the active beehives. The trees associated with the next 24 two-digit numbers will get the empty beehives, and the remaining 24 trees will remain empty. Compare the damage caused by elephants to the trees with active beehives, those with empty beehives, and those with no beehives.

13. **(a)** It is not a simple random sample because not all samples were possible. For instance, given their method, they could not have had all respondents from the East Coast. **(b)** One adult was chosen at random to control for lurking variables. Perhaps the household members who generally answer the phone have a different opinion from those who don't generally answer the phone. **(c)** There was undercoverage in this survey. Those who do not have telephones and those who have only cell phones were not part of the sampling frame. So their opinions would not have been measured. Since cell-phone-only users tend to be younger, the results of the survey may not accurately reflect the entire population's opinion.

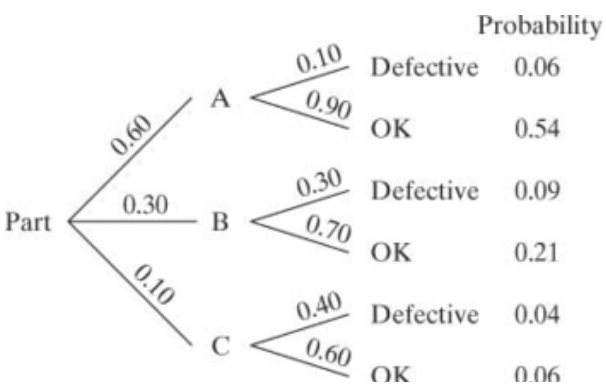
14. **(a)** Each of the 11 individuals was a block in a matched pairs design. Each participant took the caffeine tablets on one of the two-day sessions and the placebo on the other. The order in which they took the caffeine or the placebo was decided randomly. The tapping test was administered at the end of each two-day trial. The results to be compared are the differences between the caffeine and placebo scores on the tapping test. The blocking was done to control for individual differences in dexterity. **(b)** The order was randomized to control for any possible influence of the order in which the treatments were administered on the subject's tapping speed. **(c)** It is possible to carry out this experiment in a double-blind manner. This means that neither the subjects nor the people who come in contact with them during the experiment (including those who record the number of taps) have knowledge of the order in which the caffeine or placebo was administered.

Chapter 5

1. C 2. D 3. C 4. B 5. B 6. C 7. E 8. E 9. B 10. C

11. (a) Since each outcome is equally likely and there are 48 outcomes, each outcome has probability $\frac{1}{48}$. There are 27 ways in which the teacher wins (all boxes above and to the right of the diagonal line indicating ties. So the probability that the teacher wins is $\frac{27}{48}$. (b) We use the fact that $P(a \cup B) = P(a) + P(B) - P(a \cap B)$. From part (a), $P(A) = \frac{27}{48}$ There are 8 outcomes in which "you get a 3," so $P(B) = \frac{8}{48}$ and there are 5 outcomes in which "you roll a 3" and the teacher still wins, so $P(A \cap B) = \frac{5}{48}$. Putting all this together gives $P(A \cup B) = \frac{27}{48} + \frac{8}{48} - \frac{5}{48} = \frac{5}{8}$. (c) From part (b), we have $P(A) = \frac{27}{48}$, $P(B) = \frac{8}{48}$, and $P(A \cap B) = \frac{5}{48}$. Since $P(A \cap B) = \frac{5}{48} \neq \left(\frac{27}{48}\right)\left(\frac{8}{48}\right) = P(A)P(B)$, a and B are not independent.

12. (a) Here is a tree diagram.



(b) To get the probability that a part randomly chosen from all parts produced by this machine is defective, add the probabilities from all branches in the tree that end in a defective part. P(defective) = 0.06 + 0.09 + 0.04 = 0.19. (c) First compute the conditional probabilities that the part was produced on a particular machine given that it is defective. $P(a \mid \text{defective})$ = 0.3158. $P(B \mid \text{defective})$ = 0.4737. $P(\text{C} \mid \text{defective})$ = 0.2105. Since the largest of these three conditional probabilities is for Machine B, given that a part is defective, it is most likely to have come from Machine B.

13. (a) $P(\text{gets cancer} \mid \text{smoker}) = \dfrac{P(\text{gets cancer} \cap \text{smoker})}{P(\text{smoker})} = 0.32$.

(b) The event that an individual either smokes or gets cancer is the complement of the event that the individual neither smokes nor gets cancer. So P(smokes ∪ gets cancer)= 1 − P(does not smoke ∩ does not get cancer) = 0.29. (c) First find the probability that one person gets cancer. From part (b) we know that P(smokes ∪ gets cancer) = 0.29. But we also know that P(smokes) = 0.25 and P(smokes ∩ gets cancer) = 0.08.Now, using $P(S \cup C) = P(S) + P(C) - P(S \cap C)$ we get 0.29 = 0.25 + P(gets cancer) − 0.08. Solving for P(gets cancer) we get 0.12. Finally, P(at least one gets cancer) = 1 −P(neither gets cancer) = 0.2256.

14. **(a)** Assign the numbers 01–17 to represent cars with out-of-state plates and the remaining two-digit numbers between 00 and 99 to represent other cars. Start reading two-digit numbers from a random number table until you have found two numbers between 01 and 17. Record how many two-digit numbers you had to read in order to get two numbers between 01 and 17. Repeat many times for the simulation.**(b)** The first sample is 41 **05 09.** The numbers in bold represent cars with out-of-state plates. In this sample, it took 3 cars to find 2 with out-of-state plates. The second sample is 20 31 **06** 44 90 50 59 59 88 43 18 80 53 **11,** wherein it took 14 cars to find 2 with out-of-state plates. The third sample is 58 44 69 94 86 85 79 67 **05** 81 18 45**14,** wherein it took 13 cars to find 2 with out-of-state plates.