

Solutions to Review Problems for the Math 125 Final

1. The height of the block is 2% per thousand dollars. Each thousand-dollar interval between \$15,000 and \$25,000 contains about 2% of the families in the city. There are 7 of these thousand-dollar intervals between \$18,000 and \$25,000. The answer is $7 \times 2\% = 14\%$. About 14% of the families in the city had incomes between \$18,000 and \$25,000.

The problem shows that with the density scale, the areas of the blocks come out in percent. The horizontal units—thousands of dollars—cancel. The area of the block shown is therefore:

$$2\% \text{ per thousand dollars} \times 10 \text{ thousand dollars} = 20\%.$$

And then, using a proportion, about $7/10$ of the area of that block will be the answer since the base of a block for the incomes between \$18,000 and \$25,000 will be $7/10$ of the base of the block pictured and we may assume that the density for the smaller block is approximately the same as the density for the block pictured. (There's no reason to assume otherwise.)

Answer: about 14% of the families in the city had incomes between \$18,000 and \$25,000.

2. The height of the block must be 15% per \$100. The expression for it is $\frac{100\% - (10\% + 20\% + 5(5\%))}{\$300}$.

The median wage is about \$333 per month, and the 75th percentile wage is about \$500 per month.

The total area of the missing block is 45%. Adding $4/9$ of 45% to $10\% + 20\%$ will give 50%. That means $\$200 + \frac{4}{9}(\$300)$ is the median wage.

3. (a) False. The area of the rectangle for 4 or 5 children will be 7%. The height will be 7% divided by the base: 3.5% per child.
 (b) False. The area of the rectangle represents the percentage of women. So the area is 12%. Furthermore, the rectangle is 1 child wide, running from 2.5 to 3.5. There is no 3 at all in the dimensions of the rectangle.
 (c) False. As in part (a) the area of the rectangle represents the percentage of women. So the area is 7%.
 (d) False. The height of the rectangle is found by dividing its area (the percent of women) by the width of the base.

For 4 or 5 children: the height is $\frac{7\%}{2 \text{ children}} = 3.5 \text{ percent per child}$.

For 6 children: the height is $\frac{3.5\%}{1 \text{ child}} = 3.5 \text{ percent per child}$.

The two blocks have equal height.

- (e) True.

4. (a) 34
 (b) There is no mode (most common value on the list).
 (c) The average is 33; the deviations are 1, 4, -3, -6, -4, 3, 5;
 the SD is

$$\sqrt{\frac{1^2 + 4^2 + (-3)^2 + (-6)^2 + (-4)^2 + 3^2 + 5^2}{7}} = \sqrt{\frac{1 + 16 + 9 + 36 + 16 + 9 + 25}{7}} = \sqrt{\frac{112}{7}} = \sqrt{16} = 4.$$

- (d) $\frac{1}{2}$ SD is $\frac{1}{2} \times 4 = 2$. Within 2 of 33 is the range 31 to 35, containing one of the ages.
 (e) 1.75 SDs equals $1.75 \times 4 = 7$ years. Within 7 of 33 is the range 26 to 40. This range contains all 7 of the ages.
 (f) For the observed value of 30,

$$\text{standard units} = \frac{30 - 33}{4} = \frac{-3}{4} = -0.75.$$

5. (a) The average is 10; the deviations are 3, -1, 1, -3, 0;
 the SD is

$$\sqrt{\frac{3^2 + (-1)^2 + 1^2 + (-3)^2 + 0^2}{5}} = \sqrt{\frac{9 + 1 + 1 + 9 + 0}{5}} = \sqrt{\frac{20}{5}} = \sqrt{4} = 2.$$

- i. Within 2 of 10 is 8 to 12. There are 3 entries on the list in that range: 9, 10, and 11.
 ii. Within 2 SDs is within $2 \times 2 = 4$; that's 6 to 14. All of the entries of the list are in that range.
 (b) For the observed value of 13,

$$\text{standard units} = \frac{13 - 10}{2} = \frac{3}{2} = 1.5.$$

In standard units the list is +1.5, -0.5, +0.5, -1.5, 0.

- i. The converted list has an average of 0 and an SD of 1. (This is always so; when converted to standard units, any list will average out to 0 and the SD will be 1.)

6. (a) The standard units for 72.4 is $\frac{72.4 - 67}{4} = 1.35$.

The area under the normal curve from -1.35 to +1.35 is about 82%. Each tail is about $\frac{100 - 82}{2} = 9\%$.

Add the lower tail to the middle area to get 91%, the percentile rank.

Or, equivalently, add half of 82% to 50%, getting $41\% + 50\% = 91\%$.

Answer: 91%.

- (b) If the percentile rank is 40%, the left tail is 40%. The corresponding right tail will also be 40%. That leaves $100\% - 40\% - 40\% = 20\%$ in the middle. Look up that area to find the standard units of $z = \pm 0.25$, for that picture.

In this case, the 40th percentile is below average. Use -0.25.

So $-0.25 = \frac{x - 67}{4}$, $-1 = x - 67$, and $x = 67 - 1 = 66$ inches.

- (c) The standard units for 62 and 66 inches are $\frac{62 - 67}{4} = -1.25$ and $\frac{66 - 67}{4} = -0.25$.

Then take the area between -1.25 and -0.25 under the normal curve. This will be $\frac{1}{2}(78.87\%) - \frac{1}{2}(19.74\%) = 29.565\%$.

7. (a) The standard units for 505 is $\frac{505-500}{100} = 0.05$.

The area under the normal curve from -0.05 to $+0.05$ is about 4%. Each tail is about $\frac{100-4}{2} = 48\%$.

Add the lower tail to the middle area to get 52%, the percentile rank.

Or, equivalently, add half of 4% to 50%, getting $2\% + 50\% = 52\%$.

Answer: was at the 52nd percentile of the score distribution.

- (b) At the 54th percentile, the right tail is 46%. The corresponding left tail will also be 46%. That leaves $100\% - 46\% - 46\% = 8\%$ in the middle. Look up that area to find the standard units of $z = \pm 0.10$.

Of course, the 54th percentile is above average. Use $+0.10$.

So $+0.10 = \frac{x-500}{100}$, $10 = x - 500$, and $x = 500 + 10 = 510$ points.

Note, as a check, that the score of 510 had a higher percentile rank than a score of 505, but the ranks were close.

8. (a) 95%.

$z = \frac{666-500}{100} = 1.66$. Use 1.65. Middle area is 90%. Then add one tail to get 95%, the area below.

- (b) 24%

The first step is to recognize that 0.7 SDs below average is -0.7 in standard units.

Use 0.70: middle area is about 52%. Lower tail is $\frac{100\%-52\%}{2} = 24\%$, and that's the area below -0.7 .

9. (a) The averages are 3 and 4, respectively; the SDs are 1 and 2, respectively. The average product of the values in standard units is then

$$\begin{aligned} & \frac{(0 \times 0) + (0 \times -1) + (-2 \times -1.5) + (0 \times 0.5) + (1 \times 0.5) + (1 \times 1.5)}{6} \\ &= \frac{0 + 0 + 3 + 0 + 0.5 + 1.5}{6} = \frac{5}{6} \approx 0.8333. \end{aligned}$$

$$r = 0.8333.$$

- (b) When $x = 1.5$, it is $\frac{1.5-3}{1} = -1.5$ in standard units. That means it is 1.5 SDs below average. Since r is positive, y will also be below average: by $r \times 1.5 = \frac{5}{6} \times 1.5 = \frac{5}{6} \times \frac{3}{2} = \frac{5}{4} = 1.25$ SDs of y . That is $1.25 \times 2 = 2.5$ below the average y . The predicted value of y is $4 - 2.5 = 1.5$. Answer: y is predicted to be 1.5.

- (c) r.m.s. error $= \sqrt{1 - r^2} \times \text{the SD of } y = \sqrt{1 - (0.8333)^2} \times 2 = \sqrt{1 - .6944} \times 2 = \sqrt{.3056} \times 2 = 0.5528 \times 2 \approx 1.1054$.

- (d) The slope of the regression line is

$$r \times \frac{\text{SD of } y}{\text{SD of } x} = \frac{5}{6} \times \frac{2}{1} = \frac{5}{3} \approx 1.6667.$$

10. (a) The averages are 3 and 4, respectively; the SDs are 1 and 2, respectively. The average product of the values in standard units is then

$$\frac{(1.0 \times 0) + (0 \times -1) + (1.0 \times -1.5) + (0 \times 0.5) + (-2.0 \times 0.5) + (0 \times 1.5)}{6}$$

$$= \frac{0 + 0 + (-1.5) + 0 + (-1) + 0}{6} = \frac{-2.5}{6} \approx -0.41667.$$

$$r = -0.41667.$$

The slope of the regression line is

$$\frac{r \times \text{SD of } y}{\text{SD of } x} = \frac{-0.41667 \times 2}{1} = -0.8333 \quad \text{or} \quad 2 \left(\frac{-2.5}{6} \right) = -\frac{5}{6}.$$

The equation of the regression line is

$$y = mx + b = -0.8333x + b.$$

To find b , substitute the point of averages, $(3, 4)$, into the equation, getting

$$4 = -0.8333(3) + b \text{ and } 4 = -2.5 + b; \text{ so } b = 6.5.$$

Answer: $y = -0.8333x + 6.5$.

$$(b) \text{ RMS error of regression line: } \sqrt{1 - r^2} \times (\text{SD of } y) = \sqrt{1 - (-0.41667)^2} \times 2 =$$

$$\sqrt{1 - 0.17361} \times 2 = \sqrt{.82639} \times 2 = .909 \times 2 = 1.818.$$

- (c) When $x = 1.5$, it's 1.5 SDs below 3, the average x . Negative correlation gives us y **above** average by r times -1.5 SDs. That's $(-1.5) \times (-.41667) = .625$ SDs. Each SD of y is 2, so y will be $2 \times 0.625 = 1.25$ above the average y . Answer: predict $y = 5.25$.

You also could just plug $x = 1.5$ into the equation of the regression line to get

$$y = -0.8333(1.5) + 6.5 = -1.25 + 6.5 = 5.25.$$

- (d) For $x = 1$, the point on the regression line is found by
- $$y = -0.8333(1) + 6.5 = 5.6667. \text{ So the point } (1, 5) \text{ is below the line.}$$

11. The averages are 7 and 19, respectively; the SDs are 2 and 4, respectively. The average product of the values in standard units is then

$$\frac{(1.5 \times -1) + (0.5 \times 1.5) + (0 \times 0) + (-1 \times -1.5) + (-0.5 \times -0.5) + (1 \times 1) + (-1.5 \times 0.5)}{7}$$

$$= \frac{(-1.5) + 0.75 + 0 + 1.5 + 0.25 + 1 + (-0.75)}{7} = \frac{1.25}{7} \approx 0.17857.$$

$$r = 0.17857.$$

When $x = 8.75$, it is $\frac{8.75-7}{2} = 0.875$ in standard units. Multiply that by r to get 0.15625. y will be predicted to be 0.15625 SDs above its average. That's 0.625 added to 19, or 19.625.

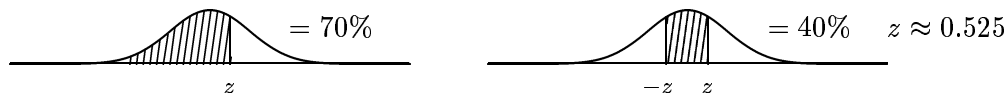
The RMS error of the regression line is calculated as

$$\sqrt{1 - r^2} \times \text{SD of } y = \sqrt{1 - .17857^2} \times 4 = \sqrt{1 - 0.031887} \times 4 = \sqrt{0.968113} \times 4 = 0.9839 \times 4 = 3.936.$$

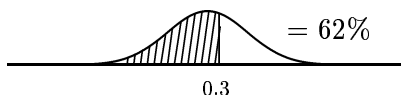
12. (a) False, the other way around. (b) True. (c) True, it's the same both ways. (d) True. (e) False, association is not causation.

13. (a) 62% (b) 50%; the point of averages is always on the regression line. (c) 50%; predict average GPA.

Work for (a):



In standard units, his SAT score was 0.525. The regression prediction for his first-year score is $0.6 \times 0.525 \approx 0.3$ in standard units.



This corresponds to a percentile rank of 62%.

Details of converting percentile to z : Take the given 70% as the area below z . That means that the upper tail must be 30% and the so-far-unpictured lower tail must be 30%. Take the lower tail from the given percentile rank and you will have the percentage value that must be looked up on the table: 40%. It's about halfway between $z = 0.50$ and $z = 0.55$, so use $z = 0.525$.

It is also possible to do the first step with a slightly different logic. Take the 70%, the given percentile rank and break it into 2 parts: the area below 0 (50%), and the area between 0 and z (20%). Then double the 20% to get the middle area (from $-z$ to z) for looking up on the table. Again look up 40% on the normal table and find z .

Details of converting z to a percentile, as required in the last step: $z = 0.3$ has 24% in the middle. Below it will be the 50% below 0 and the 12% from 0 to z . (This is half of the middle area for $z = 0.3$.) The sum is 62%.

Alternative logic: 24% in the middle, means $100\% - 24\% = 76\%$ for the two tails. One will be 38%. We need all the area except one tail = $100\% - 38\% = 62\%$.

14. (a) 78. A score of 90 is 2 SDs above average. However, r is only 0.6. If you take the students who are 2 SDs above average on the midterm, their average score on the final will only be about $r \times 2$ SDs = 0.6×2 SDs = 1.2 SDs above average on the final, that is, 1.2×15 points = 18 points. So, the estimated average score on the final for this group is $60 + 18 = 78$ pts. The regression estimates always lie on the regression line. This could be done by finding the equation of the regression line and then plugging 90 into the equation found.

The slope of the regression line is $r \times \frac{\text{SD of } y}{\text{SD of } x} = 0.6 \times 1 = 0.6$. Then $y = mx + b$ and b is found by plugging in the point of averages, (60, 60). This gives $60 = 0.6(60) + b$ and $b = 24$. Plug 90 into $y = 0.6x + 24$ and it yields $0.6(90) + 24 = 54 + 24 = 78$.

(For this problem the method presented originally requires much less work.)

- (b) 78. Part (b) is about individuals; part (a) was about groups. The arithmetic is the same for both parts.
- (c) False. This is the regression fallacy. There are two regression lines, depending on which is given: the midterm score or the final score.
See pictures in the Freeman text on pages 175 and A-63.

15. (a) In a run of 1 SD, the regression line rises $r \times \text{SD}$. The slope is $0.57 \times 4/30 = 0.076$ inches per pound. The intercept is $67 - 0.076 \times 152 = 55.448$. So the equation is
predicted height = $0.76 \times \text{weight} + 55.448$.

The prediction when the weight (x) is 200 pounds is $0.076 \times 200 + 55.448 = 70.648$ inches.

- (b) The slope of the regression equation is 0.076 inches per pound. If one man is 30 pounds heavier than another, he is expected to be 30×0.076 inches, or 2.28 inches, taller.
- (c) $\sqrt{1 - 0.57^2} \times 4 = 3.2866$ inches.

(Part (a) calculations:) The slope of the regression line is $r \times \frac{\text{SD of } y}{\text{SD of } x} = 0.57 \times 4/30 = 0.076$.

Then $y = mx + b$ and b is found by plugging in the point of averages: (152, 67).

This gives $67 = 0.076(152) + b$ and $b = 55.448$.

16. (a) with, independent.
 (b) with, mutually exclusive.
 (c) without, independent.
17. (a) Yes. (b) Yes. (c) Yes.
18. • $P(A) = 1/10 = 10\%$ and $P(B) = 1/10 = 10\%$.
 They are independent but not mutually exclusive.
- $P(\text{both}) = P(A) \times P(B) = (1/10) \times (1/10) = 1/100 = 1\%$.
 Since they are independent the chance of both is found by simply multiplying the unconditional probabilities.
- $P(A \text{ or } B) = 1 - P(\text{the opposite thing})$.
 Adding the chances of A and B is not correct because they are not mutually exclusive. The draws are, however, independent. So, no conditional probabilities are needed.
 The actual calculation:
 The opposite of “at least one of A or B” is “neither A nor B.”
 The chance of neither A nor B is found by the multiplication rule as $(9/10) \times (9/10) = 81/100 = 81\%$. Then get $P(A \text{ or } B)$ by subtracting 81% from 100%. The answer is 19%.
19. (a) $\frac{12}{52} \times \frac{11}{51} \times \frac{10}{50} \times \frac{9}{49} \times \frac{8}{48} = \frac{95,040}{311,875,200} = \frac{33}{108,290} = 0.0003 = .03 \text{ of } 1\%$.
- (b) $\frac{40}{52} \times \frac{39}{51} \times \frac{38}{50} \times \frac{37}{49} \times \frac{36}{48} = \frac{78,960,960}{311,875,200} = \frac{2109}{8330} = 0.253 = 25.3\%$.
- (c) It's the opposite of the answer to (b), so it is
 $1 - \frac{2109}{8330} = \frac{6221}{8330} = 1 - 0.253 = 0.747 = 100\% - 25.3\% = 74.7\%$.
- (d) It's the opposite of the answer to (a), so it is $1 - \frac{33}{108,290} = \frac{108,257}{108,290} = 1 - .0003 = 0.9997 = 100\% - 0.03\% = 99.97\%$.
- (e) It's the opposite of no aces. The chance of getting no aces is $\frac{48}{52} \times \frac{47}{51} \times \frac{46}{50} \times \frac{45}{49} \times \frac{44}{48} = \frac{205,476,480}{311,875,200} = \frac{35673}{54145} = 0.6588 = 65.88\%$.
 The chances of getting at least one ace are: $1 - \frac{35673}{54145} = \frac{18472}{54145} = 1 - 0.6588 = 0.3412 = 100\% - 65.88\% = 34.12\%$.
20. $(5/6)^{10} \approx 0.16$, or 16%.
 The chance of not getting a six in one roll is 5/6. It has to be all non-sixes, and they are independent, so use the multiplication rule ten times, without the need of conditional probabilities. (or use the binomial formula.)
21. The events are not mutually exclusive, so you cannot add 1% up 150 times (and it wouldn't work anyway because it's more than 100%).
 The opposite is that all of the people did not give a dollar. That chance for each person is $1 - 0.01 = 0.99$. (First convert the given 1% to 0.01, because it is not possible to multiply percents. Then subtract that from 1.) Then, the trials being independent, to get the chance of “all” just multiply unconditional probabilities. Repeat the rule 150 times to get:
 $.99^{150} = 0.22145 \approx .22 = 22\%$.
 The original question is that “at least one gave” which is the opposite of “none gave.”
 Subtract 22% from 100% to get 78%, the answer.

22. True.

23. The chance is

$$\frac{7!}{2!5!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^5 \approx 23.44\%.$$

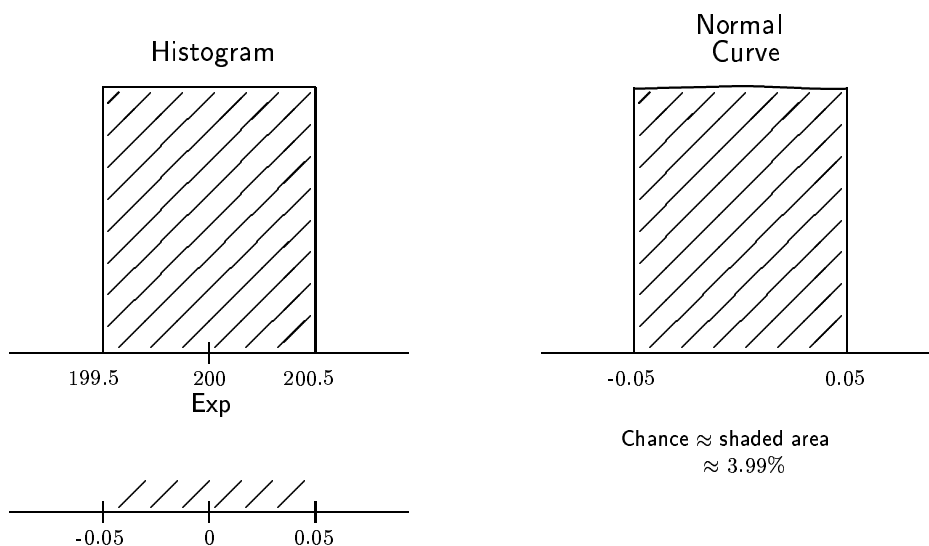
24. Ignore the 22 tails; if it comes out 15 heads, there must be 22 tails. The chance is

$$\frac{37!}{15!22!} \left(\frac{1}{2}\right)^{15} \left(\frac{1}{2}\right)^{22} \approx 6.8\%.$$

25. The expected number of heads is 200; the SE is 10. Since the SD of a zero-one box with one 0 and one 1 is 0.50, and the SE is the product of the square root of the number of draws and the SE, we get $\sqrt{400} \times 0.50 = 20 \times 0.50 = 10$. With a large number of tosses, the normal approximation is valid.

We seek the area of the rectangle over 200, and will need to find the standard units for its endpoints, 199.5 and 200.5. That gives us $\frac{199.5 - 200}{10} = -0.05$ and $\frac{200.5 - 200}{10} = 0.05$.

(The formula: std. units for the sum of draws = $\frac{\text{endpoint} - \text{EV of the sum}}{\text{SE for the sum}}$.)



Comment: the exact chance is $\frac{400!}{200! \times 200!} (.5)^{200} (.5)^{200} \approx 3.99\%$.

When doing the approximation, the curve doesn't curve much. The area under the normal curve is nearly rectangular, so there's not much error.

26. (a) To set up this problem as a sum of draws, it is necessary to classify the results as
- the ticket with the $\boxed{1}$
 - and
 - any other ticket.

The count will go up by 1 only when the result is the ticket with the $\boxed{1}$.

Make a new box with one $\boxed{1}$ and 4 $\boxed{0}$'s.

Its average is 0.2, and the SD = $(1 - 0)\sqrt{0.2(0.8)} = 0.4$.

For 160 draws, the EV = $160(0.2) = 32$ and the SE = $\sqrt{160}(0.4) = 5.059644$.

Since the number of draws is large and the EV is $32 > 5$ and it is for a sample percentage, the normal curve may be used to approximate the chances (The Central Limit Theorem). This will be an approximation to a probability histogram and, because it is a counting number, it is necessary first to adjust the endpoints of the block: $27 \pm .5 = 26.5$ and 27.5 .

Next, convert these endpoints to standard units:

$$z_1 = \frac{26.5 - 32}{5.059644} = -1.087 \approx -1.10. \text{ and } z_2 = \frac{27.5 - 32}{5.059644} = -0.889 \approx -0.90.$$

Final step to get approximating area: look up 1.10, take half the listed area; look up 0.90, take half the listed area; then subtract.

$$\frac{72.87\%}{2} - \frac{63.19\%}{2} = 4.84\% \approx 5\%$$

- (b) Here the **original** box with the $\boxed{2}$ must be used.

Average of the box = $\frac{0+0+0+1+2}{5} = 0.6$.

$$\begin{aligned} \text{(no short-cut for SD) SD of the box} &= \sqrt{(-.6)^2 + (-.6)^2 + (-.6)^2 + .4^2 + 1.4^2} \\ &= \sqrt{.36 + .36 + .36 + .16 + 1.96} = \sqrt{3.3} = 1.78885. \end{aligned}$$

Answers: EV of the sum = $160(.6) = 96$; observed sum = 111;

chance error = observed sum - EV sum = $111 - 96 = 15$;

SE for the sum = $\sqrt{160}(1.78885) = 22.627$;

$$\text{in standard units the observed sum} = \frac{111 - 96}{22.627} = 0.6629.$$

- i. The observed sum is subject to chance error. The EV sum is known exactly.

27. Before you can calculate the expected value and standard error, you should set up the required box. A box for a count or a percentage is always a 0–1 box. Here put one 1 and five 0's in the box, since we need to classify and count the results.

The average of the box is $1/6$, and the SD is

$$\sqrt{1/6 \times 5/6} = 0.372678.$$

- (a) The expected value for the sum of the draws equals the number of draws times the average of the box $= 720(1/6) = 120$.

The standard error of the sum of the draws equals the square root of the number of draws times the SD of the box $= \sqrt{720}(0.372678) = 10$.

- (b) The expected value for the sample percentage is just equal to the percentage of 1's in the box we have set up: $16\frac{2}{3}\%$.

Or find it by multiplying the average of the box by $100\% = (1/6) \times 100\% = 16\frac{2}{3}\%$.

The standard error (give or take) for the sample percentage is just the SD of the counting box times 100% and divided by the square root of the number of draws. This gives

$$\frac{0.372678 \times 100\%}{\sqrt{720}} = 1.3889\%.$$

- (c) The normal approximation applies and the median is at the expected value.

So: “bigger than $16\frac{2}{3}\%$.”

- (d) The standard units for 18% is just $\frac{18\% - 16\frac{2}{3}\%}{1.3889\%} = 0.96 \approx 0.95$.

The area under the normal curve above 0.95 is $\frac{100\% - 65.79\%}{2} = 17.105\% \approx 17\%$.

28. The expected value for the percentage of 1's among the draws and the percentage of 1's in the box are known exactly. In fact, they are always equal.

The observed percentage of 1's among the draws is subject to chance error, and can only be approximately predicted before the draws are made.

29. (a) False. All you need for any confidence interval is a single random sample with a large number of draws. Such a sample will be representative of the population, subject to chance error. Because it is only one sample, it is not correct to make probability statements based on it—but confidence intervals are fine.

- (b) The sample percentage is $(543/1,000) \times 100\% = 54.3\%$. The SE for the sample percentage of Democrats is figured as $\frac{\sqrt{0.543 \times 0.457}}{\sqrt{1000}} \times 100\% = 1.6\%$.

The formula for a 95%-confidence interval is:

sample percentage plus or minus twice the standard error for the percentage.

So $54.3\% \pm 3.2\%$ is a 95%-confidence interval for the percentage of Democrats in the city.

30. (a) True: the SD of the box is estimated from the sample data and the SE is obtained by dividing it by the square root of the number of draws.

- (b) False. As long as the sample was randomly chosen with replacement from a box and the sample size is reasonably large, confidence interval methods are appropriate.

A confidence interval is based on the results of a single sample of a particular size.

- (c) True. The sample average plus or minus 2 SEs is the definition of a 95%-confidence interval.

- (d) False. For instance, if household size followed the normal curve, there would be many households with a negative number of occupants; we're not ready for that.

- (e) True. Even though household size does not follow the normal curve, you can still use the normal curve to approximate the probability histogram for the sample average.

31. (a) True. This will be a fairly large random sample with replacement from a box.
 (b) No.
 (c) To get the standard units, he should have used the standard error for the average, which is

$$\frac{\text{SD of the box}}{\sqrt{\text{number of draws}}} = \frac{6}{\sqrt{100}} = \frac{6}{10} = 0.6.$$

Then the standard units of the endpoints are $\frac{79.1-80}{0.6} = -1.5$ and $\frac{80.9-80}{0.6} = 1.5$ and the chance is about 86.64%, the area under the curve between -1.5 and 1.5 .

32. The SD is 10 and the standard error for the average is $10/\sqrt{64} = 1.25$.
- (a) 300,007 (the sample average); 1.25 (the SE for the sample average)
 (b) The 95%-confidence interval is “average \pm 2 SEs,” which here comes out to $300,007 \pm 2.5$. That’s between 300,004.5 and 300,009.5 kilometers per second. (Always center a confidence interval at the *sample* average.)
33. The SD of the box can be estimated as 18, so the SE for the average of 400 draws is estimated as $18/\sqrt{400} = 0.9$. If the average of the box is 150, then the average of the draws is 2.67 SEs above its expected value. This isn’t plausible. P is 0.80%, using a two-tailed test (one-tailed: 0.40%).
34. (a) Null hypothesis: the number of correct guesses is like the sum of 1,000 draws from a box with one ticket marked 1 and nine 0’s.
 (b) $\sqrt{0.1 \times 0.9}$. The null hypothesis tells you what is in the box. Use it.
 (c) $z \approx (173 - 100)/9.5 \approx 7.7$, and P is tiny. $\text{SE sum} = \sqrt{0.1 \times 0.9} \times \sqrt{1000} = 0.3 \times \sqrt{1000} \approx 9.5$.
 (d) Whatever it was, it wasn’t chance variation.
35. (a) Tossing the coin is like drawing at random with replacement 90 times from a 0–1 box, with 0 = tails and 1 = heads. The fraction of 1’s in the box is unknown.
 Null hypothesis: this fraction equals $1/2$.
 Alternative: the fraction is bigger than $1/2$.
 The number of heads is like the sum of the draws.
 (b) $z = 1.58$, $P \approx 5.48\%$.
 Calculation: $\text{EV sum} = 90 \times 1/2 = 45$ and
 $\text{SE sum} = \sqrt{90} \times 0.5 = 9.4868 \times 0.5 = 4.7434$.
 That’s because the average and the SD of a 0–1 box with one 0 and one 1 are both equal to 0.5.
 The normal approximation for a count approximates a probability histogram that is made up of blocks. In this case, a correction of $\pm \frac{1}{2}$ is needed, as the sample size is rather small.
 To include 53 in the tail, start the block at 52.5.
 Then $z = \frac{\text{observed sum} - \text{EV sum}}{\text{SE sum}} = \frac{52.5 - 45}{4.7434} = \frac{7.5}{4.7434} = 1.58$.
 To get P , the area of the right tail, look up 1.60 and subtract that 89.04% from 100% to get 10.96%. Then divide by 2 to get $P = 5.48\%$. If you interpolate, P will come out to 5.71%.
 (c) It looks like chance variation is at work, so conclude that the chance of heads is 50%.

36. $\chi^2 = 15.42$, 5 deg.fr., $P \approx 1\%$. Conclude that the die is biased at the 5% level.

Calculations: all expected values are $600/6 = 100$.

$$\frac{(108 - 100)^2 + (93 - 100)^2 + (114 - 100)^2 + (120 - 100)^2 + (93 - 100)^2 + (72 - 100)^2}{100} = \frac{64 + 49 + 196 + 400 + 49 + 784}{100} = \frac{1542}{100} = 15.42.$$