Testing Whether a Coin is Fair

Suppose we flip a coin 10 times, and it lands tails 9 times. We suspect that the coin isn't fair. How strong is our evidence for this?

Let π be the probability that the coin lands tails. We wish to do a hypothesis test for

$$H_0: \ \pi = 1/2, \quad H_a: \ \pi \neq 1/2$$

What test statistic should we use? If T is the number of times the coin lands tails in n tosses, a reasonable choice of test statistic is

$$D \ = \ |T - n/2|$$

This is the absolute value of the difference between the number of tails we saw and the number we would expect to see if the coin is fair. Big values of D lead us to doubt H_0.

The P-Value for Our Test

In our example, $n = 10$, $T = 9$, and $D = 4$.
The P-value for our test with this data will be

$$\begin{aligned}
&P(D \geq 4) \\
&\quad = \ P(|T - n/2| \geq 4) \\
&\quad = \ P(T = 0 \text{ or } T = 1 \text{ or } T = 9 \text{ or } T = 10) \\
&\quad = \ P(T = 0) \ + \ P(T = 1) \ + \ P(T = 9) \ + \ P(T = 10) \\
&\quad = \ \left(\frac{1}{2}\right)^{10} \ + \ 10\,\frac{1}{2}\left(\frac{1}{2}\right)^{9} \ + \ 10\left(\frac{1}{2}\right)^{9}\frac{1}{2} \ + \ \left(\frac{1}{2}\right)^{10} \\
&\quad = \ 0.0215
\end{aligned}$$

Is this strong enough evidence to convince you that a coin you got in change from a vending machine is not fair?

Exact Sampling Distribution of Counts

The exact sampling distribution of the number of tails, T, in n coin flips is binomial(n,π), where π is the probability of a tail.

We can use this distribution to find exact P-values for tests of hypotheses such as $H_0: \ \pi = 1/2$.

We can do this equivalently in terms of either the *count* of the number of tails (out of n), or in terms of the *proportion* of tails:

$$p \ = \ T/n$$

The count has a binomial distribution; the proportion has a distribution that we can figure out from the binomial.

The Normal Approximation to the Sampling Distribution of p

When n is large, we saw earlier that the binomial(n,π) distribution is close being normal (especially if π isn't close to 0 or 1). The distribution of the sample proportion is also normal.

Specifically, when n is large, the sample proportion, p, has approximately a normal distribution with

$$\mu_p \ = \ \pi$$
$$\sigma_p \ = \ \sqrt{\pi(1-\pi)/n}$$

We can use this approximation to do hypothesis tests and find confidence intervals when n is large.

Hypothesis Test for a Proportion

To test $H_0: \pi = \pi_0$ versus $H_a: \pi \neq \pi_0$, we compute the statistic

$$z = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Under the null hypothesis, z will have close to the standard normal distribution, provided $n\pi_0$ and $n(1-\pi_0)$ are both fairly large (more than about 10).

First, let's see how well it works when n is *not* all that large. For the 9 tails out of 10 coin flips, a test of $H_0: \pi = 1/2$ gives

$$z = \frac{9/10 - 1/2}{\sqrt{(1/2)(1-1/2)/10}} = 2.53$$

This gives a two-sided P-value of 0.0114 (using Table B.2), compared to the exact P-value of 0.0215 that we found earlier.

Example: Girls and Boys on the Titanic

Recall that we looked earlier at data on passengers of the Titanic.

Among the $n = 109$ child passengers were 45 girls and 64 boys. Is the difference significant? What would it mean if it is? What is the population?

We will test the hypothesis that in the population from which these children are drawn girls and boys are equally likely. The proportion of boys in the sample is

$$p = 64/109 = 0.587$$

To test $H_0: \pi = 1/2$, we compute

$$z = \frac{0.587 - 1/2}{\sqrt{(1/2)(1-1/2)/109}} = 1.82$$

This gives a two-sided P-value of 0.069.

Testing the Population Median

Rather than test if the mean, μ, of the population has some given value, we sometimes may test whether the median, θ, has some given value, θ_0. That is, for a two-sided test, we test

$$H_0: \theta = \theta_0, \quad H_a: \theta \neq \theta_0$$

For instance, this may be appropriate when the distribution is highly skewed, or has heavy tails, since the median may then be a more meaningful measure of location. (But consider your purpose in looking at the data as well as the distribution.)

We can test whether the median is θ_0 by testing whether the proportion of the population greater than θ_0 is one-half.

Example: Bias of a Scale

To test if a scale has systematic error, as measured by the median, we weigh a 100kg weight 25 times, with results as follows:

```
 95 | 9
 96 |
 97 | 459
 98 | 89
 99 | 1349
100 | 037789
101 | 0789
102 | 378
103 |
104 |
105 | 8
106 | 1
```

Discarding the one observation of exactly 100, the proportion of the remainder that are greater than 100 is 14/24=0.5833. We'll use the normal approximation to test for the true proportion being 1/2:

$$z = \frac{0.5833 - 1/2}{\sqrt{(1/2)(1/2)/24}} = 0.82$$

The two-sided P-value is 0.41. We have no reason to think the scale has systematic error.