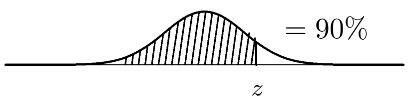
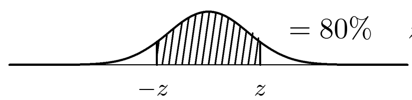
Answers to the Study Guide for the Fall 2020 Final

1. (a) around 85% (b) 40–80 cigarettes (c) 0–10 cigarettes (d) 30%
2. (a) around 99% (b) 140–150 mm (c) 135–140 mm (d) About $5 \times 2.1\% = 10.5\%$
3. (a) B (b) 20% (c) 70% (d) 70 (e) 70 (the same) (f) 87 (g) 28 points
4. (a) Approximately equal to the area under the curve between -1.5 and 0.5, or 62.465%.
   (b) Approximately equal to the area under the curve below 1.5, or 93.32%.
5. (a) 79% (b) 38% (c) 50%; the point of averages is always on the regression line. (d) 50%; predict average GPA.

Work for (a):

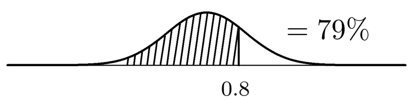


= 90%



= 80% $z \approx 1.30$

In standard units, his SAT score was 1.30. The regression prediction for his first-year score is $0.6 \times 1.30 = 0.78 \approx 0.80$ in standard units.



= 79%

This corresponds to a predicted percentile rank of 79% for that student's GPA.

Details of converting percentile to z: Take the given 90% as the area below z. That means that the upper tail must be 10% and the so-far-unpictured lower tail must be 10%. Take the lower tail from the given percentile rank and you will have the percentage value that must be looked up on the table: 80%. That gives $z = +1.30$, because we want the value that is positive.

It is also possible to do the first step with a slightly different logic. Take the 90%, the given percentile rank and break it into 2 parts: the area below 0 (50%), and the area between 0 and z (40%). Then double the 40% to get the middle area (from $-z$ to z) for looking up on the table. Again look up 80% on the normal table and find z.

Details of converting z to a percentile, as required in the last step: $z = 0.8$ has 58% in the middle. Below it will be the 50% below 0 and the 29% from 0 to z. (This 29% is half of the middle area for $z = 0.80$.) The sum is 79%. That's the area below z, so by definition it is the percentile rank.

Alternative logic: 58% in the middle, means $100\% - 58\% = 42\%$ for the two tails. One will be 21%. We need all area except one tail $= 100\% - 21\% = 79\%$.

This corresponds to a predicted percentile rank of 79% for that student's GPA. (Usually round off to nearest whole perecntage.)

Work for (b): The lower tail is given to be 30%. The upper tail will also be 30%. That leaves 40% in the middle. Look up that percentage value on the table: 40%. It gives $z = -.525$, because we want the value that is negative and we go halfway between $z = 0.50$ and $z = 0.55$, because 40% is about halfway between 38.29% and 41.77%.

It is also possible to do the first step with a slightly different logic. Take the 30%, the given percentile rank and subtract it from 50% to get the area between $-z$ and 0 (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, using the negative endpoint.

Next multiply $-.525$ by r, getting $0.6 \times -0,525 = -0.315$, so use -0.30 as the endpoint.

Details of converting z to a percentile, as will be required in the last step: $z = -0.30$ has slightly less than 24% in the middle. The two tails together will add to slightly more than 76%. Divide by 2 to get 38%, to the nearest percent, for the lower tail. And the lower tail is the percentile rank.

The predicted percentile rank for that student's GPA is 38%. (Generally round the answer to the nearest whole number.)

Answer: 38%.

6. (a) $\left(\frac{1}{6}\right)^{13} \approx 0\%$. (It is actually about one chance in thirteen billion.)

   (b) $1 - \left(\frac{5}{6}\right)^{13} \approx 90.654\%$.

   (c) $\left(\frac{5}{6}\right)^{13} \approx 9.346\%$.

   Note. The outcome in part (b) was just the opposite of the outcome in part (c), so find the answer to part (c) first; then subtract it from 100%.

   (d) The opposite event as part (a), so $100\% - 0\% \approx 100\%$.

   (e) This is logically the same thing as part (d), so 100%.

7. (a) False; these events aren't mutually exclusive, so you can't add the chances. (To find the chance, read chapter 14, section 4, on the paradox of the Chevalier de Méré.)

   The actual chance is about 42%.

   (b) False: same reason.

   Actually, we know that the chance of getting a tail on the first toss is 50%; obviously the chance of getting at least one head is less than 100%, since you could also get a tail on the second toss after getting a tail on the first toss, and that would be a case of not getting at least one head among the two tosses. That proves that getting at least one head among the two tosses is not a sure thing and therefore does not have a chance of 100%. (To find the chance, read chapter 14, section 4, on the paradox of the Chevalier de Méré.))

   The actual chance is 75%.

8. $1 - (35/36)^{36} \approx 64\%$.

9. This is like Chevalier de Méré; the chance is $1 - (3/5)^4 \approx 87\%$.

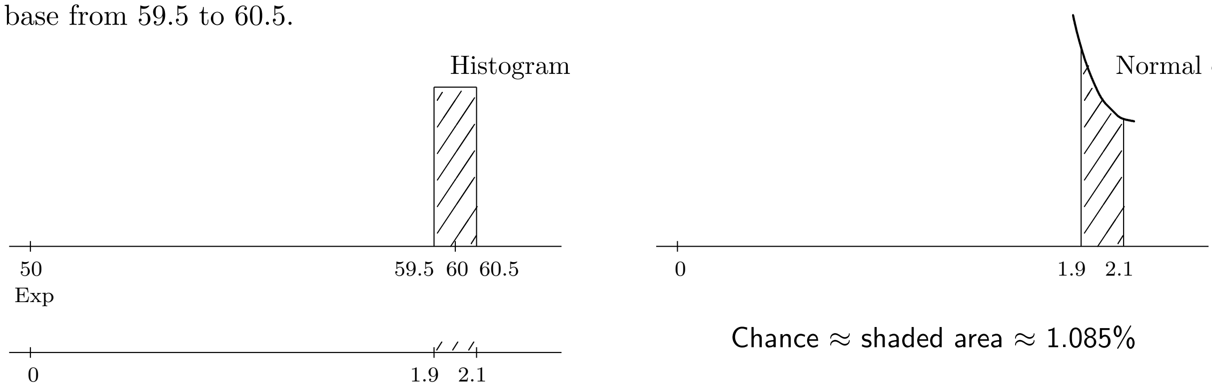
10. $1 - 81/256 = 175/256 \approx 68.36\%$.

    The Hints gave: (i) 3/4; (ii) $(3/4)^4 = 81/256 \approx 31.64\%$; (iii) the same as part (ii): $81/256 \approx 31.64\%$; then get the final answer by subtracting the answer to (iii) from 100%.

    This problem (a paradox of the Chevalier de Méré) is best solved by making these observations:

    - The four things are not mutually exclusive, so the Addition Rule does not apply.
    - It is with replacement, so the draws are independent.
    - Simple multiplication without conditionals suffices for the chance of "all."
    - The results "all the tickets are blank" and "there is no star in the four tickets drawn" are the same.
    - The result "at least one ticket has the star" and the result in (iii) are logical opposites.
    - Since "at least one" and (iii) are opposites, to get the chance of at least one, just subtract the chance of (iii) from 100% (or from 1 if (iii)'s chance is in fractional or decimal form). That answers the question.

11. The expected number of heads is 50; the SE is 5. You want the area of the rectangle over 60 in the probability histogram pictured in figure 3, page 315. That rectangle, being for counting numbers, has a base from 59.5 to 60.5.



Use standard units of the endpoints on the normal curve to get a valid approximation. The area under the normal curve between 1.9 and 2.1 is approximately 1.085%.

(The exact chance from the binomial formula is: 1.084%, so that is the precise area under the block in the probability histogram.)

12. The expected number is 30, and the SE is about 3.5.

13. The number of aces in 720 rolls of a die is like the sum of 180 draws from the box | 1 | 0 | 0 | 0 | 0 | 0 | (with 1 ticket marked "1" and 5 tickets marked "0"). The number of aces will be around 120, give or take 10 or so. There is about a 87.89% chance that the number of aces will be in the range 105 to 135. About 87.89% of the people should get a number in that range. (Because the number of draws is large, the Central Limit Theorem justifies using the normal curve to approximate the chances of the probabilty histogram.)

Because the probability histogram has blocks centered at each number with endpoints of the block $\pm\frac{1}{2}$ away, we need to use 104.5 to 120.5 to calculate the standard units for the endpoints of the region under the normal curve. If these corrections were not made, the answer would have been found as 86.64%.

14. The fraction of 1's in the box is 0.25. The expected value for the sum is $300 \times 0.25 = 75$. The SD of the box is $\sqrt{0.25 \times 0.75} \approx 0.433013$. The SE for the sum is $\sqrt{300} \times 0.433013 \approx 7.5$.

    (a) The expected value for the percent is 75 out of 300, or 25%. (Or just note that the expected value for the percentage of 1's among the draws is equal to the percentage of 1's in the box: 20,000/80,000/×100% = 25%.)

    The SE for the percent is 7.5 out of 300 times 100%, which comes out to 2.5%.

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

    $$\frac{\text{SD of the box}}{\sqrt{n}} \times 100\% = \frac{0.433013}{\sqrt{300}} \times 100\% = 2.5\%.$$

    (No surprise about the expected value: 25% of the tickets in the box are 1's.)

    (b) The percentage of 1's among the draws will be about 25.0%, give or take 2.5% or so. Parts (a) and (b) require the same calculations; in (b) you have to interpret the results.

    (c) This is ±2 SE; the chance is about 95.45% (no continuity correction; percentages are already continuous).

15. The sample is like 100 draws made at random from a box which has one ticket for each employee, showing the number of days that employee was absent. Null hypothesis: the average of the box is 6.3 days. Alternative hypothesis: the average of the box is less than 6.3 days. The SD of the box is estimated at 2.9 days, so the SE for the average is 0.29 days, and $z \approx (5.5 - 6.3)/0.29 \approx -2.8$, so $P \approx 0.3$ of 1%. This is strong evidence against the null; chance variation will not explain the drop in absenteeism.

16. The TA's null hypothesis: the scores in his section are like 90 draws from a box containing all 900 scores. (There is little difference between drawing with or without replacement, because the box is so big.) The null hypothesis specifies the average and the SD of the box, 63 and 20. The EV for the average of the draws is 63, and the SE is 3.65. So $z = (\text{obs} - \text{exp})/\text{SE} = (55 - 63)/3.65 \approx -2.2$, and $P \approx 1\%$. The TA's defense is not good.

17. $\chi^2 = 13.2$, $d = 5$, $1\% < P < 5\%$; actually, $P = 2.2\%$.

    d = degrees of freedom.

    Comment. The data do not fit the model so well.

    Calculation of χ^2*:* $(5^2 + 3^2 + 7^2 + 6^2 + 2^2 + 3^2)/10 = (25 + 9 + 49 + 36 + 4 + 9)/10 = 132/10 = 13.2.$

18. This is data snooping. If 85 different hypotheses are tested—some results are likely to be significant. To single out those five results and ignore the other 80 results is a major error in statistics. When 85 tests are run at the 5% significance level, there is a 5% chance that a fair lottery will be judged unfair by the test. Out of 85 tests run, around 4 or 5 of them would indicate that those lotteries are not truly random. So chance error is a very reasonable explanation, when all 85 tests are considered together.

    Comment. The trouble with data snooping is that it makes significance levels close to meaningless. Those who snoop, find—even when nothing is goint on.