Answers to the Study Guide for the Spring 2013 Final

1. (a) i. average = 4; deviations = $-3,-1,0,1,3$; SD = 2.
   ii. average = 9; deviations = $-3,-1,0,1,3$; SD = 2.

   (b) List (ii) is obtained from list (i) by adding 5 to each entry. This adds 5 to the average, but does not affect the deviations from the average. So, it does not affect the SD. Adding the same number to each entry on a list does not affect the SD.

2. (a) i. average = 4; deviations = $-3,-1,0,1,3$; SD = 2.
   ii. average = 12; deviations = $-9,-3,0,3,9$; SD = 6.

   (b) List (ii) is obtained from list (i) by multiplying each entry by 3. This multiplies the average by 3. It also multiplies the deviations from the average by a factor of 3, so, it multiplies the SD by a factor of 3. Multiplying each entry on a list by the same positive number just multiplies the SD by that number.

3. (a) i. average = 2; deviations = $3,-6,1,-3,5$; SD = 4.
   ii. average = -2; deviations = $-3,6,-1,3,-5$; SD = 4.

   (b) List (ii) is obtained from list (i) by changing the sign of each entry. This changes the sign of the average and all the deviations from the average, but does not affect the SD because the average is obtained from the squares of the deviations..

4. (a) 93% (b) 72% (C) $12\frac{1}{2}$% (d) 50% (e) $33\frac{1}{3}$%

5. (a) 99.38% (b) 57.63% (c) 93.945% (d) 11% (e) about $13\frac{1}{4}$% (f) 87.5%

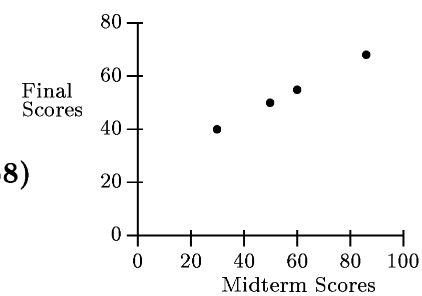
6. (a) 40 (b) 68 (c) 55 (d) 50

   Work for (a). A score of 30 is 2 SDs below average. However, r is only 0.5. If you take the students who are 2 SDs below average on the midterm, their average score on the final will only be about 0.5×2 SDs below average, that is, $1 \times 10 = 10$ points. So, the estimated average score on the final for this group is $50 - 10 = 40$.

   Labels of points plotted
   (listed from left to right)

   (30, 40), (50, 50), (60, 55), (86, 68)

   (50, 50) is the point of averages



   Comment. The regression estimates always lie on a line—the regression line. It is possible to get the answers by finding the equation for that regression line and then plugging in the values for x. However, there's more potential for introducing error than the regression method used above.

7. (a) 128. Work: 135 is $2\frac{1}{3}$ SDs above average at age 18, so the estimate at age 35 is above average by $r \times 2\frac{1}{3}$ SDs. As a fraction $r = 4/5$ and $2\frac{1}{3} = 7/3$. This becomes $\frac{28}{15} \times 15 = 28$ points.

   (b) 88.

   (c) 100, the average.

   (d) Regression is toward the mean and the mean is 100 points. In (a) and (b) the score at age 35 is indeed closer, while in part (c) the score is already at the mean, so it stays there.

8. (a) 73.7 inches (b) 73.7 inches (the same calculation)

9. In a run of one SD of x, the regression line rises

$$r \times \text{SD of } y.$$

The slope is

$$\text{rise/run} = 0.40 \times 0.6/80 \;=\; 0.003 \quad \text{GPA points per SAT point.}$$

The intercept is

$$2.6 \;-\; 0.003 \times 550 \;=\; 0.95.$$

[Or solve the equation

$$y = mx + b$$

for b, using the average values of 550 and 2.6 for x and y.

We get

$$2.6 = 0.003(550) + b,$$

so

$$b = 2.6 - 0.003(550) = 0.95,$$

as before.]

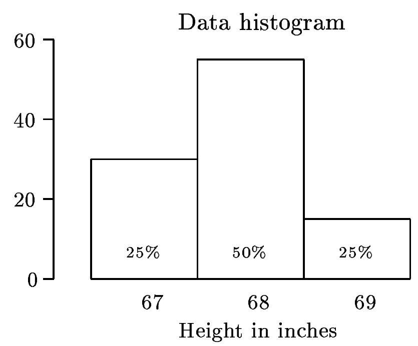
So the equation is

$$\text{predicted first-year GPA} \;=\; 0.003 \times \text{Math SAT score} \;+\; 0.95\,.$$

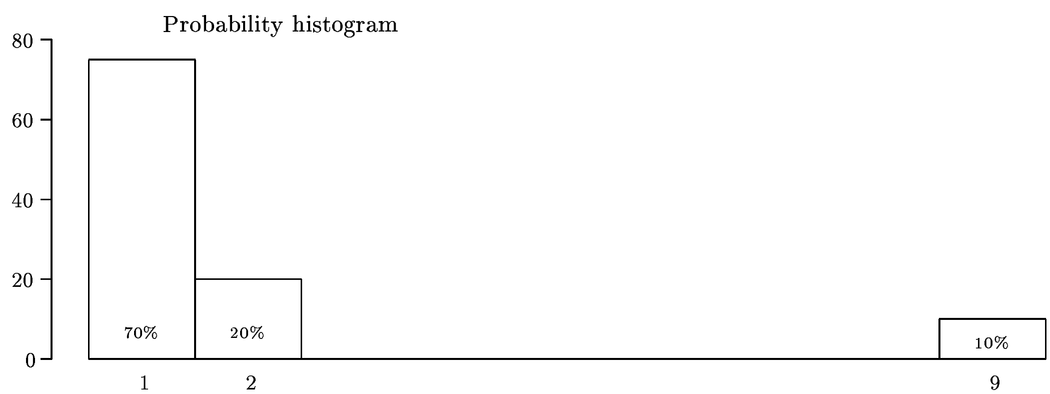
10. weight $=$ 6 $\times$ height $-$ 240 pounds.

11. height of husband $=$ 0.27 $\times$ height of wife $+$ 50.99 inches.

12.



13.



14. The histogram for the data will get closer to the probability histogram for the number of heads in 100 tosses of a coin (section 18.2). This probability histogram is not the normal curve—although it's close (chapter 18, figure 3). Also see pages 326 to 327.

15. (a) True (section 18.5)

(b) False. The histogram for the numbers in the box does not change, and could have any shape—depending on how the box was set up in the first place.

(c) False. The histogram for the draws gets more and more like the histogram for the numbers in the box. You have to distinguish between (i) the probability histogram for the sum of the draws, and (ii) the histogram for the draws as data.

(d) True.

16. $(5/6)^{30} \approx 0.42\%$.

As a decimal: about 0.004212720233.

As an exact fraction: impossible on a calculator—too many digits to display all of them.

As a fraction is the form $1/n$: about 1/237, so one chance out of 237.

17. (a) The question is about the second ticket, not the first; see part (a) of example 2 in section 2 of chapter 13, pages 226 and 227.

    The answer is 1/4.

    (b) False. While it is true that the multiplication rule applies, the two draws are not independent. The correct application of the multiplication rule in this case requires a conditional probability for the second draw, which assumes that the first draw resulted in the 3.

    The correct chance is $1/4 \times 1/3$.

18. (a) dependent and not mutually exclusive.

    (b) multiply.

    required.

    (c) $1/5 \times 1/4 = 1/20 = 0.05 = 5\%$.

19. (a) about 1.7% (b) about 88% (c) about 8% (d) about 4% (could subtract (b) plus (c) from 100%)

20. (a) True. It is a properly classified 0–1 counting box, and the percentage requires dividing the sum of the draws by the number of draws and then multiplying by 100%.

    Here it gives: sum of the draws $\times$ 1/500 $\times$ 100%. Simplify because $1/500 \times 100\% = 0.2\%$.

    (b) True. The expected value for the sample percentage equals the population percentage.

    (c) True. It is a large sample, so the central limit theorem applies and the normal curve may be used as an approximation, once the percents of 1's among the draws have been converted to standard units.

21. (a) sample, population. (b) False. They are exactly equal.

22. The first thing to do it to set up a box model. There should be 30,000 tickets in the box, one for each registered voter; 12,000 are marked 1 (Democrat) and 18,000 are marked 0. The number of Democrats in the sample is like the sum of 1,000 draws from the box. The fraction of 1's in the box is 0.4. The expected value for the sum is $1{,}000 \times 0.4 = 400$. The SD of the box is $\sqrt{0.4 \times 0.6} \approx 0.49$. The SE for the sum is $\sqrt{1000} \times 0.49 \approx 15.5$.

    (a) The expected value for the percent is 400 out of 1,000, or 40%. (Or simply observe that the expected value for the percentage of Democrats in the sample is equal to the percentage of Democrats in the town: $12{,}000/30{,}000 \times 100\% = 40\%$.)

    The SE for the percent is 15.5 out of 1,000 or 1.55%.

    The SE for the percent may also be calculated directly from the SD of the box and the sample size by

    $$\frac{\text{SD of the box}}{\sqrt{n}} \times 100\% = \frac{0.49}{\sqrt{1{,}000}} \times 100\% = 1.55\%.$$

    (No surprise about the expected value: 40% of the registered voters are Democrats.)

    (b) The percentage of Democrats in the sample will be about 40.0%, give or take 1.55% or so. Parts (a) and (b) require the same calculations; in (b) you have to interpret the results.

23. 71.4%, 68%.

Comment. The expected value is computed from the box; the observed value, from the sample.

24. SE for average ≈ 1.25, so $z \approx (52.7 - 50)/1.25 \approx 2.16$ and P is approximately the area to the right of 2.16 under the normal curve. From the table, this is about 1.6%. The difference is hard to explain as chance variation. The alternative hypothesis is looking good. (This assumes that the line is drawn at 5%.)

25. $\chi^2 = 1.0$, $d = 5$, $95\% < P < 99\%$. The die looks fair. χ^2 was $\frac{1^2+1^2+0^2+2^2+2^2+0^2}{10}$.

26. The sample size is large, so the sum of the draws will follow the normal curve by the Central Limit Theorem.

The sample percent is just the sum of the draws divided by the size of the sample, then multiplied by 100%. It will follow the normal curve.

The standard error is rather small, because the sample size is large. Using the standard deviation of the sample (the bootstrap method) to calculate the SE is allowed in view of the large sample size. The square root law applies because the sample was random and has a box model with numbered tickets. The factor correcting for the fact that the sample was not taken with replacement should be close to 1, in view of the large size of the population relative to the size of the sample.

Finally, one can be fairly confident that the sample percentage is close to the population percentage. In essence, they should generally end up fairly close together, so the sample is a good estimate of the population. A result within, say 3 SEs, should happen very often (99.73%) on the normal curve.

That is the basis of our confidence.