Lawton Chiles Statewide Statistics Team Solutions

1. When you type the data into a list, the mean is 93 and the median is 72. When you subtract the mean from each value, square the differences and add them up, you get 67682. Divide by 14 (n – 1), to get the variance of 33841/7. The standard deviation is the square root of that, or $\frac{\sqrt{236887}}{7}$. The interquartile range is 166 – 30 = 136.

2. For part A, $\frac{88-74}{8} = 1.75.$ This leads to a p value of .0400591135, which rounds to .0401. For part B, $\frac{58-74}{8} = -2.$ This leads to a p value of .022750062, which rounds to .0228. For part C, $\frac{62-74}{8} = -1.5\ and\ \frac{85-74}{8} = 1.375.$ This leads to p values of .0668072287 and .9154342212 when both are done with less than as the direction. The difference between them is .8486269925, which rounds to .8486. For part D, $\frac{94-74}{8} = 2.5.$ This leads to a p value of .9937903201. The probability that the student scored higher than 77 and is less than 94 is .3538302873 – (1 - .9937903201) = .3476206074. The final answer is $\frac{.3476206074}{.9937903201}$ = .3497927081, which rounds to .3498.

3. Since the coefficient of determination is .7225, the correlation coefficient is -.85 because you were told it's a negative relationship. Plugging the values in produces $y - 70 = (-.85)\left(\frac{12}{1.5}\right)(x-3)$. When you solve for y, the equation becomes y = -6.8x + 90.4. So part A = -6.8 and B = 90.4. For part C, plug 5 in for x and solve for y. y = (-6.8)(5) + 90.4 = 56.4. For part D, plug 1 in for x and solve for y. You get a predicted value of 83.6. The actual score was 85, so the residual is 85 – 83.6 = 1.4

4. For part A, since they are independent, $P(X \cup Y) = P(X) + P(Y) - P(X)P(Y) = .64 + .28 - .64(.28) = .7408$. For part B, since the part outside the sets is .41, $P(X \cup Y) = .59$. Therefore, .59 = .64 + .28 - $P(X \cap Y)$. Solving gives $P(X \cap Y) = .33$. For part C, since the intersection is .22, X only is .42 and Y only is .06. $P(X'|Y) = \frac{P(X' \cap Y)}{P(Y)}$ = $\frac{.06}{.28}$ = $\frac{3}{14}$. For part D, since the two events are mutually exclusive, their union equals their sum. .64 + .28 = .92.

5. For part A, 3(90) – 2(78) = 270 – 156 = 114. For parts B and C, the standard deviation is equal to the square root of the sum of the variances of the new sets because they are independent sets.
B = $\sqrt{((6)(6))^2 + (3(12))^2}$ = $\sqrt{2592}$ = 36$\sqrt{2}$. C = $\sqrt{(3(6))^2 + (4(12))^2}$ = $\sqrt{2628}$ = 6$\sqrt{73}$.
For part D, 10(78) – 12(90) = 780 – 1080 = -300.

6. This is a binomial distribution. For part A, the mean is 150(.46) = 69. For part B, $\sqrt{150(.46).54}$ = $\sqrt{\frac{1863}{50}}$ = $\frac{9\sqrt{23}}{5\sqrt{2}}$ = $\frac{9\sqrt{46}}{10}$. For part C, binompdf(150, .46, 75) = .0401827195, which rounds to .0402. For part D, 1 – binomcdf(150, .46, 75) = .1435169504, which rounds to .1435.

7. This is a geometric distribution. For part A, the mean is 1/(.25) = 4. For part B, $\sqrt{\frac{1-.25}{.25^2}}$ = $\sqrt{12}$ = 2$\sqrt{3}$. For part C, geometcdf(.25, 4) = .68359375, which rounds to .6836. For part D, $(1-.25)^3$ = .421875, which rounds to .4219.

8. The row totals are 100, 75, 25 and the columns totals are 83, 64, 53. There are 200 total in the chart. For part A, there are 100 Spanish students out of 200 total, so A = .5. For part B, $P(C \cup F) = \frac{64+75-20}{200} = \frac{119}{200}$. For part C, there are 53 students in Physics and 3 of them take Japanese, so C = $\frac{3}{53}$. For part D, there are 117 students not in Biology and 50 of them take French, so D = $\frac{50}{117}$.

9. For part A, the standard deviation goes from 18 to 8. Therefore, you must multiply the data by $\frac{8}{18} = \frac{4}{9}$. When you multiply the mean by $\frac{4}{9}$, you get $\frac{244}{9}$. You must add $\frac{431}{9}$ to get up to the new mean of 75. So A = $\frac{4}{9}$ and B = $\frac{431}{9}$. For part C, plug 88 into the equation to get y = $\frac{4}{9}(88) + \frac{431}{9}$ = 87. For part D, plug 91 into the equation to get 91 = $\frac{4}{9}x + \frac{431}{9}$. Solving gives x = 97.

10. For parts A and B, create two equations with the given information. For each equation, the z-score was rounded to 2 decimal places using invNorm. They are 90 – mean = 1.64SD and 65 – mean = -1.04SD. Solving these equations by eliminating the mean produces an exact standard deviation of 625/67. When you plug this into either equation, you get an exact mean of 5005/67. For part C, plug into the z-score formula. z = $\frac{80-\frac{5005}{67}}{\frac{625}{67}}$ = .568. This leads to a p value of .2850174436, which rounds to .2850. For part D, plug into the z-score formula. z = $\frac{70-\frac{5005}{67}}{\frac{625}{67}}$ = -.504. This leads to a p value of .3071306798, which rounds to .3071.

11. When you plug the information into a Venn diagram, there are three open spots for each group that takes exactly two of the classes. Let A = History and Math, B = Math and English and C = English and History. Create three equations with the information. A + B = 155, B + C = 135 and A + C = 110. Solving these equations produces A = 65, B = 90 and C = 45. For part A of the question, when you add all the numbers inside the circles of the Venn diagram, you get 620. Therefore, there are 30 students who don't take any of the three classes. For part B, there are 65 + 90 + 45 + 140 = 340 students who take at least two of the classes. For part C, there are 425 students who take English. 230 of them (90 + 140) take Math from that group. So part C = $\frac{230}{425} = \frac{46}{85}$. For part D, there are 225 students who don't take English (650 – 425). 115 of them (65 + 50) take History from that group. So part D = $\frac{115}{225} = \frac{23}{45}$.

12. To find the mean of a discrete distribution multiply the value by its probability and add the products up. For part A, 1(.05) + 2(.18) + 3(.43) + 4(.22) + 5(.12) = 3.18. To find the exact standard deviation, subtract the mean from each value, square the differences, multiply the square difference by its probability, add them up and take the square root. The standard deviation equals $\sqrt{\frac{2619}{2500}} = \frac{3\sqrt{291}}{50}$. For part C, the mean is 1(.01) + 2(.02) + 3(.1) + 4(.13) + 5(.53) + 6(.16) + 7(.05) = 4.83. For part D, when you go through the process, the standard deviation is $\sqrt{\frac{12011}{10000}} = \frac{\sqrt{12011}}{100}$.

13. For part A, the solution is (.48)(.37) = .1776. For part B, there are two ways for a student to be in extracurricular activities. The solution is (.52)(.45) + (.48)(.63) = .5364. For part C, we know the student is in extracurricular activities. The probability that it is a boy is

$\frac{.52(.45)}{(.52)(.45)+(.48)(.63)} = \frac{65}{149}$. For part D, we know the student is not in extracurricular activities. The probability that it is a girl is $\frac{.48(.37)}{.52(.55)+.48(.37)} = \frac{444}{1159}$.

14. Part A is false. For the data set 1, 2, 3, 4 the median is 2.5. That median is not in the data set. Part B is true. The only way to show causation is through a controlled experiment to eliminate lurking variables. Part C is false because the fraction should be $\frac{s_y}{s_x}$. Part D is false because when two events are mutually exclusive, they have nothing in common, so their intersection is zero. In order for two events to be independent, they must have an intersection.