The Statistical Method

Will Perkins

February 24, 2013

What is statistics?

- A method for answering questions using data
- Qualitative and Quantitative answers
- Based on probability theory

A Simple Example

Question: Is this coin a fair coin?

How would you answer?

1. Flip the coin 100 times
2. Record the number of times it comes up heads.
3. What can you conclude?

Null Hypothesis

To answer any statisitcal question you need to compute a probability. To compute a probability you need a completely specificed probability space and probability measure.

In our example, this distribution should be 100 flips of a **fair** coin.

This probability distribution is called the *Null Hypothesis* and sometimes written H_0.

Alternate Hypothesis

The *Alternate Hypothesis* is the opposite of the null hypothesis and is denoted H_1.

There are two varieties of Null Hypotheses:

1. Two-sided Alternative: The coin is *not* fair. I.e. the coin is biased in favor of heads or in favor of tails.
2. One-sided Alternative: The coin is biased in favor of heads.

The choice of alternate hypothesis depends on the question you're trying to answer.

A Statistic

We also need a random variable on the probability space defined by the null hypothesis. In this case, the only sensible choice is $X =$ the number of heads.

In more complex situations there are many choices of a statistic, and you are free to choose your own. But to make the statistical method valid, you should choose your statistic before you collect your data.

Collecting Data

After you've specified a null hypothesis (probability space), alternate hypothesis, and test statistic (random variable), you collect your data.

You can think of this as taking a sample of your random variable X.

In our case, let's say we flip the coin 100 times and get 65 heads.

P-values

Under the distribution defined by the null hypothesis, our statistic (random variable) X has a fully specified distribution. We want to calculate the probability that under the null hypothesis, X takes a value as extreme or more extreme than the value under the observed data.

'Extreme' is defined in terms of the null hypothesis.

Say we had the one-sided null hypothsis: $H_1 =$ the coin is biased in favor of heads. Then we would calculate:

$$p = \Pr[X \geq 65]$$

P-values

If we had the two-sided alternative $H_1 =$ the coin is not fair, then we would calculate

$$p = \Pr[X \leq 35 \text{ OR } X \geq 65]$$

These probabilities are called **p-values**.

P-values

Here's a formal definition of a p-value. Let x be the value of the test statistic under the observed data.
One-sided alternative:

$$p = \Pr_{H_0}[X \geq x]$$

Two-sided altenative:
Trickier: often statisticians simply double the one-sided p-value (good for statistics with symmetric distributions). In the case of a statistic with a unimodal distribution, can also add the probabilities of all values of the test statistic less or equally likely than the observed.
Remember that the probabilities are always taken with respect to the probability measure defined by the null hypothesis.

How to Compute Probabilities

We need to calculate

$$\Pr[Bin(100, 1/2) \geq 65]$$

Three possible ways to do this:

1. Calculate it exactly (with a computer):

$$p = \sum_{j=65}^{100} \binom{100}{j} 2^{-100} = .0017588$$

2. Approximate the binomial by a normal distribution:

$$p \approx \Pr[Z \geq 3] = .0013499$$

3. Monte Carlo simulation: run 100,000 simulations of 100 coin flps with a random number generator.

$$p \approx$$

Error decays like $\frac{1}{\sqrt{T}}$ where T is the number of simulations.

Conclusions

Once we've computed a p-value, what does it mean?
The answer is somewhat vague, philosophical, and unsatisfying.

Some guidelines:

1. The lower the p-value, the less confidence we can have in the null hypothesis.
2. A very low p-value may make us reject the null hyopthesis as inconsisent with the data. (What is very low? .01? .00001?)
3. A high p-value, on the other hand, does not tell us that we can accept the null hypothesis.
4. The statistical method, like the scientific method, only gives negative results.
5. We can always report the *p*-value instead of simply 'reject' or 'not enough information'

(Discrete) Goodness of Fit

Here's a slightly more complicated statistical test. Say there are n categories, maybe favorite colors of people in America. We might want to ask: is the distribution of favorite colors uniform?
Let p_i be the probability a randomly chosen person prefers color i.
We can specify the null hypothesis that the distribution is uniform: $p_i = 1/n$ for all i.

We sample 100 people, ask their favorite color. How can we perform a statistical test?

(Discrete) Goodness of Fit

We need a statistic that measures how far the observed data is from uniform.
One choice:

$$X = \sum_{i=1}^{n} (E_i - O_i)^2$$

where E_i is the expected number of observations of category i (under the null hypothesis) and O_i is the observed number.

Now we need to know how to compute $\Pr[X \geq x]$.

(Discrete) Goodness of Fit

Again there are multiple ways to do this.

1. Approximate with Chi Square distribution.
2. Monte Carlo simulation.

Tests of Independence

Another example: we want to know whether two characteristics are independent of one another. Say, is gender independent of preference for Coke or Pepsi.
In this case our data will be a 2 by 2 contigency table, with 4 numbers: the number of men who prefer Coke, men who prefer Pepsi, women who prefer Coke, women who prefer Pepsi.
Say the above numbers are 34, 22, 47, 33.

Our null hypothesis would be that gender and preference are independent, and the alternate hypothesis is that they are not independent.

Tests of Independence

We can calculate the *expected* number of each of the 4 pairs of categories under the null hypothesis.

The expected number of men who prefer coke would be computed:

$$\frac{\text{Num. of Men} \cdot \text{Num. prefer Coke}}{\text{Tot. Num of Samples}}$$

And likewise for the other 3 pairs.

Tests of Independence

Again our statistic can be

$$X = \sum_{i=1}^{4} (E_i - O_i)^2$$

How to calculate p-value?