Homework 5 Solutions

p329 # 12

Part (a)

No. To estimate the chance you need the expected value and standard error. To do get the expected value you need the average of the box and to get the standard error you need the standard deviation of the box.

Part (b)

You can do it without the actual numbers as long as you have the average and the standard deviation. Since the numbers are bounded between -10 and 10 you can assume that with 1000 draws the probability histogram for the sum will be approximately normal.

To use the normal distribution you need to know the expected value and the SD of the the normal distribution (also called the SE). The expected value of the sum of 1000 draws is just 1000(AVG of the box). The SD for the sum is $\sqrt{1000}(SD$ of the box). So you do have enough information.

p352 # 11

Part (a)

In this question the box that the car is in is fixed, since the car has already been hidden. Suppose the car is in box B. Let X_i be a random variable that counts whether the i-th coin flip is heads. Then we are looking at the sum $B+X_1+X_2+\cdots+X_{100}$. So if the sum is 65 then we have $65 = B+X_1+\cdots+X_{100}$ so we can solve for B to get $B = 65-(X_1+X_2+\cdots+X_{100})$.

Now we are looking for the probability histogram for this sum. We can compute the expected value:

$$\begin{aligned}\mathbb{E}(65-(X_1+\cdots+X_{100})) &= 65-\mathbb{E}(X_1+\cdots+X_{100})\\ &= 65-100\mathbb{E}(X_1)\\ &= 65-50\\ &= 15\end{aligned}$$

Since this sum is going to follow a normal distribution (by the Central Limit Theorem) the Porsche would be most likely to be in the boxes closest to the expected value. So you should look in 10,11,12,13,14,15,16,17,18,19 and 20.

Part (b)

Here the expected value for B is computed just like in the last question, which comes out to be 45. So you would look in boxes 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50.

Part (c)

In general if the sum is N you would look in boxes $N-55$ to $N-45$.

Part (d)

To figure out the chances you need the SD. Remember, that the SD is not changed when we add a constant to each element in a list. So we can compute the SD as follows:

$$\begin{aligned}SD(65-(X_1+\cdots+X_{100})) &= 65-SD(X_1+\cdots+X_{100})\\ &= \sqrt{100}SD(X_1)\\ &= \sqrt{100}\frac{1}{2}\\ &= 5\end{aligned}$$

The chance then would be the area under the normal curve from $N-55.5$ to $N-44.5$. In SU these points become -1.1 and 1.1. So the area under the curve, and thus the chance to get the Porsche, is 72.87%.

p371 # 1

	Number of heads		Percent of heads	
Number of tosses	Expected value	SE	Expected value	SE
100	50	5	50%	5%
2,500	1,250	25	50%	1%
10,000	5,000	50	50%	0.5%
1,000,000	500,000	500	50%	0.05%

p372 # 5

We are looking at the sum of 50 random guests. Let X_i denote the weight of the i-th guest. Then we are trying to estimate the sum $X_1 + \cdots + X_{50}$. We can compute the expected value and the SD as usual:

$$\mathbb{E}(X_1 + \cdots + X_{50}) = 50\mathbb{E}(X_1) = 50(150) = 7500$$

$$SD(X_1 + \cdots + X_{50}) = \sqrt{50}SD(X_1) = \sqrt{50}(35) \approx 247.5$$

Now 4 tons is 8000 pounds, so we are interested in the chance that this sum will be over 8000. In SU 8000 is $\frac{8000-7500}{247.5} \approx 2$, so we are looking for the area to the right of 2 in the normal distribution. This is about 2.5%. Sounds like the tower of terror :P

p383 # 3

Part (a)

The expected value in 100 draws is 1 red marble. ($\mathbb{E}(X_1 + \cdots + X_{100}) = 100\mathbb{E}(X_1) = 100\frac{1}{100}$). The SE is:

$$SE = SD(X_1 + \cdots + X_{100}) = \sqrt{100}\sqrt{\frac{1}{100}\frac{99}{100}} \approx 1$$

Part (b)

You can't draw fewer than 0 red marbles, so the chance is 0.

Part (c)

The area in the normal approximation that corresponds to drawing less than 0 red marbles is the area left of -0.5. In SU this is $\frac{-0.5-1}{1} = -1.5$. This is 6.68%. (The book gets its answer by using 0 instead of -0.5, but here it is important to pay attention to the endpoints, because it changes the answer so much since the SD and expected value are so small.)

Part (d)

No. There haven't been enough draws to make the normal curve look like the probability histogram. There are two factors that affect whether the normal curve approximates the probability histogram; how unbalanced the box is and how many draws have happened. If the box is really unbalanced like this one you need more draws for the probability histogram to be normal. In this case with 10000 draws it would be much closer.

p383 # 4

False. That is how you would regularly compute the confidence interval, but as is shown in the last problem, the normal curve does not approximate the probability histogram well here. So we can't compute the 95% confidence interval in this way :(

p391 # 2

Part (a)

Here you just compute the percent in the sample, which is also 47%. It is likely to be off by 1 SE, which is $\sqrt{\frac{(0.47)(0.53)}{500}} \approx 2.2\%$.

Part (b)

The previous sample is a simple random sample, so a 95% confidence interval can be computed as usual. We computed the SE in the previous step, so a 95% confidence interval is $47\% \pm 4.4\%$.

p392 # 4

In the households interviewed 379 had one or more cars. This is 75.8%. The SE is $\sqrt{\frac{(0.758)(0.242)}{500}} = 1.9\%$.

p393 # 13

In this box $\mathbb{P}(1) = 0.25$, $\mathbb{P}(2) = 0.5$ and $\mathbb{P}(5) = 0.25$. So the probability histogram is (ii). The probability histogram for the sum of 100 draws should look like a normal distribution according to the Central Limit Theorem, so it is (iii). (i) is irrelevant.

Worksheet 1

We know their winning proportion is currently 0.375. So the SE for this after 16 games is $\sqrt{\frac{(0.375)(0.625)}{16}} = 0.121$.

Part (a)

The 95% confidence interval is just the observed proportion $\pm\ 2SE$. This is 0.375 ± 0.242.

Part (b)

The 99.7% confidence interval is the observed proportion $\pm 3SE$, which is 0.375 ± 0.363.

Part (c)

Finally a 50% confidence interval is the observed proportion $\pm 0.68SE$, which is 0.375 ± 0.082.

Worksheet 2

This question is a tricky one. It is easiest if we think about it like question 4 from last week. Let X_i be a random variable that counts whether or not the Red Sox won the i-th game and similarly let Y_i count wins for the Yankees. Now let $Z_i = Y_i - X_i$. We are interested in how many games we need to play for a 95% confidence interval for the variable $\frac{Z_1+\cdots+Z_n}{n}$ to not include 0. This means that we want $0 = \mathbb{E}(\frac{Z_1+\cdots+Z_n}{n}) - 2SD(\frac{Z_1+\cdots+Z_n}{n})$. The first term can be rewritten:

$$\begin{aligned}\mathbb{E}(\frac{Z_1+\cdots+Z_n}{n}) &= \mathbb{E}(\frac{(Y_1-X_1)+\cdots+(Y_n-X_n)}{n}) \\ &= \mathbb{E}(\frac{Y_1+\cdots+Y_n}{n}) - \mathbb{E}(\frac{X_1+\cdots+X_n}{n}) \\ &= 0.500 - 0.375 \\ &= 0.125\end{aligned}$$

Unfortunately the SD is a foul beast. First we compute $SD(Z_1)$. Z_1 has outcomes -1, 0 and 1. -1 happens when the Yankees lose and the Red Sox win, which happens with probability $(0.5)(0.375) = 0.1875$. 0 happens when they both win or both lose, which happens with probability $(0.5)(0.375)+(0.5)(0.625) = 0.5$. Finally 1 happens when the Yankees win and the Sox lose, which happens with probability $(0.5)(0.625) = 0.3125$. So now we can compute the expected value for Z_1:

$$\mathbb{E}(Z_1) = -1(0.1875) + 0(0.5) + 1(0.3125) = 0.125$$

Note: there is a much easier way to compute this. $\mathbb{E}(Z_1) = \mathbb{E}(Y_1 - X_1) = \mathbb{E}(Y_1) - \mathbb{E}(X_1) = 0.5 - 0.375 = 0.125$.

Now we can compute the SD... ugh :(Z_1 has outcomes -1, 0 and 1, so $(Z_1 - \mathbb{E}(Z_1))^2$ has outcomes $(-1.125)^2 = 1.26563$, $(-0.125)^2 = 0.015625$ and $0.875^2 = 0.765625$. So $SD(Z_1) = \sqrt{\mathbb{E}((Z_1 - \mathbb{E}(Z_1))^2)} = \sqrt{1.26563(0.1875) - 0.015625(0.5) + 0.765625(0.3125)} = \sqrt{0.484375} = 0.695971$.

Ok, now we can solve the problem. We are interested in the differences of their winning percentages. That is given by the value $\frac{Z_1+\cdots+Z_n}{n}$. Now we compute the expected value and

SD of this:

$$\begin{aligned}
\mathbb{E}(\frac{Z_1+\cdots+Z_n}{n}) &= \frac{\mathbb{E}(Z_1+\ldots Z_n)}{n} \\
&= \frac{\mathbb{E}(Z_1)+\ldots\mathbb{E}(Z_n)}{n} \\
&= \frac{n\mathbb{E}(Z_1)}{n} \\
&= \mathbb{E}(Z_1) = 0.125
\end{aligned}$$

$$\begin{aligned}
SD(\frac{Z_1+\cdots+Z_n}{n}) &= \frac{SD(Z_1+\cdots+Z_n)}{n} \\
&= \frac{\sqrt{n}SD(Z_1)}{n} \\
&= \frac{SD(Z_1)}{\sqrt{n}} \\
&= \frac{0.695971}{\sqrt{n}}
\end{aligned}$$

So after n games the 95% confidence interval for the difference of the winning percentages is $0.125 \pm 2\frac{0.695971}{\sqrt{n}}$. We are interested in when the left endpoint of this interval is 0, i.e. $0 = 0.125 - \frac{1.39}{\sqrt{n}}$. We solve this for n and get $n = 124$. Whew!

Worksheet 3

If they keep winning at a .375 rate then their 95% confidence interval would be $.375 \pm 2\sqrt{\frac{(0.375)(0.625)}{n}}$. So we want to solve for when the right endpoint for this interval is 0.5. This happens when $n = 60$.

Worksheet 4

For the Red Sox to win 90 games they would need to win 84 out of the remaining 146 which is a proportion of 0.575. So if their true winning proportion is this or less I will lose. Right now we have a sample of 16 games. For this sample the observed winning proportion is 0.375 with a SE of 0.121. In SU 0.5 becomes $\frac{0.575-0.375}{0.121} = 1.65$. So we are looking for the area to the left of 1.65 on the normal distribution, which is about 95%. Thus we can say that the Red Sox's true winning percentage is less than 0.575 with 95% confidence.

For the bet this means that the expected value of the bet is $1000(0.05) - 2(0.95) = 48.1$. So I should take it!

Worksheet 5

After 80 games the Red Sox are still winning at a rate of 0.375. Now our SE is changed though; it is $\sqrt{\frac{(0.375)(0.625)}{80}} = 0.05$. To get a 99.99% confidence interval we need to take 3.9 SE's, which gives us the interval 0.375 ± 0.195.

Worksheet 6

First let's compute our confidence that the Red Sox winning percentage is greater than 0.500. This is the area to the right of 0.500 in SU, which is $\frac{0.500-0.375}{0.05} = 2.5$. The area is then 0.62%. (This means that in only 0.62% of samples is the true proportion 2.5 SEs greater than the sample proportion.)

Now we compute the same thing for the Yankees. First their winning percentage is $\frac{46}{80} = 0.575$. Now, their SE is $\sqrt{\frac{(0.575)(0.425)}{80}} = 0.06$. So in SU 0.5 is $\frac{0.500-0.575}{0.06} = -1.25$. So, in the Yankees case we are 10.6% confident that their true winning percentage is less than 0.500.

These two events are independent, so to find our confidence that both would occur at the same time we can just multiply our confidence in each event. So we are $100(0.0062)(0.106) = 0.07\%$ confident that the Yankees winning percentage is less than 50% and the Red Sox's is greater than 50%.