Solutions to Practice Problems for the Final Exam

(Final on Dec. 15)
Math 125, Fall 2023

Multiple Choice. 5 points for each correct response, 1 point deducted for each wrong answer

1. The categories on the horizontal scale are: (B) number of medals.

   The categories are based on the number of medals, so that sets the unit on the horizontal scale. The population is sorted according to medal counts.

2. The height of the block for 12 to 15 medals is: (C) 6% per medal.

   The height is found by dividing the percentage by the width of the base. The result will be densities: percent per unit category. Here it is in % per medal count.

   (Distributing the % over the four separated counts (12, 13, 14, and 15) produces the density: % per medal. Each of those bases is 1 medal wide.)

   Here the endpoints are 11.5 medals and 15.5 medals, as the counting categories are discrete. (Refer to the last paragraph on page 43 and observe the histogram pictured on page 44.) The width is 15.5 medals − 11.5 medals = 4 medals.

   So the answer is 24% divided by 4 medals (see problem 1), which gives 6% per medal.

3. The endpoints of the base of the block for 8 to 11 medals are: (B) 7.5 to 11.5.

   The same argument applies as in problem 2. (Refer to page 43 (bottom) to page 44.)

4. The endpoints of the base of the block for 35 to 40 are: (B) 35 to 41 years.

   Ages are truncated, so 35-year-olds have ages from 35.000 to 35.999. Similarly for 40-year-olds: their ages run from 40.000 to 40.999 years.

5. The height of that block for 35 to 40 is: (C) 2.5% per year.

   Find the percentage of students in that category: $36/240 \times 100\% = 15\%$.

   The height of the block is the percentage divided by the width of the base: 15%/6 years = 2.5% per year. The base was 41 − 35 years = 6 years. (Be careful: the block for 40 years ends at 41.)

6. The vertical scale is a density scale. Answer: (C).

   See section 3.3 in the text.

7. The endpoints of the base of the block for 72 to 77 inches, inclusive, are: (D) 71.5 and 77.5 inches.

   The heights are rounded, so those with height 72 inches have heights from 71.5 to 72.5 inches. Similarly, those with height 77 inches have heights from 76.5 to 77.5 inches.

8. The height of that block for 72 to 77 is: (D) 1% per inch.

   The height is the density found by dividing the percentage by width of the base. So find 6%/6 inches = 1% per inch. The width of the base is found by subtracting the endpoints: 77.5 inches − 71.5 inches = 6 inches.

   (Do not make the mistake of subtracting 72 from 77 to get the width of the base. The width is found from the endpoints, not from the nominal categories.)

9. The area of the block for 72 to 77 inches (inclusive) represents: (E) percent of women.

   This is the basic definition. See the box on page 32 of the text.

10. The area of the block for 72 to 77 inches (inclusive) is: (D) 6%.

    It is the percentage given for the category in question.

11. The average is $\frac{3+6+10+10+12+19}{6} = \frac{60}{6} = 10$.

    The SD is $\sqrt{\frac{49+16+0+0+4+81}{6}} = \sqrt{\frac{150}{6}} = \sqrt{25} = 5$.

    The standard units equals $\frac{\text{obs-avg}}{\text{SD}} = \frac{3-10}{5} = \frac{-7}{5} = -1.4$.

    The answer is: (C) -1.4.

    Alt. Method to get SD (top of pg. 74): avg of squares is (9+36+100+100+144+361)/6=750/6=125.

    Then the square of the average is $10^2 = 100$. SD $= \sqrt{125-100} = \sqrt{25} = 5$.

12. (D) all but 19

    The average is $\frac{3+6+10+10+12+19}{6} = \frac{60}{6} = 10$.

    The SD is $\sqrt{\frac{49+16+0+0+4+81}{6}} = \sqrt{\frac{150}{6}} = \sqrt{25} = 5$.

    1.5 SDs are 1.5 times 5, or 7.5.

    Within 7.5 of 10 (the average) means between $10 - 7.5$ and $10 + 7.5$: between 2.5 and 17.5. In that interval are all but 19. The answer is (C).

    Alt. Method to get SD (top of pg. 74): avg of squares is (9+36+100+100+144+361)/6=750/6=125.
    Then the square of the average is $10^2 = 100$. SD $= \sqrt{125-100} = \sqrt{25} = 5$.

13. (B) -1.361 It is the basic definition of standard units. See the box on page 79.

14. (D) 46%

    The endpoints in standard units are:

    $\frac{66.5-68}{3.6} = \frac{-1.5}{3.6} = -0.41667 \approx -0.40$ and

    $\frac{71-68}{3.6} = \frac{3}{3.6} = 0.8333 \approx 0.85$.

    The area from -0.40 to 0.85 is broken up into two parts:

    The area from -0.40 to 0 (one half of the area from -0.40 to $+0.40$), which comes out to about 15.54%, and

    the area from 0 to 0.85 (one half of the area from -0.85 to $+0.85$), which comes out to about 30.235%.

    Add the two areas to get 15.54% + 30.235%, or about 46%.

15. (A) 5th

    First find the standard unit representation for 62 inches:

    $\frac{62-68}{3.6} = \frac{-6}{3.6} = -1.6667 \approx -1.65$.

    Then find area under the normal curve to the left of -1.65, because the percentile rank is the percentage of all men with heights below 62 inches.

    From -1.65 to $+1.65$ is about 90% of the area. Subtract 90% from 100% to get the area of both tails. We need only the left tail, so divide by 2.

    The area below 62 inches is 5%. By definition, 62 inches is at the 5th percentile.

16. (D) 70.34 inches

    On the normal curve, the 74th percentile is a positive value of z that has areas of 50% below $z = 0$ and 24% between 0 and z.

    When 0 to z is 24%, the area from $-z$ to z will be double: 48%.

    From the table, $z = 0.65$ has total area in the middle of 48.43%. Use $z = 0.65$ as the standard units for the 74th percentile score.

    If the score has standard units of 0.65, by definition it is 0.65 SDs above average. That translates to 0.65 times 3.6 (or 2.34 inches) above 68 inches.

    So, the answer is 70.34 inches, (D).

Alternate Method:

Instead, once 0.65 is found to be the SD, one could use a formula, rather than the verbal approach above.

That way the formula $z = \frac{x - \text{average}}{\text{SD}}$ is used.

$0.65 = \frac{x-68}{3.6}$.

$0.65 \times 3.6 = x - 68$.

$2.34 = x - 68$.

$x = 2.34 + 68 = 70.34$ inches.

17. (B) 12%

 88% of the men are shorter than the man at the 88th percentile, so 12% of them are taller.

 or

 This man is taller than 88% of the men in the distribution. That makes him shorter than 12% of the men. Those men—12%—are taller than the 88th percentile man.

18. (C) 0.45

 First find the average x and the average y:

 avg $x = \frac{1+3+4+5+7}{5} = \frac{20}{5} = 4$ and avg $y = \frac{1+7+3+4+5}{5} = \frac{20}{5} = 4$.

 The find SD of x and SD of y:

 SD of $x = \sqrt{\frac{(-3)^2+(-1)^2+0^2+1^2+3^2}{5}} = \sqrt{\frac{20}{5}} = 2$. Same for SD of y; it equals 2.

 Now convert each of the ten numbers in the original table to standard units, and put the ten results in a new table.

 Proceed like this: $\frac{1-4}{2} = \frac{-3}{2} =$ -1.5.

 The ten results are put in the new table and each row is multiplied across to get the products.

x std. units	y std. units	product
-1.5	-1.5	2.25
-0.5	1.5	-0.75
0.0	-0.5	0
0.5	0.0	0
1.5	0.5	0.75

 Then get the average of the 5 products. They add to 2.25; divide by 5: 0.45.

 That is r. The correlation coefficient, r, for the above data set is: (C) 0.45

 Alt. Method to get SD (top of pg. 74): avg of squares is (1+9+16+25+49)/5=100/5=20.

 Then the square of the average is $4^2 = 16$. SD $= \sqrt{20-16} = \sqrt{4} = 2$.

 Alt. Method to get r (pg. 134, Techincal Note): avg of products is (1+21+12+20+35)/5=89/5=17.8.

 Then the product of the averages is $4 \times 4 = 16$. $r = (17.8 - 16)/(\text{SD of } x \times \text{SD of } y) = 1.8/(4) = 0.45$.

19. Use the simple formula:

predicted y = average y + [(x in standard units) $\times$ r $\times$ (the SD of y)].

The height of 72 inches is $\frac{72-70}{3} = 2/3$ in standard units.

The formula yields: $180 + [2/3 \times 0.4 \times 45] = 180 + 12 = 192$ pounds.

Alternative Solution without the Formula: (See the top half of page 160 in the text.)

The height of 72 inches is 2 inches above 70 inches, the average. That is 2/3 SDs above average.

These men who are 72-inches tall should be r times 2/3 SDs above average in weight. So they are $0.4 \times 2/3 = 0.26667$ SDs above average in weight.

That's 0.266667×45 pounds = 12 pounds. So their average is around 180 + 12 = 192 lbs. Use that estimated average for the prediction. (See page 165 to top of page 166 in the text.)

Answer: (B) 192 pounds

20. (C) 71st

The 91st percentile is a positive z-value that has 50% below 0 and 41% between 0 and z. So, from $-z$ to z will be double 41%: 82%.

From the table, 82% corresponds to $z = \pm 1.35$. Use +1.35, as z is positive.

Now look at Example 2 on page 166 of the text.

This man is 1.35 SDs above average in height. The regression method predicts that he will be $0.4 \times 1.35 \approx 0.55$ SDs above average in weight.

Translate 0.55 in standard units to a percentile rank.

Below 0 is 50% of the area; between 0 and 0.55 is half of 41.77% or about 21%.

The area below this individual is 50% + 21% = 71%. That is his percentile rank for weight. He is at the 71st precentile in weight.

Answer: (C) 71st precentile

21. (A) 0 inches

Association is not causation. (See page 150 in the text.)

22. (D) 41.243 pounds

The formula is $\sqrt{1-r^2}$ $\times$ the SD of the variable being predicted.

Here: $\sqrt{1-r^2} = \sqrt{1-0.4^2} = \sqrt{1-0.16} = \sqrt{0.84} = 0.916515$.

Multiply 0.916515 by 45 pounds to get 41.243 pounds, answer (D).

23. (D) $y = 6x - 240$

The slope of the regression line is: $\frac{r \times \text{SD of } y}{\text{SD of } x}$.

Here, the slope is $\frac{0.4 \times 45}{3} = 6$.

Next write the equation as $y = 6x + b$, plug in a known point (x, y) on the line, and solve for b. The point of averages is always on the regression line. That would be a good choice here. It is $x = 70$, $y = 180$.

$180 = 6(70) + b$, $\quad 180 = 420 + b$, $\quad b = 180 - 420 = -240$. $\quad$ The line is $y = 6x - 240$.

24. (C) 55%.

The chance of a picture card on any draw is 23%. But do not add the chances of a picture card on each draw to find the chance of at least one, since the results are not mutually exclusive. (See the box on page 242 in the text.) It is possible to get a picture card on the first draw and a picture card on the second draw.

Instead, look to the opposite event. The opposite of getting at least one picture card is getting no picture cards: getting all others.

Calculate that chance using conditional probabilities: $\dfrac{40}{52} \times \dfrac{39}{51} \times \dfrac{38}{50} = \dfrac{38}{85} \approx 45\%$.

The chance of the original outcome is 100% minus the chance of the opposite.

So, the probability originally sought is 55%, answer (C).

25. (E) 99%

The chance that not all 3 are picture cards is the opposite of all 3 being picture cards. Find the chance that all 3 are picture cards and subtract that result from 100% to answer the original question.

The chance that all 3 are picture cards is: $\dfrac{12}{52} \times \dfrac{11}{51} \times \dfrac{10}{50} = 11/1105 \approx 1\%$.

The chance that not all 3 are picture cards is $100\% - 1\% = 99\%$, answer (E).

26. (D) 94.3%

The chance that not all 3 are even-numbered is the opposite of all 3 being even-numbered. Find the chance that all 3 are even-numbered and subtract that result from 100% to answer the original question.

With replacement, the results are independent. Just multiply their unconditional probabilites (See page 232.)

The chance that all 3 are even-numbered is: $(20/52)^3 = 0.0569 \approx 5.7\%$.

The chance that not all 3 are even-numbered is $100\% - 5.7\% = 94.3\%$, answer (D).

27. (B) 23.3%.

The draws are independent, so the chance that all three cards are not even-numbered is found by multplying the chances. Each card has 32/52 chance to be non-even-numbered. To find the chance of all three being non even-numbered, find the product of the 3 factors $32/52 \times 32/52 \times 32/52$: $(32/52)^3 = (8/13)^3 = (0.61538)^3 = 0.233 \approx 23.3\%$, answer (B).

28. (C) 40%

Do not add up seven chances of the train being late. The outcomes are not mutually exclusive; the train could be late more than once. (See the box on page 242 in the text.)

Instead look to the opposite outcome. The opposite of being late at least once is not being late at all. That outcome is "the train is on time all 7 days."

Since the results are independent, just multiply together seven factors, each of them the unconditional chance of being on time (0.93). (See the box on page 232 in the text.)

We have just found the chance of the opposite of "being late at least once" to be equal to $(0.93)^7 \approx 0.60 = 60\%$. (Now refer to the second box on page 223 of the text.)

The chance originally requested (of being late at least once) is $100\% - 60\% = 40\%$, answer (C).

29. (B) 32%

This is a binomial probability. The clue is the word "exactly" in the question. Find n, k, and p.

$n = 7$, $k = 1$, $n - k = 6$, $p = 0.07$, and $1 - p = 0.93$.

The binomial formula is: The probability $= \frac{n!}{k! \times (n-k)!} p^k (1-p)^{n-k}$. (page 259)

Here $\frac{7!}{6! \times 1!}(0.07)^1(0.93)^6 \approx 0.32 = 32\%$, answer (B).

30. (B) 100 tosses. It is better for him if the SE for the sum of the draws is small, and the SE for the sum of the draws gets larger as the number of tosses increases. (See figure 1 on page 275 of the text.)

31. (B) 100 tosses. It is better when the standard error for the percentage tends to be larger.

    When the number of tosses increases, the observed percentage tends to get closer to 50%, making the result of between 55% and 60% less likely. So the player is better off with the smaller number of tosses. (See figure 2 on page 276 of the text.)

32. (C) 5%

    This is a binomial probability. The clue is the word "exactly" in the question. Find n, k, and p.

    $n = 30$, $k = 19$, $n - k = 11$, $p = 0.5$, and $1 - p = 0.5$.

    The binomial formula is: The probability $= \frac{n!}{k! \times (n-k)!} p^k (1-p)^{n-k}$.

    Here $\frac{30!}{19! \times 11!} (0.5)^{19} (0.5)^{11} \approx 0.0509 = 5\%$, answer (C).

    (No shortcuts or cancellation, just use a calculator to get the decimal answer.)

    (The normal approximation also gives 5%, as the standard units of the endpoints of the block turn out to be about 1.30 and 1.65.)

33. (E) 6.47%

    This is a binomial probability. The clue is the word "exactly" in the question. Find n, k, and p.

    $n = 20$, $k = 6$, $n - k = 14$, $p = 1/6$, and $1 - p = 5/6$.

    The binomial formula is: The probability $= \frac{n!}{k! \times (n-k)!} p^k (1-p)^{n-k}$.

    Here $\frac{20!}{6! \times 14!} (1/6)^6 (5/6)^{14} \approx 0.064705 = 6.47\%$, answer (E).

    (No shortcuts or cancellation, just use a calculator to get the decimal answer.)

    (The normal approximation is not that accurate here: the sample size (20) is not large.)

34. (B) 1.885%

    This is a binomial probability, with the clue the word "exactly."

    However, most calculators will not return the value of 125 factorial (125!).

    The normal approximation to the probability histogram for the binomial should be used, as the number of tosses is large enough.

    Set up the box model: one 0 and one 1. The average and the SD of this box are both 1/2.

    The number of heads is modeled by the sum of the draws from this 0–1 box.

    Find the EV for the sum of the draws and the SE for the sum of draws.

    They are: $125 \times 0.5 = 62.5$, and $\sqrt{125} \times 0.5 = 5.59$.

    The box of the probability histogram for 72 heads runs from 71.5 to 72.5 heads.

    (See page 317, example 1(a).)

    The standard units for the endpoints are:

    $(71.5 - 62.5)/5.59 \approx 1.60$ and $(72.5 - 62.5)/5.59 \approx 1.80$.

    The area under the normal curve between 1.60 and 1.80 $= 92.81\%/2 - 89.04\%/2 = 1.885\%$.

35. (E) 1.3%

To get the proper counting box needed to model the percentage, put 1 on the ticket with the 7 and put 0 on all the other 31 tickets.

The give or take for the sample percentage is just the SE for the sample percentage, which has the formula: $\frac{\text{SD of the box}}{\sqrt{\text{no. of draws}}} \times 100\%$.

(See the technical note on page 362 of the text.)

Use the shortcut to find the SD of the box:

The fraction of tickets with 1 is 1/32, and the fraction of tickets with 0 is 31/32.

The formula gives: $(1-0)\sqrt{0.03125 \times 0.96875} = \sqrt{0.03027} = 0.17399$.

(See the formula on page 298.)

Then the answer is: $\frac{0.1739}{\sqrt{179}} \times 100\% = 1.3\%$, answer (E).

36. (A) 8.606 to 9.994%

(a) Set up a 0–1 box where 1 is on each Republican ticket and 0 on the others.
(b) Find the sample percent: $(651/7000) \times 100\% = 9.3\%$.
(c) Use the bootstrap to estimate the fraction of 1's in the box: 651/7000 or 0.093.
(d) Estimate the SD of the box by the shortcut: $\sqrt{0.093 \times 0.907} = \sqrt{0.084351} = 0.290432$.
(e) Estimate the SE for the percentage: $\frac{\text{SD of box}}{\sqrt{\text{no. of draws}}} \times 100\% = \frac{0.2904}{\sqrt{7000}} \times 100\% = \frac{29.04\%}{83.666} = 0.347\%$.
(f) Write 95%-confidence interval as sample percent plus or minus two SEs.
(g) Answer is: 9.3%±0.694%. That's 9.3%-0.694% to 9.3%+0.694%, or 8.606% to 9.994%, answer (A).

37. (C) \$623

The standard error for the average is (using technical note on page 415) the SD of the box, divided by the square root of the number of draws. With a large random sample, the SD of the sample may be used to estimate the SD of the box.

Here it gives \$39,400/$\sqrt{4000}$=\$39,400/63.246≈\$623, answer (C).

(The second bullet on page 417 explains what the SE says about how far off the average income in the entire city (the population average) is from the sample average.)

38. (F) 99.968%

The standard error for the average is (using technical note on page 415) the SD of the box, divided by the square root of the number of draws.

Here it gives $5/\sqrt{36} = 5/6 \approx 0.8333$.

A sample of size 36 should be large enough for the normal approximation to apply.

Find the standard units for 97 and 103: $(97-100)/0.8333 = -3.6$ and $(103-100)/0.8333 = 3.6$.

The area under the normal curve from -3.6 to $+3.6$ is just the table entry: 99.968%, answer (F).

39. (C) 208.9 to 209.1

The standard error for the average is (using technical note on page 415) the SD of the box, divided by the square root of the number of draws. With a large random sample, the SD of the sample may be used to estimate the SD of the box.

Here it gives SE for the average of the draws $= 2/\sqrt{1600} = 2/40 = 0.05$.

The 95%-confidence interval is sample avg $\pm$ 2 $\times$ SE avg: $209 \pm 2(0.05) = 209 \pm 0.1$, which is 208.9 to 209.1, answer (C).

40. (B) a through c

    a. The errors come from an error box; the box is the same for each weighing.

    b. The errors are drawn from the box at random.

    c. The average of the error box is 0.

    The Gauss model does not need a large number of draws; that only comes into play when the normal approximation is to be used. And the Gauss model is silent as to whether the tickets in the error box follow the normal curve. In fact, even when applying a confidence interval and a normal curve for the sample averages, the tickets themselves need not follow the normal curve.

    Only the first three assumptions are needed to apply the Gauss model.

41. (A) a through d

    a. The errors come from an error box; the box is the same for each weighing.

    b. The errors are drawn from the box at random.

    c. The average of the error box is 0.

    d. There are a large number of draws.

    These four are necessary for the confidence interval. The first three allow the Gauss model; the large number of draws is necessary for the normal curve to apply. That sets up the Gauss model confidence interval.

    The last two assumptions are not necessary. The first of them (letter e.)—a normal distribution for the tickets themselves—is clearly not needed. (See the box on page 412.)

    The 2nd of them (letter f.) is not needed either. (See the box on pg 416.) That estimate of the SD of the error box is good enough for the confidence interval when the number of draws is large (letter d.).

    e. The tickets in the error box are normally distributed. **(Not needed.)**

    f. The SD of the errors must have been known in advance. **(Not needed.)**

    Answer: (A) a through d

42. (C) 17 pounds

    The SD of the weighings estimates the SD of the error box.

    (See the box on page 451 in the text.)

    A weighing is modeled by drawing a ticket from the error box and adding its value to the exact weight. A typical ticket will be off from 0 (the average of the error box) by about the SD: 17 pounds. (See the box on page 67 in the text.)

43. (A) 1.49 pounds

    The actual weight of the truck is the population average for weighings. The sample averge is 12,347 pounds. Now refer to the second bullet on page 417 of the text.

    The SE for the sample average (using the Technical Note on the top of page 415) is found to be $\dfrac{\text{SD of box}}{\sqrt{\text{no. of draws}}} = \dfrac{17}{\sqrt{130}} = 1.49$ pounds. (Refer to page 416 to estimate the SD of the box.)

    So, the sample average (12,347 pounds) is off from the population average (the actual weight of the truck) by about 1.49 pounds.

    Answer: (A) 1.49 pounds

44. (C) 12,344.02 to 12,3349.98 pounds

    This problem continues the previous problem (#43). Refer to the answer above to find the SE for the average.

    The 95%-confidence interval is just the sample average ± 2 SE averages.

    It is $12{,}347 \pm 2(1.49) = 12{,}345 \pm 2.98$.

    That is 12,344.02 to 12,349.98 pounds, answer (C).

45. (C) $p = 6.055\%$, fair

The null hypothesis is that the coin is fair, with a box having one 0 and one 1.

The alternative hypothesis is that the coin gets too many heads.

Model the number of heads for the coin by the sum of the draws from the box.

The average and SD of the box are both 1/2.

The EV for the sum of the draws is number of draws $\times$ average of the box.

The SE for the sum of the draws is $\sqrt{\text{number of draws}} \times$ SD of the box.

Here: EV $= 34 \times (1/2) = 17$, and SE $= \sqrt{34} \times (1/2) = 2.915$.

Now, to approximate the chance of getting 22 or more heads, draw the block for 22 heads for the probability histogram. It runs from 21.5 to 22.5 heads. Then the endpoint for 22 or more heads (the tail) is 21.5. The chance for this result or any result more extreme is needed.

Using 22 in the standard units calculation is wrong. The continuity correction ($\pm 1/2$) is needed here because the number of tosses is small (large enough for the normal curve, but at the same time small enough to need the $\pm 1/2$ correction). Read the first paragraph on page 317 of the text and review example 1 below.

The normal curve may be used to approximate the chances for the sum of the draws because the number of draws is large enough (not too large but more than 25). See the Central Limit Theorem in the box on page 325 of the text.

The standard units are: $\frac{21.5-17}{2.915} \approx 1.55$.

The central area for $z = \pm 1.55$ is 87.89%. Two tails are 100% $-$ 87.89% = 12.11%.
The right tail—which is P—is half of that: 6.055%.

Since $P > 5\%$, conclude it was just chance error from a fair coin.

Answer: (C) $P = 6.055\%$, fair

46. (A) $p = 3.215\%$, not fair

The null hypothesis is that the die is fair, with a box having five 0's and one 1.

The alternative hypothesis is that the coin gets too many fives.

Model the number of fives for the die by the sum of the draws from the box.

The average of the box is 1/6, and the SD of the box is $(1-0)\sqrt{(1/6)(5/6)} = 0.37268$,

The EV for the sum of the draws is number of draws $\times$ average of the box.

The SE for the sum of the draws is $\sqrt{\text{number of draws}} \times$ SD of the box.

Here: EV $= 90 \times (1/6) = 15$, and SE $= \sqrt{90} \times 0.37268 = 3.53555$.

Now, to approximate the chance of getting 22 or more fives, draw the block for 22 fives for the probability histogram. It runs from 21.5 to 22.5 fives. Then the endpoint for 22 or more fives (the tail) is 21.5. The chance for this result or any result more extreme is needed.

Using 22 in the standard units calculation is wrong. The continuity correction ($\pm 1/2$) is needed here because the number of tosses is small (large enough for the normal curve, but at the same time small enough to need the $\pm 1/2$ correction). Read the first paragraph on page 317 of the text and review example 1 below.

The normal curve may be used to approximate the chances for the sum of the draws because the number of draws is large enough (not too large but more than 25). See the Central Limit Theorem in the box on page 325 of the text.

The standard units are: $\frac{21.5-15}{3.53555} \approx 1.85$.

The central area for $z = \pm 1.85$ is 93.57%. Two tails are 100% $-$ 93.57% = 6.43%.
The right tail—which is P—is half of that: 3.215%.

Since $P < 5\%$, conclude it was not chance error, but a real difference.

Answer: (A) $P = 3.215\%$, not fair. (Note that using $\frac{22-15}{3.53555}$ would have led to a wrong value of P.)

47. (B) 4%, unfair

    There are 5 degees of freedom: one less than the number of categories.

    The expected values are all $120 \times (1/6)$, or 20.

    For each category, subtract the expected value from the observed, square the result, and then divide by 20, the expected.

    Here is the first calculation: $\dfrac{(15-20)^2}{20} = 25/20 = 1.25$.

    Add the five results. This is χ^2: $\quad 1.25 + 2.45 + 0.05 + 0.45 + 4.05 + 4.05 = 12.3$.

    Or, since they are all over 20, add the six squares of the (observed − expected) and then divide by 20: $\frac{25+49+1+9+81+81}{20} = \frac{246}{20} = 12.3$.

    From the χ^2-table, the value of P is estimated at 4%. It's between 1% and 5%, closer to 5%.

    Answer: (B) 4%, unfair

48. (a) False.
    (b) True.
    (c) True.
    (d) True.
    (e) False.
    (f) False.
    (g) True.
    (h) False.
    (i) False.

    A: boxes on pages 40 and 41.

    B: See the lower third of page 43, last paragraph, sentence starting with "With a discrete variable $\cdots$, and Figure 6 on page 44..

    C: box on page 67.

    D: box on page 79 and Example 1 on page 80.

    E: box on page 340.

    F: False. It will be off by the SE for the average, not the SD of the box.
    See the last sentence on page 410 and summary note 2 on page 437.

    G: See figure 2 on page 276 and the last comment by *Assistant* on pages 276 to 277.

    H: False, because of chance error. The sample percentage is likely to be close to the population percentage, but not exactly equal. The SE for the percentage sys how far off you can expect to be.
    See the box on page 386.

    I: See to top box on page 412.

49. Because the sample was random and rather large, it is *representative* of the population as a whole. That's the theory of inferential statistics, based on the Central Limit Theorem, the square root law, the bootstrap, and the fact that a percentage works like the sum of the draws, when put into standard units. We can be confident in that sample and its associated interval, knowing that chance errors for the sample percentage are normally distributed.