

Randomness Exercises

E1. Of the following, which appears to be the most indicative of the first 10 random flips of a fair coin?¹

- a) HTHTHTHTHT
- b) HTTTHTHTTT
- c) HHHHTTTTTT
- d) THTHTHTHTH

E2. Of the following, which appears to be the most indicative of the first 36 random rolls of a fair die?

- a) 123456123456123456123456123456123456
- b) 111111222222333333444444555555666666
- c) 166126623344455661234554321553232141
- d) 312415443445252624464226255211232365

E3. Refer back (or forward, as the case may be, if you haven't done it yet) to quiz 1. Assuming the mp3 player is set on "random", about how many songs would you have to hear until you're reasonably sure you've heard each song at least once?

E4. Assuming the mp3 player is set on random, about how many songs would you have to hear until you're reasonably sure you'll hear a duplicate song?

E5. The problems in **E3** and **E4** (as well as quiz 2) are classically known as a "coupon collector" type. Do some Googling and find out how to solve them without running the program I gave you. If you're the first one to submit a correct solution to me, I'll give you 5 extra credit points.

¹ I guess this is a good place to mention this:

All 4 of those results up there are equally likely. In fact, if you take MTH 243, you'll learn that each one of them has likelihood of 1 in 1024. However, if you actually saw the first, third, or 4th ones, you'd likely think something was up. Now, that might seem counterintuitive to the fact that they're all equally likely, but the fact of the matter is that the second one is different – it has somewhat asymmetric clumping, which you would expect in a truly random set of coin flips. The other ones seem to have a pattern underlying them – the antithesis of randomness.


E6. While you're at it, here's another: check out the "infinite monkey" theorem (sometimes also called the "law of infinite probability"). See how it applies to all of this?²

E7. And, while you're on a roll...do you think that the decimals in the expansion of the number π are random ($\pi \approx 3.14159265359\dots$, but the digits go on forever)? For those who enjoy poker, [this might interest you!](#)



Answers.

E1. You can rule out "a" and "d", because they show "too perfect" of the 50:50 ratio. That is, if your coin alternated back and forth perfectly between heads and tails, you'd most likely suspect something was up. In much the same way, "c" looks a little fishy to me. I'd say "b" has the right amount of minor clumping.

Now, I know what you're going to say..."What the hell! What do you mean 'right amount of clumping?'" Good for you! [Here's a link to an Excel spreadsheet](#) that simulates the experiment of "flipping a coin 10 times and keeping track of what happens." If you hold down the  button, you will see (rapidly!) many, many trials, and, in the cells to the right, the results. You'll then see three cells that describe the nature of three outcomes we're interested in:

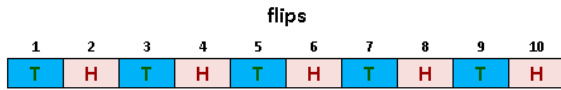
- "Alternating" means either HTHTHTHTHT or THTHTHTHTH – choices a) and d);
- "5 followed by 5" means either HHHHHTTTTT or TTTTTHHHHH – choice c) plus its mirror;
- "Clumpy Enough...but not too much" means the run has to have at least one least one clump of 3, but no clumps of 4 or more³.

² Here's a great meme I saw about it:



³ This is a *tad* bit arbitrary, but there's good mathematical basis in it (as you'll learn later in MTH 105, and also in MTH 243, if you take that journey). For now, think of it this way – in ten flips of a coin, there are more ways to get runs of 3 (roughly 25% of the time) than runs of 4 (only about 13% of the time). For example, you can have one run of 3, two runs of 3, or 3 runs of three. You can have either one or two runs of 4 – but they're each pretty rare, by comparison. However, a run of 4 is still more likely than either the "alternating" or "5 and 5" options!

Here's an "alternating" one:

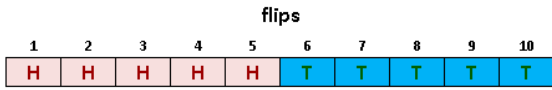


Alternating?
Yes!

5 followed by 5?
No.

"Clumpy Enough...but not too much."
No.

Here's a "5 followed by 5":

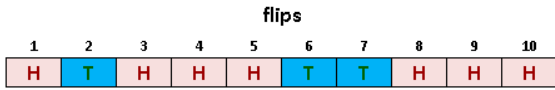


Alternating?
No.

5 followed by 5?
Yes!

"Clumpy Enough...but not too much."
No.


And here's a "clumpy enough"



Alternating?
No.

5 followed by 5?
No.

"Clumpy Enough...but not too much."
Yes!

I encourage you to explore the sheet by holding down  and watching the results. You'll then see why I answered the way that I did!⁴

E2. After reading the answer to **E1**, I'll bet you'll see why I picked "d". 😊

E3, E4: Isn't the internet *wonderful*?

.....

⁴ Formally, the chance of getting either an alternating series or a "5 and 5" is about 1 in 500. Later, you'll see that this means you'll get one, on average, every 500 series you try. Don't blink! You might miss it! Compare that with the roughly 75% of the time you get no more than (but including) a run of 3.

Quiz 1.

- (2 points)** I'm an old man – I still keep mp3's on my phone⁵. 😊 Suppose my mp3 player is loaded up with exactly 12 songs (call them song '1' through song "12"), and then I play exactly twelve songs on "random". Of the following 12-song playlists, which order would be the most indicative of true randomness? In case you're unsure of the difference between "randomness" and "shuffling", [take a look here](#)).

 - 1, 5, 4, 6, 9, 10, 2, 12, 11, 8, 7, 3
 - 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5
 - 3, 4, 2, 12, 3, 1, 10, 10, 11, 8, 3, 8
 - 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
- (3 points)** Why did you pick the way you did in #1?
- (2 points)** Now, suppose that your mp3 player was set to "shuffle" and not "random". Which of the above playlists would be more likely to be heard now?
- (3 points)** Why?


⁵ Don't even get me *started* about the LPs I keep at the house.

Quiz 2.

This quiz'll take us a little farther into randomness, and also how Excel helps us tame it. In class, we tossed coins and drew cards, and analyzed how long, on average, it took us for a few things to happen. This quiz will extend that out to another classic tool of the casino (and chance in general): a 6-sided die.

Go ahead and open up the spreadsheet that accompanies this quiz. In it, you'll see two tabs: one tab to model the experiment "Roll a die until you get the same number twice", and the other to model the experiment "Roll a die until you see all the sides." Just like in class, the first experiment is, on average, in fewer trials than the second – but how *many* fewer? That's (partially) what this quiz is all about. The other part is some basic graph reading⁶⁷.



This first thing you have to do is run the simulator (remember – use the  button) . Like, a whole lot. "What's 'a lot'?", you might ask. Well, until the graph stops "bouncing around" and settles into a consistent shape.

1. **(2 points)** Include screenshots of both graphs as your answer to number 1 here.
2. **(2 points)** Like we did in class, calculate the average number of rolls it takes to get a repeat.
3. **(2 points)** Repeat #2 for the average number of rolls it takes until you see all 6 sides.

Now, in both of those experiments, the graphs of the results are called *skewed*. In its simplest definition, "skewed" just means "asymmetric", and, as some of you will see in future stats classes, asymmetry can really mess with certain measures that you try to use. Let's look at the second experiment ("Roll a Die Until You See All The Sides") to see what I mean.

4. **(2 points)** Refer to your average from #3. Round that puppy off to the nearest whole number, and, using your data set, tell me what percentage (to the nearest whole percent) of the data from your experiment are less than or equal to that rounded average.
5. **(2 points)** What percent are larger?

Now, looking at those percentages, you realize that they're nowhere *near* $\frac{1}{2}$...which means that the median (which we've discussed in class) is pretty different than the average (not to mention the mode, which is way lower than either of the other two). And that's where future statistics classes will come in – you'll analyze data sets by "shape" to decide which measure(s) are OK to use without seeming misleading.

6. **(extra 2 points)** Google up the "Coupon Collector Problem" and tell me how you could have used a simpler (but much trickier) mathematical expression to get your answer from #3.

Quiz 3.

⁶ I used to do these experiments in class, too. But, as you'll see, they often require WAAAAAAAY huge numbers of trials, and my students got really bored rolling dice for, like, half an hour. 😊

⁷ Actually, the main point of the quiz is, like, 5 things. That's just how I roll. #holycrap #worstpunever



My son (and, um, me) loves to build with LEGOs. Once, when we went to the store, we saw the display at right. There are 16 different minifigures included in “series 11”. You buy them one at a time, and you can’t tell which one you get until you open it. Max wanted to collect all 16, of course...and then he grabbed 16 bags.

1. (2 points) What’s wrong with Max’s logic? That is, why did grabbing 16 bags most likely *not* satisfy what he was trying to do?

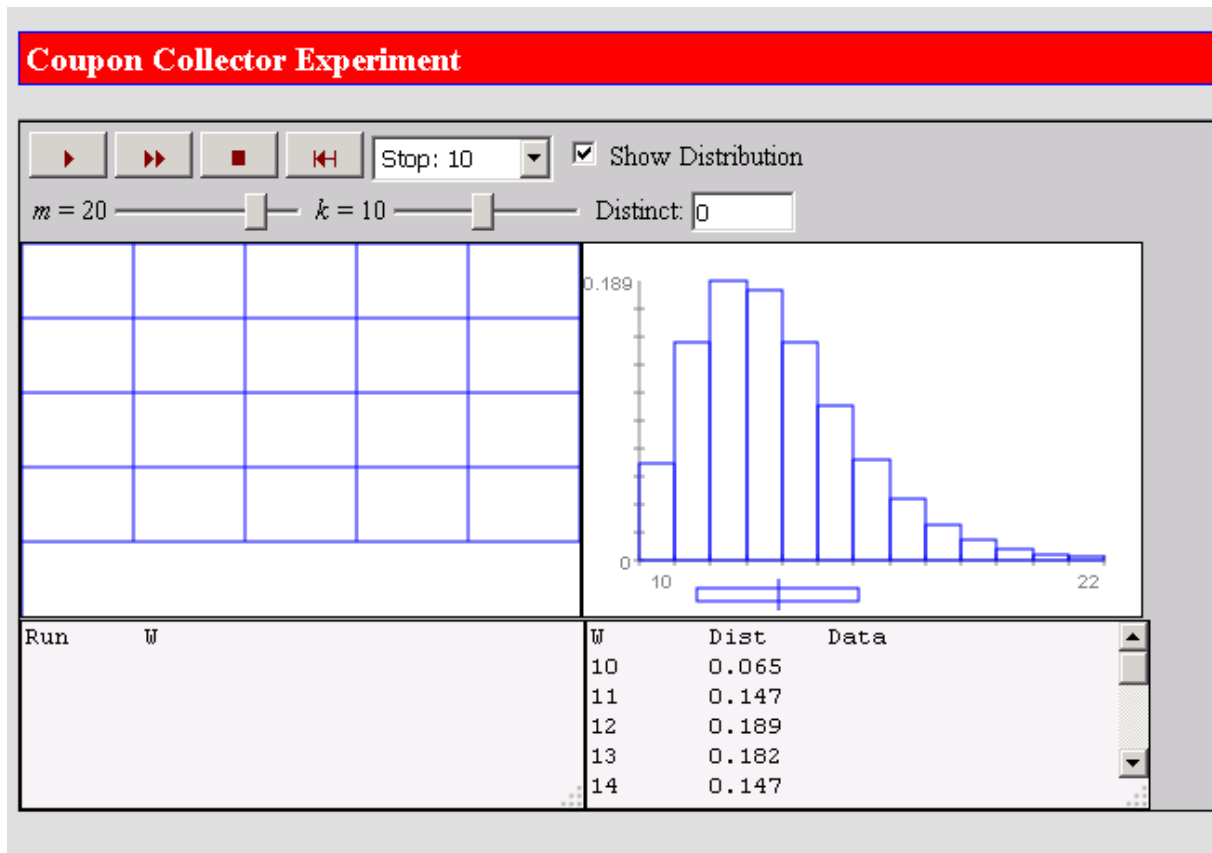
Now – assuming that the figures are uniformly distributed in the stores’ displays (a quick Google search can’t disprove this), we’re going to do the math problem Max and I pondered: how many bags of these minifigures, on average, would you have to buy before you were reasonably sure you had them all?

Like any good scientists, we’re going to do an experiment. Start by opening up this webpage:

<http://www.randomservices.org/random/apps/CouponCollectorExperiment.html>

(tech heads-up: don’t use IE or Edge!⁸) This page contains a simulator to help you explore this version of the Coupon Collector’s Problem (CCP). The CCP is relatively straightforward: given a certain number of distinct objects (“coupons”), how many would you have to collect, on average, until you’re sure you’ve seen them all? See how it’s our LEGO minifigure problem⁹?

OK – when you open that page, you see this:



⁸ EVER.

⁹ Or our “flip until you see both sides” experiment? Or our “deal cards until you see all suits” one? Or Quiz #2? ☺

Start by adjusting the page until it looks like this:

Run	W

W	Data
16	
17	
18	
19	
20	

If you get stuck here, email me!

Now, press the “play” button . When I do this, I get the following:

4	1	5	6
3	2	4	1
4	1	2	4
2	4	2	8

Run	W
1	53

W	Data
16	0.000
17	0.000
18	0.000
19	0.000
20	0.000

This screen tells me lots!

- See the bottom left rectangle that says “Run 1, W 53”? That tells me that I had to “collect” 53 different minifigures until I got them all! That means *lots* of repeated figures.
- See the rectangle cut up into 16 smaller rectangles, each with numbers inside of them? That’s the distribution of the 16 different minifigures. See how we got multiples of some? That’s to be expected!
- At bottom right is a “running total” of the outcomes of your experiment. If I scroll down, you’ll see the 53:

Run	W
50	0.000
51	0.000
52	0.000
53	1.000
54	0.000
55	0.000
56	0.000

Cool! Press the play button again and see how many minifigures it takes this time!


Run	W
1	53
2	89

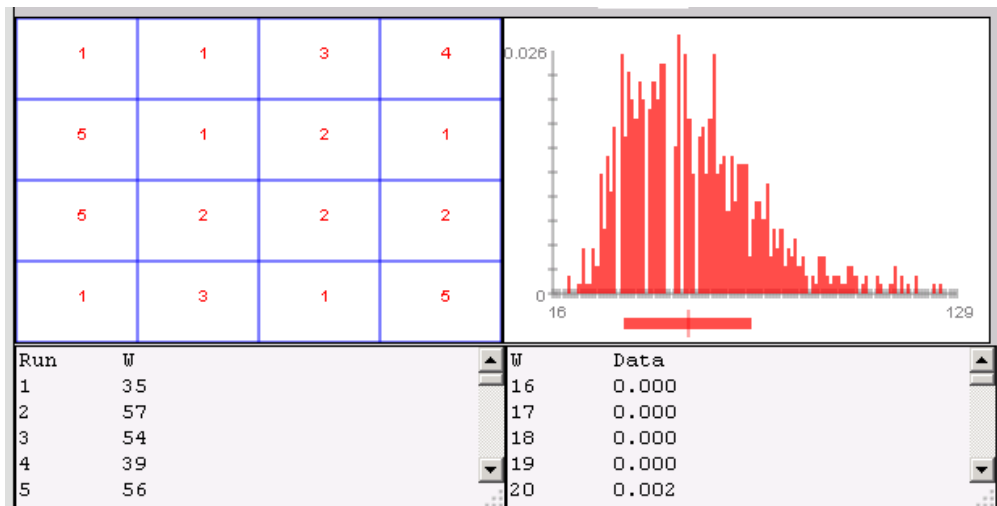
Whoa! 89?!?!? That's crazy! One more time before we get super serious:

Run	W
1	53
2	89
3	84

OK – so, if I asked you average those three values, you'd have no problem, right? You'd get $\frac{53 + 89 + 84}{3}$, which equals 75.3 (approximately), and you'd tell Max he needs to buy about 75 bags (at \$4 a pop!). But I wonder – how accurate is an average based only on these three values?

2. (2 points) Give one reason why these three values might not give a good representative average.

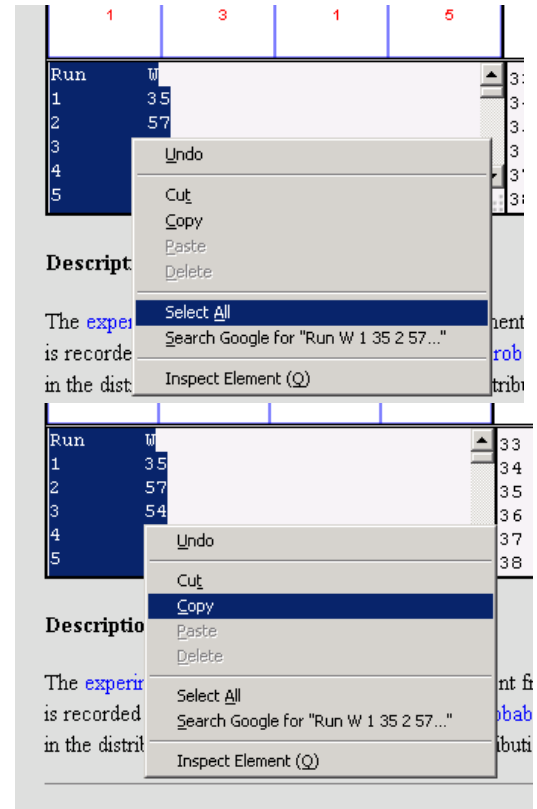
So, here's what I propose: let's do this experiment a bunch. How's 1000 times sound? Ready! Cool! Press the "Fast Forward" button . What you'll see is the experiment being done, over and over again, rapidly. The website is keeping track of the results and also graphing the results. After mine ran 1000 times, here's what I saw:



- (1 point)** Before we go any further, take a screen shot of your results and include it for your answer to this question!

At top left, you can see the 1000th experiment results. At top right, a graph of the results of all 1000 experiments (you might notice that the fewest number of minifigure bags needed would be 16. Technically, there's no maximum number (you could, for example, never get that pesky 16th bag. But, the applet tops out at 129. Safely large enough of a number, I guess, since the chance of having to go 129 bags to get all 16 is about 0.00026...FYI, all those "chances" are in the bottom right rectangle). Cool shaped graph, huh? You'll study them in some detail in MTH 243!

- OK! Let's do some math! Start by right – clicking in that bottom left rectangle. From the menu that pops up, choose "select all". All the data should highlight.
- Right click again in that rectangle while the data is highlighted, and select "copy". FYI: this is a GREAT way of grabbing text from webpages!
- Open up a blank Excel sheet. Left – click in cell A1 and press the "Ctrl" and "V" buttons at the same time. Voila! Your data has been pasted!



	A	B
1	Run	W
2	1	35
3	2	57
4	3	54
5	4	39
6	5	56
7	6	28
8	7	66
-	-	...

Now – we're gonna learn a few cool Excel commands. Start by answering the question we originally posed: on average, how many bags would we need to buy before we got all the minifigs?

Left – click in cell D3 and type the following formula:

=average(B2:B1001)

Don't forget the equals sign! Now press Enter!

4. (2 points) What's that average? You can round to the nearest whole number!

5. (2 points) Why is it so different from the one we got earlier (75)?

(You might be wondering why you can just use the "average" command here, but in the past few averages we've done in class, you couldn't. It's because here, the data in that bottom left window is given to you all listed out – I call this in "raw form". Averages in raw form are pretty simple – it's the **weighted** averages that you need to be careful with!)

Let's learn one last Excel command: something called "standard deviation". Please allow me to motivate why!

When I check the average I got from my 1000 experiments (I got an average of about 53) against how often that average actually **happened** over the course of my 1000 experiments (2.6% of the time...26 times out of 1000), I begin to realize that the exact average, but itself, is a little TOO precise! What we need is a "plus or minus" around that average¹⁰. And that's exactly what a standard deviation is! It's a measure of how far away, on average, the 1000 data points are from the average.

Type this formula in to cell D4:

=STDEV(B2:B1001)

6. (1 point) What do you get for your data's standard deviation? Nearest whole number's fine!

So, when I did it, I got a standard deviation of 18. We'll study both the average AND the standard deviation more in MTH 243, but, for now, you can think of them as this: I can expect a sizable percentage of my data to lie within 18 units up or down from 53 (that is, between 27 bags and 71 bags). I checked, and I found that 84% of my data lies between those two values!

So there's a quiz that primed you not only for some good statistical analysis, but also some web cutting/pasting and some more Excel tools. Fun!

¹⁰ We'll talk more about this in class soon!

Quiz 4.

In this quiz, we'll examine the wonderful number π . Remember that number from high school geometry class? It's the numerical value of the distance around any circle divided by the diameter of that circle. It's about equal to 3.14, but (as you might or might not remember) it's actually what's called an "irrational" number (not because it flies off the handle or anything; "irrational" just means it can't be written as an exact ratio of two integers).


Over time, folks have had a blast with computers, calculating more and more digits of π . Here are the first 100 digits:

3.14159265358979323846264338327950288419716939937510
58209749445923078164062862089986280348253421170679

Now, we've started studying randomness, and one of things we've discovered is that *truly* random numbers are impossible to predict. For example, if I asked you to guess at which digit came after that last "9" up there (no Googling!), and you truly guessed randomly,


1. **(1 point)** what percent of you would get that digit right?
2. **(1 point)** OK – *now* Google the next digit. What is it? 😊

What we'll do in the remainder of the quiz is analyze just *how* random π appears to be. And to do that, we'll attack this as we should – using statistics! Go ahead and open up the spreadsheet that accompanies this quiz. Left – click on the tab that's called "10 digits". There, you'll see the first ten digits of pi (vertically) in Column A. In column B, I had Excel generate 10 random digits, and place these right next to the digits of pi.

Now, press the  button. Each time you do this, you'll get a new random string of 10 digits in column B. Excel then compares these random digits to the digits in pi in the first column....if they match, Excel returns a "yes!" in column C; if not, that space stays blank. Make sense?

Now, as you do this, pay attention to the cell marked "Percentage of Matches". Next to that cell, Excel tells you what percentage of the time your random digit matches its fixed place in pi.

Bored yet? OK!

3. **(1 point)** Press the  button 10 times, and write down the 10 percentages you get. Then, average them (that is, add them up and divide by 10 – this is OK, since they're all equally weighted). How does the average you get compare to the percentage you got in #1?
4. Repeat what you just did in #3, but use
 - a. **(1 point)** the "100 digits" sheet.
 - b. **(1 point)** the "1000 digits" sheet.
5. **(2 points)** Do you notice that the values you got in 3 and 4 (not to mention 1!) are fairly close in value? What does that tell you?

Now, we'll test π 's randomness one last way¹¹...by exploring the number of "runs" in π . A "run" in statistics is defined to be a string of characters that are all the same. So, for example, the number 1111122222 has 2 runs. The number 11111222223 has 3 (even though there's only one "3", it's a run, by definition). So, in the first 10 digits of π , we see, actually, 10 different runs!


3.141592653

Since there are no identical digits adjacent to each other in the first ten digits, they're all individual runs.

Head over to the "100000 digit" sheet (if you're not already there). In this sheet, I've added two columns: column D tells you where the runs are in π , and column E tells you where the runs are in 100000 random digits.

All randomness has "clumpiness" – that is, where certain characteristics repeat. What we'll do now is analyze these clumped runs in π and see if...well, π is "clumpy" (or, "runny") enough.

Take a look at how many runs π actually has in it (cell G9) and then take a look at how many runs 100000 random digits *should* have (in cell G10). Now look at cell G11, where I have Excel calculate the percent error of those two numbers (percent error, if you've forgotten, is the "percent offness" the two numbers show. If the numbers are very close in value, the percent error should be small)

6. (1 point) What's the largest percent error you can find? Press  a few times to see it cycle.
7. (1 points) Is that error small enough to cause you to believe that pi is clumpy enough? Just yes or no for this one. 😊

¹¹ There are actually many more ways than these two to test randomness – take MTH 244 to find out how!