Method in Action Case Studies:

Title

Binomial tests and randomization approaches: the case of US presidential candidate height and election outcomes

Gert Stulp[a] & Louise Barrett[b]

[a] University of Groningen

Department of Sociology

Grote Rozenstraat 31

9712 TS Groningen

The Netherlands

Phone: 0031(0)50-363 6326

Fax: 0031(0)-50 363 6304

Email: g.stulp@rug.nl

[b] University of Lethbridge

Department of Psychology

4401 University Drive West

Alberta T1K 3M4

Canada

Contributer biographies

Dr. Gert Stulp

Gert Stulp is a behavioural biologist by training, and is currently investigating modern human behaviour from an evolutionary perspective. Having worked on personality in rats, lateralization in chickens, and intelligence in crows, for his PhD he turned to the role of human height in natural and sexual selection in contemporary populations. The type of research he does is wide ranging, from experimental work on both humans and non-human animals, along with naturalistic observational and questionnaire-based studies of human social behaviour. The diversity of his research is accompanied by a diversity in statistical techniques covered, in which he has substantial interest. His future plans are to investigate fertility decision by humans in contemporary western populations from an evolutionary point of view. Find out more about Gert here: https://sites.google.com/site/gertstulp/.

Prof. Louise Barrett

Louise Barrett studies the behavioural ecology and social cognition of human and non-human primates. She has a long-standing interest in the use of observational and statistical methods to test hypotheses in natural social settings, as opposed to controlled experimental environments. She has conducted fieldwork in the UK, Uganda and South Africa, on both human and non-human populations, with an emphasis on cooperation, reproductive decision-making, life history theory and parental investment. She now plans to extend this work to consider fertility decisions in contemporary western human societies. Find out more about Louise here:

https://dl.dropboxusercontent.com/u/3098948/B-HWEB/index.html.

Relevant disciplines

Psychology, sociology, political science

Academic level

Introductory Undergraduate

Methods used

quantitative research; Inferential statistics; Nonparametric statistics; Probability theory; Statistical packages: e.g. SPSS; Permutation; Probability; Statistical significance; Null hypothesis; Random values;

Keywords

probabilities; presidential candidate height; randomization tests; binomial tests; sample size

Link to the research output

Stulp, G., Buunk, A.P., Verhulst, S., & Pollet, T.V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. The Leadership Quarterly 24(1): 159-171. http://dx.doi.org/10.1016/j.leaqua.2012.09.002.

Abstract

Media reports often emphasize how the height of US presidential candidates plays an important role in election outcomes, with taller candidates winning an overwhelming number of elections across time. These claims are, however, rarely backed up by any form of reliable statistical analysis. Even when statistical methods are employed, they are often inadequate to the task. This case study discusses how to analyse Presidential candidate height in relation to election success, showing you how to calculate probabilities by hand, how to conduct a binomial test, and how to perform randomization tests.

Learning outcomes

By the end of the case you should:

1) Possess an increased sense of scepticism towards unsubstantiated claims made in the media;

2) Be able to carry out binomial tests by hand and by computer;

3) Understand the value of a randomization approach;

4) Better understand the importance of conducting statistical tests; particularly when sample size is small.

Height and Presidential election outcomes: a tall story?

We humans like to think of ourselves as complex, thoughtful creatures, capable of rational thought and well reasoned decision-making. At the same time, we're also fascinated by the suggestion that we're very easily swayed by (apparently) inconsequential factors when it comes to making certain kinds of decisions. Take voting patterns. As soon as the candidates for the US presidency are announced, political commentators delight in pointing out that, over the course of history, the taller of the two prospective candidates has won each election. At first glance, the evidence on which this claim is based seems to be quite strong. One study, for example, found that the taller candidate won all the elections between 1900 and 1968, while another, using data from 1952 to the year 2000, showed that the taller candidate won 10 of the 13 elections during this period. A more comprehensive study that used data from all US elections available (i.e., from 1789 and 2008), found that the taller man won 58% of the time. The weight of evidence, then, does seems to lie in favour of taller candidates winning.

But let’s look more closely. First, the number of elections included in these assessments varies, with some using a sample as small as 13 while others include all the elections that have ever taken place (56 in total). These studies also differ in the particular time period covered, and it’s often unclear why a particular set of years was chosen. The actual success of taller candidates also varies, from 58% to 100% of all elections selected, and it is apparent that when more elections included in the analysis, the success of the taller candidate drops. Finally, there is often no attempt to assess these patterns statistically (after all, 58% is only slightly more than half of all elections, so is this really any different from what one would expect if we simply ignored height, and tossed a coin to predict how often each candidate would win?). As you can see, once we consider the numbers in more detail, the results become rather murky; it is not at all clear that we should accept that height matters to election outcomes, or at least it is not clear that it matters all that much.

We should be concerned by this. After all, if the media are making large claims about the importance of height in Presidential elections, we need a way to satisfy ourselves that these claims have some factual basis, and aren't just plausible fictions or urban myths. How can we decide if we should take them seriously? In what follows, we explore the case of presidential height, showing how to accurately calculate the probability that the taller candidate wins more than one would expect by chance and how to perform the relevant statistical tests, in three different ways. Before we embark on this, however, we need to get some feel for probability distributions. More specifically, we need to have some idea of how to tell whether something occurs at random, or according to some underlying pattern.

Recognising randomness

Humans are extremely good at recognising patterns, but this ability can sometimes lead us astray. This is because we don't have a very good grasp of what randomness looks like. We tend to think of a random sequence of events as one where no pattern of any kind is ever apparent to us, but this isn't true. For example, while a lottery throwing up the numbers 6, 2, 4, 3, 1, 5 is what we expect to see if the lottery is fair and numbers are being selected at random, a lottery that produces a sequence of 1,2,3,4,5,6 seems rigged; this sequence doesn't look random at all (and of course, it isn't: the numbers occur in order of magnitude, which is obviously meaningful to any human being that can count). Yet, the latter sequence is just as likely to be *generated at random* as the first. We tend not to see this because we tend to lump together all other random sequences (6,4,5,2,1,3; 6,1,2,5,4,3; 3,5,6,2,1,4) and regard them as equivalent, even though they are, of course, just as different from each other as 1,2,3,4,5,6 is different from any other sequence. In other words, the perceived non-randomness of the ordered sequence of numbers obscures our ability to recognise that this sequence can be produced by a completely random process, and that it occurs just as frequently as any other unique sequence of six randomly generated numbers.

The point to focus on here is that any such sequence occurs only very rarely by chance alone. This means that, if a sequence tends to occur repeatedly, we have good reason to suppose something other than a random process is generating it. What we need, then, is some means of working out how often we can expect certain patterns to occur randomly, so we can decide whether we have a meaningful pattern or not. Basically, this is what statistical analysis does for us; it helps us overcome our natural bias to see patterns by providing us with the means to compare our observations to what we can expect by chance. So, when we read that around 70% of all US elections have been won by the taller candidate, we immediately see this as a pattern: it seems too much of coincidence to expect that the taller candidate would just happen to win such a large percentage of presidential races. But, as you should now recognise, we can’t simply trust our intuition here. We need to perform a more rigorous test to determine whether height influences election outcomes, or whether winning an election is really just a lottery with respect to height.

Winning by a head? What tossing coins can tell us about election success

To demonstrate how statistical analyses work, we’ll run through the analyses presented in a recent academic article, written by one of us (Stulp et al., 2013), that tested whether taller candidates are genuinely more successful in US elections. A reference to the article is given at the end of this piece and, if you look it up, you will be able to download all the data that was used in the paper, and reanalyse them for yourself, following the procedures we outline here. This means that, in what follows, we haven’t used the results of the most recent election in 2012, between Barack Obama and Mitt Romney, as this election wasn’t included in the original paper. So, to business.

The first thing to note is that most claims about the influence of height relate to more recent elections. This is mainly for reasons of convenience: heights of presidential candidates are more readily available for recent years. A more principled reason for focussing on more recent elections reflects the influence of television. From 1960 onwards, all presidential debates have been televised, allowing people to directly compare candidates against each other in a variety of ways, including, of course, their relative heights. The first televised debate took place between Richard Nixon and John F. Kennedy, with a further 12 elections taking place since (not including the most recent election in 2012). Of these elections, 67%, or exactly $^2/_3$, were won by the taller candidate (i.e., 8 out of 12, in the remaining thirteenth election, the candidates were of equal height, so this one was excluded). As we note above, this looks like more than just coincidence, but how do we go about testing this? One way to do so is to calculate the likely outcomes, which requires nothing more than some straightforward high school maths. Before we get into the details of these calculations, we need to take a small detour to introduce the concept of "permutations".

Imagine we flip a coin three times. How likely is it that we'll throw heads (H) twice, and tails (T) once? A specific sequence, let's say HHT, occurs with a probability of $(^1/_2)^3$= $^1/_2$*$^1/_2$*$^1/_2$= 0.125. But we are interested in any possible sequence, as long as there are two heads and one tail. So, we are also interested in cases of HTH and THH. Each of these sequences has the same chance of occurring, namely $(^1/_2)^3$. To work out the probability that any sequence containing two heads and a tail will occur in any order, we simply need to combine these probabilities, thus: 3 * $(^1/_2)^3$. These different combinations of sequences (HHT, HTH, THH) are known as permutations. We can calculate the number of possible permutations for any given number of trials using the following general formula: if there are N binomial (or Bernoulli) trials (that is, either an event occurs or it doesn't: head or tail, failure or success), and x of these trials are successes (so that, by definition, N-x are failures), then the number of permutations will be equal to $\frac{N!}{x!(N-x)!}$. Here, the exclamation mark indicates a specific mathematical procedure that we need to follow, known as the "factorial" function. The exclamation mark tells us that we need to multiply all the numbers in the sequence up to and including the number specified in front of the exclamation mark. So 3! (“3 factorial”) means that we have to calculate 3*2*1 = 6. Thus, if we toss a coin three times, and want to work out how many permutations include two heads and one tail (or two successes and one failure), we simply plug our numbers into the formula above, and this tells us what we can expect.

Now we can return to our question regarding the likely success of the 8 taller candidates in 12 elections. This is exactly equivalent to working out the probability that we’ll get 8 heads in a series of 12 coin flips. First, we calculate the number of possible permutations we can expect: $\frac{12!}{8!4!}$ = 495. Then, we multiply this number by the probability of throwing 8 heads, i.e., $(^1/_2)^8$ and the probability of throwing 4 tails, i.e., $(^1/_2)^4$. This gives $495*(^1/_2)^8*(^1/_2)^4 = 0.12085$. So, in any series of 12 throws, we can expect to see 8 heads occurring just over 12% of the time. To put this back in presidential terms, the likelihood that, in 8 out of 12 elections, the taller candidate can be expected to win is, purely by chance, more than twice as often as the conventional threshold of significance used in statistical analysis (which is set at a probability of 0.05 or 5%). In other words, in our series of 12 elections, we actually have no reason to believe that the taller candidate won by anything other than chance alone (at least with respect to height).

The value we have just calculated is, however, an underestimate of the true probability in which we are interested. This is because our hypothesis was that taller candidates would win more elections than shorter candidates, not that they would win precisely $^8/_{12}$. It is quite possible that the taller candidate could, theoretically speaking, have won more than 8 of the 12 elections, and we need to factor this into our probability calculations. That is, we also want to know the probability that the taller candidate could have won $^{9}/_{12}$, $^{10}/_{12}$, $^{11}/_{12}$, and $^{12}/_{12}$ elections. This means that we need to calculate the probability that the taller candidate will win *at least* 8 out of the 12 possible elections. Using our more formal notation, we can write these probabilities as P(x=8), P(x=9), P(x=10), P(x=11), and P(x=12). Using the formula given above, we can then calculate these probabilities, which we have done for you in the sixth column of Table 1. If we sum all these probabilities ($P(x\geq8)=0.12085+0.053711+0.016113+0.00293+0.000244$), we get a value of 0.193848. That is, the chance that the taller candidate would win an election purely by chance alone is now about 20% or one in five. The pattern relating to height that we discern in these wins is, therefore, just as misleading as our assessment of the non-random sequence of lottery numbers.

Calculating probabilities by hand is a fairly simple procedure, but scientists, just like everybody else, tend to look for easier ways of achieving the same results. In this case, statistical software (like the widely used R and SPSS) can calculate these binomial probabilities for us in a fraction of the time needed to do so by hand. This isn't just a matter of being too lazy to calculate for ourselves: imagine you now wish to test how likely it is that you'll see 423 wins or fewer across 1000 events -- that's a lot of calculations to do purely by hand. In these cases, it makes more sense to use statistical software, and perform a 'binomial test' on your data, which is the formal name given to the procedure we have just carried out (if you want to carry out a binomial test in SPSS (IBM SPSS Version 20), you simply use the drop-down menus to select the following options: analyze -> non-parametric statistics -> Legacy dialogs -> Binomial. To do the same in R, simply type in: “binom.test(8, 12, 0.5)”). Both SPSS and R will give you a two-sided p-value of 0.3876953125. A two-sided p-value is calculated for a non-directional hypothesis (in this case, height will affect election outcome in some way) while a one-sided p-value is calculated for a directional hypothesis (here, taller candidates will win more elections). If we halve the two-sided p-value, we obtain the one-sided p-value, 0.19384765625, which is identical to the value we calculated by hand (we did some rounding, but the value is identical). Of course, it’s good to know that the statistical output from these programmes is identical to our calculations by hand, showing both that we did it correctly and that the statistical software has implemented the binomial test correctly (a minor note of caution: in SPSS, when sample sizes are larger than 25, the package uses a short-cut (a “normal approximation” of the binomial distribution), rather than calculating exact binomial probabilities. In these situations, the output would be close to our calculations by hand, but not identical). As with our calculations by hand, the use of statistical software also demonstrates that 8 out of 12 wins for the taller of two candidates winning is not significantly different from chance.

Randomizing election outcomes

Although we’ve now shown quite convincingly that height isn’t a significant factor in deciding election outcomes, there is another useful method we can use to calculate probabilities. We’ll run through this with you now because, as you’ll see in the next section, we actually require this alternative method for certain kinds of analyses (the binomial test just won’t cut it in these cases). This third method involves what is known as a ‘randomization’ procedure. In essence, we let a computer ‘flip coins’ for us over and over again to produce a very large number of randomly generated sequences, and we can then look to see how often a particular sequence appears. This is the equivalent of calculating binomial probabilities. In other words, we can use the computer to simulate real-world events and see how often they can be expected to happen by chance. So, to use our presidential example, we can get the computer to randomly generate 12 numbers (representing the sequence of 12 elections), each with a value between 0 and 1. We decide that any number falling above 0.5 equals T or ‘taller candidate wins’ and every number falling below 0.5 equals S or ‘shorter candidate wins’. To show you what we mean, here is a sequence of randomly generated numbers obtained in just this way: 0.216, 0.521, 0.313, 0.698, 0.948, 0.685, 0.420, 0.776, 0.835, 0.492, 0.437, and 0.976. We can translate this into “candidate wins”, giving us the sequence: S, T, S, T, T, T, S, T, T, S, S, T. In other words, in this particular computer-generated run of random numbers, the taller candidate wins 7 of 12 “elections”.

Of course, this is only one sequence and, in itself, it is just as likely to occur as 12 shorter candidate wins or 12 taller candidate wins, or 6 wins each. That is, the probability that this sequence was produced works out as $(^1/_2)^{12}$, as we calculated previously. A single sequence therefore doesn’t tell us very much. What if we carry out another run of 12 numbers, and then another? What if we carry out a total of 10,000 computerised runs? This is easy work for a computer. In fact, it took our computer exactly 0.34 seconds to randomly generate a sequence of 12 numbers 10,000 times, and sum the values that were higher than 0.5 in each case. We stored these 10,000 summations as output, with values that range between 0 (all values in a given sequence fell below 0.5) to 12 (all values in a given sequence fell above 0.5). In Table 1 column 7, we have entered the results of all the 10,000 runs (that is, how often in 10,000 runs do we get $^{0}/_{12}$ taller candidate wins, $^{1}/_{12}$ taller candidate wins, $^{2}/_{12}$ taller candidate wins, and so on, across each run). As expected (given that we randomly determine, with a chance of $^{1}/_{2}$, whether the taller or a shorter candidate will win), the most frequently observed value was 6 (i.e., where there are 6 values were higher than 0.5 and 6 were lower): this occurred in 2246 runs out of 10,000, which we can express as a probability of $^{2246}/_{10000}$ = 0.2246. This is very similar to the probability calculated by hand. Remember that the chance of 6 wins out of 12 can be calculated as: $\frac{12!}{6!6!}*(^{1}/_{2})^{6}*(^{1}/_{2})^{6}$, which equals 0.225586. Another example: the chance that all values will fall above 0.5 (or only taller candidates will win the election) can be calculated as $(^{1}/_{2})^{12}$ (remember, twelve taller candidate wins in a row can occur only in one sequence, namely TTTTTTTTTTTT) = 0.000244, which is a very low probability. In our 10,000 randomized sequences, this happened only twice, i.e., with a probability of $^{2}/_{10000}$ = 0.0002 -- again almost identical! When we compare all the probabilities obtained through our randomization approach (column 8) to those obtained through the calculations by hand (column 6), we can see that they are all very similar. This is, of course, exactly what we would predict, and it gives us confidence that our computer indeed generated its numbers at random. Now, let us consider our research question again: how often do we find 8 out of 12 tall candidate wins in our randomized sequences? Exactly 1224 times out of 10,000, i.e., a probability of 0.1224. Again, this is very similar to our previously calculated value (0.12085), and it is clearly non-significant. As above, though, we are not interested in the probability of exactly 8 out of 12 wins, but the probability of *at least* 8 out of 12 wins (or, to put it another way, the chance of 8 or more wins). Using our table, we can calculate that 8 or more wins occurred exactly (1224+570+186+35+2) = 2017 times (i.e., we sum the probabilities of 8, 9, 10, 11 and 12 wins), which is a probability of 0.2017, or a chance of 1 in 5 that the taller candidate wins by chance, just as we calculated above. So, our randomization procedure also indicates that we have no reason to conclude that the taller candidate won 8 out of 12 elections by anything other than chance.

Does a bigger sample mean more taller winners?

This isn't the end of the story, however, because televised campaigns represent only a subsample of all possible US elections. Maybe height plays a role if we consider all the elections that have ever taken place. As we've noted, there have been 56 US elections in total since 1789. We immediately run into some problems, however, when we try to include data from all these elections. Most importantly, not all the data that we need are available. For the elections in 1804, 1808, 1816, and 1868, for example, we don't have information on the heights of all the candidates. This is an unavoidable problem, so we just have to make the best of things, and run our analyses on those elections for which we do have data. Another issue is that in 1789, 1792, and 1820, the candidates ran unopposed, which means their height was irrelevant; being 4'6" or 6'4" couldn't make any difference here, these men were going to be elected no matter what. Finally, in 1832, 1884, 1940, and 1992, there was no taller candidate because the opponents were approximately the same height. Excluding all these problematic years, we're left with 45 elections for analyses.

In these 45 elections, the taller candidate was elected president 26 times. So, in 58% of all elections for which data are available, the taller candidate prevailed. A binomial test reveals that a value of 58% is not significantly different from chance ($^1/_2$), with a one-tailed p-value of 0.187. There is, however, a problem using the binomial test in this way. To be blunt, what we have just done is completely wrong. The reason is simple: in five elections (namely those taking place in 1824, 1836, 1856, 1860, and 1912) there were more than two candidates (in 1824, for instance, there were four). The binomial test we've used assumes that the probability of winning (or losing) is equal to $^1/_2$, but, of course, this cannot be the case whenever there are more than two candidates involved. For instance, if there are four candidates, each candidate will only have a $^1/_4$ chance of winning, not $^1/_2$ (assuming the election outcomes are decided randomly, of course). To illustrate this, consider the 12 televised debates we used for the above analysis. If there had been four candidates running in each of these, and 8 out of 12 had been won by the tallest candidate, we would now have a statistically significant result (Binomial test, $p = 0.002782$, one tailed).

As you can see, the number of candidates running changes the probabilities of winning. How, then, do we go about testing our hypothesis, if we have to contend with a variable number of candidates across different elections? This is where our randomization approach comes in handy. Following the procedure above, we can generate 45 random numbers representing each election, but for those elections that had more than two candidates, we do not use $^1/_2$ as our cut-off value. Instead, we calculate a new value based on the number of candidates. So, when there are three candidates, we look at value above and below $^{1}/_{3}$ to determine the election outcome. This allows us to determine a frequency distribution of the elections that the tallest candidates would win by chance. We can then compare this distribution to the actual number of times the tallest candidate won, and determine the likelihood that this result could happen. Using this approach (see the paper for further details), the tallest candidate was predicted to win 26 times or more in 1142 of our 10,000 random samples. That is, our calculated (one-tailed) p-value of 0.1142 is not significantly different from chance at the conventional threshold of significance of 0.05. Once again, even when we use data from all possible elections, we find no evidence to suggest that height has any significant influence on election outcomes.

There is perhaps one final point to highlight. If we look again at the percentage of taller candidates that have won elections, we can see that, when all elections are considered, taller candidates win a lower percentage of elections (58%) than when we included only those elections since 1960 (67%), yet the p-value is *lower* for this smaller percentage and *higher* for the larger percentage. How can this be? It is because of the difference in sample size. To illustrate this point, imagine that 60% of coin flips result in heads. When we flip a coin only ten times, this would mean we got 6 heads and 4 tails (not a very surprising outcome when you think about it), and a binomial test would reveal a one-sided p-value of 0.378. In contrast, if we flip a coin 100 times to yield 60 heads and 40 tails, this would be a rather surprising finding intuitively and, indeed, a binomial test reveals a one-sided p-value of 0.0284. In other words, a single sequence of a small number of coin flips is very likely to throw up something that looks unusual, but actually doesn't deviate from chance. In a larger sample, with a greater variety of permutations (you can view 100 coin flips as 10 sequences of 10), a single sequence of 6 heads and 4 tails will no longer stand out, but if a certain sequence is repeated sufficiently often (for example, 10 sequences of 6 heads and 4 tails), it is clear that something other than chance is operating. This is why we should be particularly cautious when dealing with small sample sizes, and always ensure we perform a statistical test.

Conclusion

In this case study, we have shown that, contrary to popular belief, the tallest of two candidates doesn't always win the US presidential election. To do so, we used some very simple statistical techniques, namely the binomial test and randomization procedures. Interestingly enough, although claims on the relationship between height and election outcomes are unsubstantiated as we have seen, candidate height *does* matter when examining the number of votes received by each candidate (for further details see the paper below). We hope this example has convinced you that it is unwise to rely on our intuitions about certain patterns of events, and that a good knowledge of probability and statistics is helpful in preventing us being led astray by seemingly convincing numbers.

Exercises and Questions

1) In our worked examples, we excluded the most recent election of 2012. Find the heights of both 2012 presidential candidates, calculate the percentage of elections won by

the taller candidate, and calculate, by hand, the one-sided p-value to test the hypothesis that this percentage differs significantly from chance.

2) All the p-values we have calculated were one-sided. This is the p-value that is associated with the ‘directional hypothesis’ that the taller candidate is more likely to win. A less biased hypothesis would be to simply state that height has an influence on election outcomes, and to calculate a two-sided p-value. Why is this considered less biased?

3) Using the randomization results presented in Table 1, how would you go about calculating the two-sided p-value for the hypothesis that height matters in presidential elections?

4) Women, on average, are shorter than men. How would this affect our analysis if, in future elections, there were also female candidates running?

5) Why do you think height has been linked to election outcomes? Is it fair that candidates are judged on physical characteristics?

Read more

For a highly accessible and incredibly fun explanation of statistical methods that goes beyond the binomial tests presented here, try:

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics. Fourth Edition. Sage Publications Ltd. ISBN 978-1-4462-4917-8. http://www.sagepub.com/field4e/.

References

Stulp, G., Buunk, A.P., Verhulst, S., & Pollet, T.V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. The Leadership Quarterly 24(1): 159-171. http://dx.doi.org/10.1016/j.leaqua.2012.09.002.

Table 1. The probability of the taller candidate winning in x out 12 elections determined through a binomial test and a randomization approach.

Binomial test						10,000 randomizations[d]	
event	# tails	# heads	# permutations[a]	probability of sequence[b]	probability of event[c]	frequency[e]	p-value[f]
P(x=0)	0	12	$\frac{12!}{0!\,12!}=1$	$\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{12}$	0.000244	3	0.0003
P(x=1)	1	11	$\frac{12!}{1!\,11!}=12$	$\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{11}$	0.00293	29	0.0029
P(x=2)	2	10	$\frac{12!}{2!\,10!}=66$	$\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{10}$	0.016113	161	0.0161
P(x=3)	3	9	$\frac{12!}{3!\,9!}=220$	$\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{9}$	0.053711	534	0.0534
P(x=4)	4	8	$\frac{12!}{4!\,8!}=495$	$\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{8}$	0.12085	1200	0.1200
P(x=5)	5	7	$\frac{12!}{5!\,7!}=792$	$\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{7}$	0.193359	1911	0.1911
P(x=6)	6	6	$\frac{12!}{6!\,6!}=924$	$\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{6}$	0.225586	2246	0.2246
P(x=7)	7	5	$\frac{12!}{7!\,5!}=792$	$\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{5}$	0.193359	1899	0.1899
P(x=8)	8	4	$\frac{12!}{8!\,4!}=495$	$\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{4}$	0.12085	1224	0.1224
P(x=9)	9	3	$\frac{12!}{9!\,3!}=220$	$\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right)^{3}$	0.053711	570	0.0570
P(x=10)	10	2	$\frac{12!}{10!\,2!}=66$	$\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{2}$	0.016113	186	0.0186
P(x=11)	11	1	$\frac{12!}{11!\,1!}=12$	$\left(\frac{1}{2}\right)^{11}\left(\frac{1}{2}\right)^{1}$	0.00293	35	0.0035
P(x=12)	12	0	$\frac{12!}{12!\,0!}=1$	$\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^{0}$	0.000244	2	0.0002

[a] The number of permutations can be calculated using the formula $\frac{N!}{x!(N-x)!}$. *N* indicates sample size, *x* the number of "successes".

[b] The probability of the specific sequence is determined by $p^{x}(1-p)^{n-x}$, where *p* equals the chance on success (in this case the chance of 'heads', or $^{1}/_{2}$). Because the chance of success (heads) is equal to the chance of failure (tails), all probabilities of specific sequences are equal, namely $(^{1}/_{2})^{12}$.

[c] Probability is determined by $\frac{N!}{x!(N-x)!}\ p^{\ x}\ 1-p^{\ N-x}$. The summations of all probabilities in this column equal 1, given that we have listed all possible combinations and their probabilities.
[d] The computer produced 10.000 runs of 12 random numbers. For each run, we determined how many of the random numbers were larger than 0.5 (see text).
[e] The frequency of the event in the 10,000 runs. All values combined in this column add up to 10,000
[f] The p-value is calculated by dividing the frequency by 10,000