Solutions to Old Sample Final

Math 125 Spring 2009

1. (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.
   (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 $z = 0.25$.
   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.

2. (a) 33
   (b) There is no mode (most common value on the list).
   (c) The average is 33; the deviations are 0, 4, -1, -6, -5, 3, 5; the SD is
   $$\sqrt{\frac{0^2 + 4^2 + (-1)^2 + (-6)^2 + (-5)^2 + 3^2 + 5^2}{7}} = \sqrt{\frac{112}{7}} = \sqrt{16} = 4.$$
   (d) $\frac{1}{2}$ SD is $\frac{1}{2} \times 4 = 2$. Within 2 of 33 is the range 31 to 35, containing two 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 28,
   $$\text{standard units} = \frac{28-33}{4} = \frac{-5}{4} = -1.25.$$

3. (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
   $$\frac{(0 \times 0) + (0 \times -1) + (-2 \times -1.5) + (0 \times 0.5) + (1 \times 0.5) + (1 \times 1.5)}{6}$$
   $$= \frac{5}{6} \approx 0.8333.$$
   $r = 0.8333$.
   (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 $= \sqrt{1-r^2} \times$ the SD of y $= \sqrt{1-(0.8333)^2} \times 2 = \sqrt{1-.6944} \times 2 = \sqrt{.3056} \times 2 = 0.5528 \times 2 \approx 1.10554$.
   (d) The slope of the regression line is
   $$r \times \frac{\text{SD of } y}{\text{SD of } x} = \frac{5}{6} \times \frac{2}{1} = \frac{5}{3} \approx 1.6667.$$

4. (a) i. The things "getting a king on the first card" and "getting a king on the second card" are not mutually exclusive. So simply adding up the chances for all six cards drawn being a king will not give the correct answer. It will double count the chances that more than one result occurs. The answer $\frac{4}{52}+\frac{4}{52}+\frac{4}{52}+\frac{4}{52}+\frac{4}{52}+\frac{4}{52}=\frac{6}{13}\approx 46.15\%$ will be too big. Our answer will have to be smaller than 46.15%.
The best way to find the desired result is to go to the opposite of getting at least one king. That is getting all non-kings. Since there is no replacement, the draws are not independent. We need conditionals. The chance that all six cards are not kings is the chance that the first card drawn is not a king, multiplied by the conditional chance that the second card drawn is not a king, given that the first card drawn was not a king and it was not replaced, etc. (for all six draws).

$$\text{P(no kings)} = \frac{48}{52}\times\frac{47}{51}\times\frac{46}{50}\times\frac{45}{49}\times\frac{44}{48}\times\frac{43}{47} = \frac{1}{13}\times\frac{47}{51}\times\frac{23}{5}\times\frac{3}{49}\times\frac{11}{1}\times\frac{43}{47} = \frac{1533939}{2544815} \approx 60.28\%.$$

The chances of the opposite, getting at least one king, is then $1-\frac{1533939}{2544815}=\frac{1010876}{2544815}=39.72\%$. Or simply subtract 60.28% from 100% to get 39.72%.

ii. The results of getting a king on each draw are still not mutually exclusive. Our answer will have to be less than 46.15% when the cards are drawn with replacement.
The chance of the opposite, getting all non-kings, will still be needed. Here with replacement, the draws are independent and conditional probabilities are not needed. The chance of all non-kings is obtained by multiplying the chance of a non-king on any draw, $\frac{48}{52}=\frac{12}{13}$, by itself six times. That gives

$$\left(\frac{12}{13}\right)^6 = \frac{12^6}{13^6} = \frac{2985984}{4826809} \approx 61.86\%.$$

The chances of the opposite, getting at least one king, is then $1-\frac{2985984}{4826809}=\frac{1840825}{4826809}\approx 38.14\%$. Or simply subtract 61.86% from 100% to get 38.14%.

(b) i. The chance that several things will all happen equals the chance that the first will happen, mutiplied by the chance that the second will happen given the first has happened, etc. Draws made without replacement require conditionals.

$$\text{P(all numbered cards)} = \frac{36}{52}\times\frac{35}{51}\times\frac{34}{50}\times\frac{33}{49}\times\frac{32}{48}\times\frac{31}{47} = \frac{34782}{363545} \approx 9.57\%.$$

ii. Here, since the draws are made with replacement, they are independent. The chance that all six will happen equals the product of the six unconditional probabilities. (This is a special case of the multiplication rule, stated in part (b)i.)

$$\text{P(all numbered cards)} = \left(\frac{36}{52}\right)^6 = \left(\frac{9}{13}\right)^6 = \frac{9^6}{13^6} = \frac{531441}{4826809} \approx 11.01\%.$$

5. 10%—use the binomial formula. The conditions are met: two possible results on each draw, we want to find the chances that one of the results occurs k times in the n draws, n is fixed in advanced, p—the chance of the specified result occurring—is the same on each trial, and the draws can be assumed to be independent (since with replacement).

   $n = 6$, $k = 2$, $n - k = 4$, $p = 0.10$, and $1 - p = 0.90$.

   The chance that exactly two of the six will not own cars is given by

$$\frac{6!}{2! \times 4!}(0.10)^2(0.90)^4 = 15(0.01)(0.6561) = 0.098 \approx 10\%.$$

6. (a) The exact chance is $\frac{40!}{20! \times 20!}(.5)^{20}(.5)^{20} \approx 12.54\%$.
   Use the binomial formula to find the chance that an event with chance 1/2 will occur exactly 20 times out of 40. Here $n = 40$, $k = 20$, $n - k = 20$, $p = 0.5$, and $1 - p = 1 - 0.5 = 0.5$.

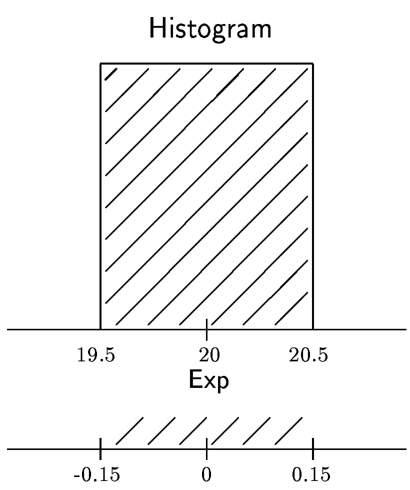
   (b) The expected number of heads is 20; the SE is 3.1623. Since the SD of a zero-one box with one 0 and one 1 is 0.50, and the SE is the product of the square root of the number of draws and the SE, we get $\sqrt{40} \times 0.50 = 6.3246 \times 0.50 = 3.1623$.

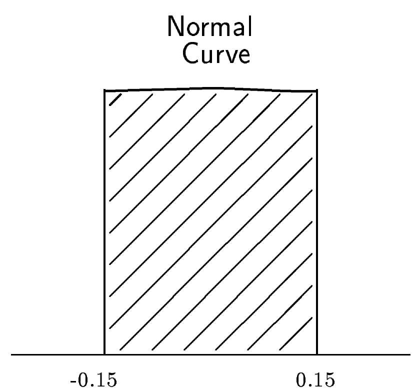
   We seek the area of the rectangle over 20, and need to find the standard units for its endpoints, 19.5 and 20.5.

   That gives us $\dfrac{19.5 - 20}{3.1623} = -0.158$ and $\dfrac{20.5 - 20}{3.1623} = 0.158$.

   For our table, round them to the nearest 0.05, getting +0.15 and −0.15.

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





Chance ≈ shaded area
≈ 11.92% (See comment below.)

Comment: the exact chance is ≈ 12.54%.

With interpolation, a better estimate for the area of the block is $11.92\% + \frac{0.008\%}{.05\%} \times (15.85\% - 11.92\%) = 11.92\% + 0.6288\% = 12.5488\%$, a really close approximation.

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.

7. Model: each measurement equals the exact elevation, plus a draw from the error box. The tickets in the box average out to 0. Their SD is unkown, but can be estimated by the SD of the data, as 30 inches. The SE for the sum of the 25 measurements is estimated as $\sqrt{25} \times 30 = 150$ inches. The SE for the average is estimated as $150/25 = 6$ inches.
   (a) 30 inches. Each reading is off the exact elevation by an error drawn from the error box. We estimate from the sample that the SD of the error box (the give or take) is about 30 inches. That's the bootstrap method and 25 readings, although not very many, should be just about enough to begin to rely on such an estimate.
   (b) False. The average is 81,411 inches exactly.
   (c) True. The interval is "average ± 2 SEs."
   (d) False. This mixes up SD and SE.
   (e) False, same issue.

8. (a) $\chi^2 = 15.42$, 5 deg.fr., $P \approx 1\%$. Conclude that the die is biased at the 5% level.
   Calculations: all expected values are $600/6 = 100$.

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

   (b) The null hypothesis is that threes come out as expected from a fair die, one in six tosses in the long run. The alternate hypothesis is that threes come out more than one in six tosses in the long run.

   Make a box with six tickets, five 0's and one 1. This will classify the draws for counting the number of threes. The average of this box is $\frac{1}{6}$ and its SD is $(1-0)\sqrt{\frac{1}{6} \times \frac{5}{6}} = \sqrt{\frac{5}{36}} = 0.372678$.

   The expected number of threes in 600 tosses is now $600 \times \frac{1}{6} = 100$ and the standard error of the sum is $\sqrt{600} \times 0.372678 = 9.1287$.

   A normal curve will approximate the exact answer, since this involves the sum of the draws and the number of draws is large.

   Now convert the observed number of threes to standard units. Assume that the null hypothesis is true, and determine how often a result of 114 or more heads should occur by chance alone. This is the P value.

   Since the exact probabilities are represented by blocks (as in problem 6(b)), it would be good to use the continuity correction. 114 or more heads includes the rectangle for 114 and all others to the right. So, the starting endpoint for this tail area is 113.5.

   The standard units for 113.5, assuming that the null is true, are:

$$\frac{\text{observed sum of the draws} - \text{expected sum of the draws}}{\text{standard error of the sum of the draws}} = \frac{113.5-100}{9.1287} = 1.48 \approx 1.5.$$

   The tail area is $\frac{100\% - 86.64\%}{2} = 6.68\%$. $P = 6.7\%$. This is larger than 5%, so we attribute it to chance error. Conclude that the die is fair. This does *not* agree with the answer in part (a).

i. There are two reasons why this test for the threes does not shed any light on the fairness of the die.
Firstly, we have already run the χ^2-test on the data and included the number of threes in that calculation.
Also, there is a serious issue as to why we have chosen the threes to compare to their expected value. Why did we not look at the sixes? This is a form of data snooping: looking at the results before deciding which test to run. There is nothing inherent in the threes that makes one want to test them instead of one of the other results. Even if the die were biased, it is possible that one or more of the categories would be somewhat close to the expected value.