Detailed Example of a Test with a Zero-one Box

A random-number generator is alleged to produce the digits from 0 to 9 with equal frequency.

A test was run which produced 750 digits. Only 63 of them were 4's.

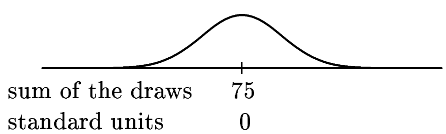
Does this support the assertion that this generator is random, or is there cause for concern based on the result?

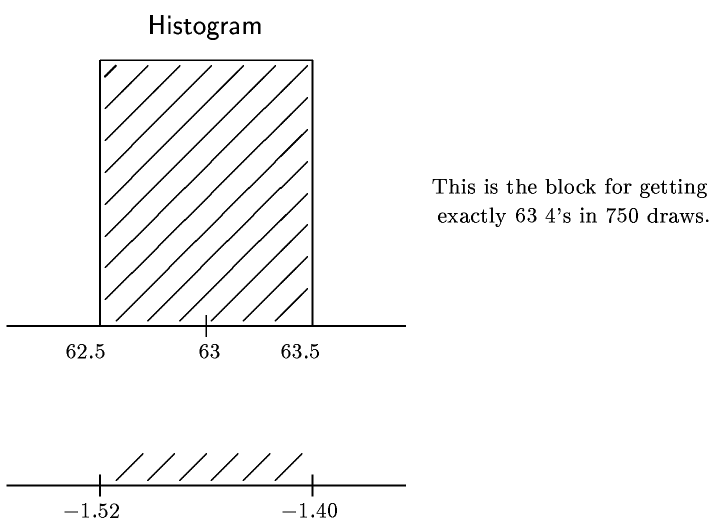
You must:

- State both the null and the alternate hypotheses.
- Set up a box model with the correct number of tickets.
- Find the average and SD of that box.
- State how the test statistic was derived from the draws from the box.
- Find the expected value and standard error of that test statistic.
- Draw the normal curve and fix its center.
- Draw an appropriate block from the probability histogram for the observed result and carefully find its endpoints. Some adjustment might be necessary.
- Using the sum of the draws scale, place that block on the graph of the normal curve.
- Explain why the height of the block will be approximated by the height of the normal curve at the position in question.
- Shade the tail area that applies to the observed result.
- Put into standard units the endpoint that forms the boundary of the tail area.
- Find the numerical value of the tail area. Use the normal table and your picture.
- Make the decision and use it to answer the question.

Answers follow on the next page.

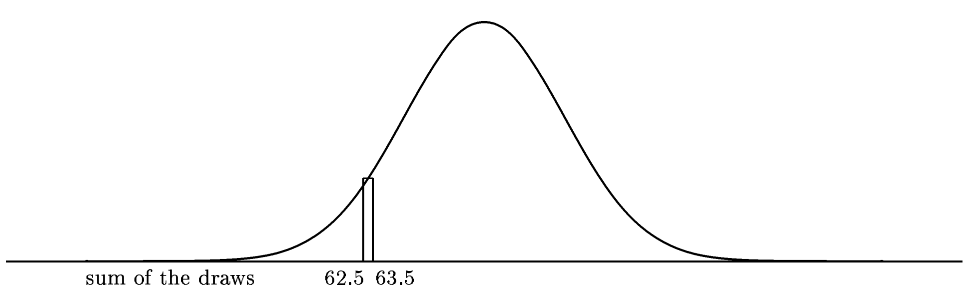
Answers.

- The null hypothesis is that the random-number generator was like 750 draws from a box of numbered tickets in which the fraction of tickets with a 4 is 1/10; the alternative hypothesis is that it was like drawing from such a box for which the fraction of 4's is more than 1/10.
- The box model has ten tickets: nine of them have a 0 on them, and one of them has a 1 on it.
- The average is 0.1; the SD is 0.3.
- It was the sum of the draws with replacement from that zero-one box.
- The expected value for the sum of the draws was $750 \times 0.1 = 75$; and the standard error for the sum of the draws was $\sqrt{750} \times 0.3 \approx 8.216$.
- 

- 

  The area of this block represents the binomial probability of getting exacly 63 4's when 750 draws are made from the box specified in our model.

  The continuity correction of $\pm 1/2$ is necessary. The numbers below show the endpoints in standard units, which will be found by proceeding with the next few steps.

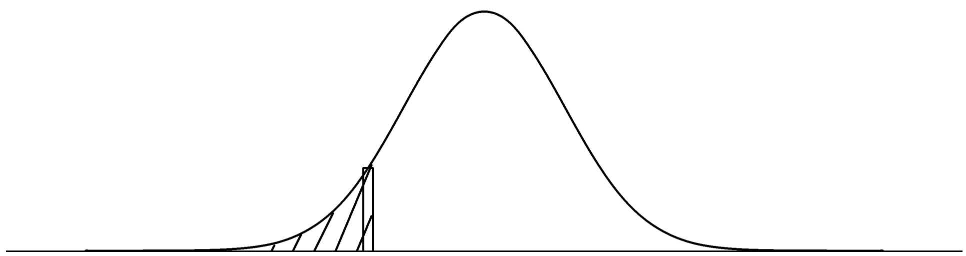
  In a test of hypothesis, the statistician is interested in finding the chance of getting a result of the observed number or more extreme, the tail area. In this example the tail is 63 4's or fewer.
- 

- The Central Limit Theorem applies.

  There is a box model for the random-number generator. The statistic, 63, is now represented as the sum of the draws from a box. The probability histogram will follow the normal curve, being that the number of draws is large. The next step will be to represent the endpoints in standard units, so that the standard normal curve, with its associated table, can be used.

  Note that the contents of the box do not follow the normal curve at all. For 0, the block is 10% in area, while for 1, the block is 90% in area.

  The blocks for the sum of the draws will follow the normal curve.

- 

- The standard units are only needed for the right endpoint of the block: 63.5

$$\frac{63.5 - 75}{8.216} = -1.40.$$

- The area is just about 8% (actually 8.075%).

  The area from -1.40 to $+1.40$ is about 84%. So the total area of the two tails is about 16%. Since the normal curve is symmetric, the two tails will ahve equal areas. One tail has about 8% area.

- It appears that the null hypothesis is true, since P is greter than 5%. An explanation of chance error is reasonable.

  We conclude that the random-number generator was fair.