Darran McManus

September 19, 2014

There are several hypothesis testing methods used in statistics. These notes contain a short description of some tests and terminology, followed by examples for each case.

1 Terminology:

Z-test: Test of the mean. The null-hypothesis is a normal distribution with known mean and standard deviation.
T-test: Test of the mean. The null-hypothesis is a normal distribution with known mean, but unknown standard deviation.
F-test: Test of the variance. The null-hypothesis is a normal distribution with known mean, but unknown standard deviation.
χ^2 **- test:** Tests observed data against data expected from a given distribution.

One sample test: Null hypothesis involves a single distribution.
Two sample paired test: Null hypothesis involves a single distribution, however this is the distribution of the difference of two quantities (see examples).
Two sample unpaired test (pooled): Null hypothesis involves two normal distributions with equal standard deviations.
Two sample unpaired test (unpooled): Null hypothesis involves two normal distributions with unequal standard deviations.

2 Examples:

2.1 One-sample Z-test:

A claim is made that the broadband in a particular area is too slow. 100 homes in this area were tested. The average broadband speed for these homes was 4.9 Mbps. It is known that the national average is 5.6 Mbps, with a standard deviation of 0.8 Mbps. Is the claim supported?

The null hypothesis is a normal distribution with $\mu = 5.6$ and $\sigma = 0.8$, the distribution describing the whole nation. That is, the data for the 100 houses is not unusual under the national distribution.

To test this we use the Z-test statistic:

$$Z = \frac{\overline{X} - \mu}{(\sigma / \sqrt{n})} \tag{1}$$

Where $\overline{X}$ is the sample mean, n is the sample size and μ and σ are the mean and standard deviation respectively of the null hypothesis distribution. Provided n is large enough, then Z is known to follow a normal distribution.
In this case the value of Z is -0.875. To work out how likely this value is, we go to Z-score table and find out the area of the normal distribution for $|Z| > 0.875$. It turns out to be 0.62.
This means there is a 62% chance of obtaining a mean of 4.9 or greater under the null hypothesis. Hence we do not reject the null hypothesis.

2.2 One-sample T-test:

It is known that the average number of accidents nationwide in factories in 2012 is 120. 20 factories in a given area were selected and the number of accidents in each recorded. The average among these factories was 135, with a standard deviation of 15. Could one conclude that these factories are significantly more dangerous than the national average?

Firstly, we do not know the national standard deviation. These means the standard deviation of the null hypothesis is unknown. So we must use the T-test statistic:

$$T = \frac{\overline{X} - \mu}{(S / \sqrt{n})} \tag{2}$$

Where S is the standard deviation of the sample. The T-statistic is known to follow the student t-distribution with $n-1$ degrees of freedom, in our case this is 19 degrees of freedom. In this case the value of the T-statistic is 4.472. We look up the area of the t-distribution (with 19 degrees of freedom) for $|T| > 4.472$. For 19 degrees of freedom values with $|T| > 3.8$ have a likelihood of 0.999. Hence we can reject the null-hypothesis with 99.9% confidence.

2.3 Paired T-test:

A researcher wants to know if, in families with two children is the second child heavier in adulthood than the first.

Initially this may seem to require a null hypothesis involving two distributions, one for the first child and one for the second. However it in fact only requires a single distribution as the variable you are interested in is the difference between the weight of the first child (X) and the weight of the second (Y), that is $D = X - Y$. The null hypothesis would be that $X - Y$ would have a mean of 0. Since the variance of the null hypothesis is unknown, this would be a t-test. You use the formula:

$$T = \frac{\overline{D} - \mu}{(S_D / \sqrt{n})} \tag{3}$$

With the $\overline{D}$ the average difference in weight in your sample and S_D the standard deviation in the weight difference in your sample. After this, it is no different than the t-test above.

2.4 Pooled two sample T-test:

A group of researchers wants to know if the average running speed of athletes on Diet X is different from those on Diet Y. They take 31 athletes on Diet X and the sample mean is $\overline{X} = 9.2m/s$ and the sample variance $S_X = 0.4$. For Diet Y, it is $\overline{Y} = 9.4m/s$ with variance $S_Y = 0.2$, using 28 athletes. What can the researchers conclude?

Firstly the null hypothesis involves two distributions, the distribution of running times for Diet X and for Diet Y. Part of the null hypothesis in this case is that the means are equal. Although one can make different assumptions, perhaps that the means differ by some amount. However we assume that the normal distributions have the same standard deviation.
One uses a modified version of the t-statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{(S_p(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}))} \tag{4}$$

Where S_p is a quantity known as the pooled sample standard deviation.

$$S_p = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{(n+m-2)} \tag{5}$$

This t-statistic is known to follow the t-distribution with (n + m - 2) degrees of freedom. In our case this is 57 degrees of freedom.
In our case $S_p = 0.103$ and $T = -7.4$. For the t-distribution with 57 degrees of freedom values of $|T| > 3.5$ occur only 0.05% (area under graph is 0.00005) of the time. We can then reject the null hypothesis.

2.5 Unpooled two sample T-test or Welch's t-test:

We can take the previous example, except this time the null hypothesis does not assume that the two distributions have the same variance. In this case we use the t-test statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\frac{S_X^2}{\sqrt{n}} + \frac{S_Y^2}{\sqrt{m}}} \tag{6}$$

This is known to follow the t-distribution with $[r]$-degrees of freedom. Where $[r]$ is the integer portion of:

$$r = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(\frac{S_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_Y^2}{m}\right)^2}{m-1}} \tag{7}$$

In the example above $r = 45.05$ and so $[r] = 45$. Also, $T = -30$.

2.6 F-test:

Suppose that a company manufacturers fridges in two different factories. The average force the product can take is supposed to be 900 pounds. A sample of 20 fridges from one factory are taken and 25 from the other. The sample deviations are $S_X = 20$ and $S_Y = 31$. What is the likelihood that the two factories have the same variance in the force the fridges can take.

Here the null hypothesis is that $\sigma_X = \sigma_Y$, we are testing the variances. This is done with the F-statistic:

$$F = \frac{S_X^2}{S_Y^2} \tag{8}$$

In our case $F = 0.64$. This follows the F-distribution with degrees of freedom $(n-1, m-1)$ in our case $(19, 24)$. The value $F = 0.64$ is inside the 0.05 significance region.
So we do not reject the null hypothesis.

This test can be used to check whether you need Welch's test or not.

2.7 Non-Gaussian:

So far all the tests have assumed a Gaussian or normal distribution for the null hypothesis. It is possible for instance for a test to reject the hypothesis that two statistics come from the same normal distribution due to the fact that they come from the same non-normal distribution. The χ-squared test can be used to test what distribution one should use.

$$\chi^2 = \sum_{X_i} \frac{(O_{X_i} - < X_i >)^2}{< X_i >} \tag{9}$$

with O_{X_i} the values in your data for those random variables and $< X_i >$ the values expected under the distribution you are assuming. The value of χ^2 is then checked against the χ^2 - distribution with degree of freedom $k = n-1$, with n the sample size.

For example, suppose a gambler at a casino plays a game where you roll three dice and win âĆň10 for every six you roll. The gambler plays this game 200 times and obtains:

Rolls with no six: 94
Rolls with one six: 68
Rolls with two sixes: 31
Rolls with three sixes: 7

The expected values of each result are:
Rolls with no six: 116
Rolls with one six: 69
Rolls with two sixes: 14
Rolls with three sixes: 1

The χ^2 statistic then has the value: 60.8.

We then check this against the table for a χ^2-distribution with $k-1$ degrees of freedom, with k the number of random variables. In this, we check against the χ^2-distribution with 3 degrees of freedom. We see that χ^2 has a probability of 0.95 of being less than 7.815. Hence we may reject the null hypothesis.

Some tests are more robust to deviations from normal/Gaussian behaviour. The T-test is reasonably good, but the F-test is not.