Answers to the Sample Final for Spring 2015

1. 2%.

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

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

4. (a) $r = -0.93$, by calculation. (b) $r = 0.82$, by calculation.
   (c) No calculation is necessary; $r = -1$. The points all lie on a line sloping down, $y = 8 - x$.

5. (a) $(5/6)^{10} \approx 0.16$, 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.)
   (b) i. 1/8
   ii. $1 - 1/8 = 7/8$
   (c) i. $\left(\frac{12}{13}\right)^{40} \approx \frac{1.46977 \times 10^{43}}{3.61189 \times 10^{44}} \approx 0.04069 \approx 4.07\%$.
   ii. $100\% - 4.07\% = 95.93\%$.
   iii. $\left(\frac{25}{26}\right)^{40} \approx 0.208289 \approx 20.83\%$.

6. The chance of getting one 5 on the red die is

$$\frac{6!}{1!5!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^5 = 3125/7776 \approx 40.18776\%.$$

The chance of getting one 5 on the blue die is

$$\frac{4!}{1!3!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3 = 125/324 \approx 38.58\%.$$

The rolls of the red die are independent of the rolls of the blue die. We want the chance that both happen; so use the multiplication rule for independent things. The answer is

$$\frac{6!}{1!5!}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^5 \times \frac{4!}{3!1!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3 = \frac{3125}{7776} \times \frac{125}{324} = \frac{3125 \times 125}{7776 \times 324} \approx 15.5\%.$$

7. The expected value is 400, the observed value is 431, the chance error is 31, the standard error is 20.

   First find the average and the SD of the box: 4 and 2. Then multiply the average of the box by the number of draws to get the expected value, and multiply the SD of the box by the square root of the number of draws to get the standard error.

   The observed value is just what sum of the draws turned out to be, and the chance error is just the observed value minus the expected value.

8. (a) is false. The percentage of 1's among the draws will be off its expected value, due to chance error. The SE for the number of 1's among the draws is $\sqrt{500} \times \sqrt{0.25 \times 0.75} = 9.7$. The SE for the percentage of 1's among the draws is $(9.7/500) \times 100\% = 2\%$, approximately.

   (b) is false: Chance error gets smaller and smaller as the sample size gets very large, but it is not eliminated. The errors become many orders of magnitude smaller and for all practical purposes they can often be ignored.

   Chance error is tamed, but not eliminated.

   Observation. It can be argued that the roots of the worldwide financial collapse of 2008 and 2009 were in the cavalier attitude of so-called "experts" to very small inefficiencies and very small probabilities. There was a feeling of confidence and self-assurance; but that tiny chance error proved to be devastating.

9. Model: there is a box with 50,000 tickets, one for each household in the town. The ticket shows the commute distance for the head of household. The data are like 1000 draws from the box. The SD of the box is unkown, but can be estimated by the SD of the data, as 9.0 miles. The SE for the sum of the draws is estimated as $\sqrt{1000} \times 9 \approx 285$ miles, and the SE for the average is estimated as $285/1000 \approx 0.3$ miles.

   Or get the SE for the average directly by dividing the SD by the square root of the number of draws: $9/\sqrt{1000} \approx 0.3$ miles.

   (a) The average commute distance of all 50,000 heads of households in the town is estimated as <u>8.7 miles</u>; this estimate is likely to be off by <u>0.3 miles</u> or so.

   (b) 8.7 miles ± 0.6 miles.

10. (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's in the box. Use it.

    (c) $z \approx (173 - 100)/9.5 \approx 7.7$, and P is tiny.
    (SE sum $= \sqrt{1{,}000} \times$ SD of box $= 31.62 \times \sqrt{0.1 \times 0.9} = 31.62 \times 0.3 \approx 9.5$.)

    (d) Whatever it was, it wasn't chance variation.

    We might explain it like this: an average person would get one in 10 correct, so that is about 100 correct out of the 1,000. This individual comes along and gets considerably more correct and claims that he has ESP. The skeptic would counter by saying that he just got lucky. The statistician with his math can tell us that people do get lucky, but not that lucky. To get that lucky would be less than one chance in a trillion. Such a position that is was chance error is untenable: it is without a doubt not a reasonable explanation.

11. (iii)

12. The <u>null</u> hypotheis says that the difference is due to chance but the <u>alternative</u> hypothesis says that the difference is real.

13. (a) True. See pages 480 to 481. Even though the hypotheses stay the same, different data give different z and P.

    (b) True. Big P is good for null.

    (c) True. Small P is bad for null.

    (d) True. See page 479. It's just the definition.

    (e) True; $z = (\text{obs} - \text{exp})/\text{SE}$, and "exp" is computed on the null. Just solve the equation for "obs."

14. It is possible that the claim could be correct. If Pedroia had a lot of at-bats in the second half of the season, while Napoli had more at-bats in the first half, then Napoli could end up with a higher batting average for the entire season.

    For example, suppose Pedroia had 60 at-bats in the first half and 200 at-bats in the seoond half, while Napoli had 240 at-bats in the first half but only 50 at-bats in the second half.

    The calculation is as follows:

    Pedroia: 260 at-bats with hits $60 \times .300 + 200 \times .250 = 18 + 50 = 68$. Batting average is $68/260 = .262$.

    Napoli: 290 at-bats with hits $240 \times .275 + 50 \times .240 = 66 + 12 = 78$. Batting average is $78/290 = .269$.

    This phenomenon is called Simpson's Paradox.