MORE ON THE EQUAL-VARIANCE, TWO-SAMPLE T-TEST

Robustness

Recall:

- ***All models are wrong; some are useful***. (G.E. Box)
- The discussion on whether the model assumptions fit in the example about comparing two computer packages suggests.

They illustrate: We can't expect the assumptions of an inference procedure to apply exactly.

A procedure is said to be *robust* to departures from a model assumption if the results are still reasonably accurate when the assumption is relaxed to some degree.

Robustness may be determined by theory or by computer simulations.

Robustness of two-sample, equal-variance t-test:

- If samples are large enough, the Central Limit Theorem (*theory*) tells us that even if X and Y are not normally distributed, the distribution of $\bar{X} - \bar{Y}$ is approximately normal, so the test statistic will still have a distribution that is approximately t with m + n -2 degrees of freedom. *Computer simulations* have shown that moderate departures of X and Y from normality have little effect on the distribution of the t-statistic.

  Computer simulations:

- Simulations have also shown that this test is relatively robust to departures from the equal variance assumption, provided the two sample sizes are equal or nearly equal.

- However, lack of independence can cause serious problems -- the results of a t-test may be very misleading.

Another perspective on the two-sample, equal-variance t-test.

This test is equivalent to a certain F-test. The F-test can be generalized to situations where we are comparing more than two means and to some sampling methods other than simple random samples.

More detail on distributions:

A *t-distribution with k degrees of freedom* is defined as the distribution of a random variable of the form $\frac{Z}{\sqrt{U/k}}$, where

- Z~N(0,1)
- U~ χ^2(k) (Chi-squared with k degrees of freedom.)
- Z and U are independent.

A *chi-squared distribution with k degrees of freedom* is defined as the distribution of a random variable that is a sum of squares of k independent, standard normal random variables.

The proof that our test statistic T for the equal-variance, two-sample t-test has a t-distribution follows from these facts:

- T = $\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S^2}{m} + \frac{S^2}{n}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} \Bigg/ \sqrt{\frac{(m+n-2)S^2}{\sigma^2(m-n-2)}}$
  (algebra)

- Z = $\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}}$ is standard normal (seen earlier)

- U = $\frac{(m+n-2)S^2}{\sigma^2}$ is chi-squared with m + n -2 degrees of freedom. (Can be proved using model assumptions)

- U and Z are independent (Can be proved using model assumptions.)

An *F-distribution $F(\nu_1, \nu_2)$ with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator* is the distribution of a random variable of the form $\frac{W/\nu_1}{U/\nu_2}$, where

- $W \sim \chi^2(\nu_1)$
- $U \sim \chi^2(\nu_2)$, and
- U and W are independent.

If we have a t random variable of the form T = $\frac{Z}{\sqrt{U/k}}$,

where U and Z are as in the definition of t-distribution, then

$$T^2 = \frac{Z^2}{U/k}.$$

Now Z^2 is a chi-squared random variable with 1 degree of freedom, and U is chi-squared with k degrees of freedom, so T^2 is an F-distribution with 1 degree of freedom in the numerator and k degrees of freedom in the denominator. So we could do any t-test (with two-sided alternative) as an F-test, by using the square of the t-statistic.

To get some insight, assume equal sample sizes and look at the square of the t-statistic for the two-sample, equal-variance t-test:

Under the null hypothesis $\mu_X = \mu_Y$, the t-statistic is

$$T = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}},$$

Assuming m = n,

$$S^2 = \frac{(n-1)S_x^2 + (n-1)S_Y^2}{(n-1)+(n-1)} = \frac{S_x^2 + S_Y^2}{2}$$

and

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(S_x^2 + S_Y^2)}{2}}\sqrt{\frac{2}{n}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(S_x^2 + S_Y^2)}{n}}}.$$

Then our F statistic is

$$T^2 = \frac{(\bar{X} - \bar{Y})^2}{\left[\frac{(S_x^2 + S_Y^2)}{n}\right]},$$

which is equivalent to

$$\frac{\frac{n}{2}\left(\bar{X}-\bar{Y}\right)^2}{\frac{1}{2}\left(S_x^2+S_Y^2\right)}.$$

Since m = n, the denominator in this re-expression is just our pooled estimator of σ^2, the common variance of the two populations.

If the null hypothesis is true, then the two distributions (of X and Y) are the same -- so we may consider our two samples to be two samples of size n from the same $N(\mu,\sigma^2)$ distribution. So we can consider their means $\bar{X}$ and $\bar{Y}$ as samples from the sampling distribution of the mean of this common distribution.

Recall:

1. The sample means of samples of size n from an $N(\mu,\sigma^2)$ distribution have an $N(\mu,\sigma^2/n)$ distribution (the sampling distribution).
2. The sample variance of a distribution is an unbiased estimator of the population variance of that distribution.

Applying (1) and (2) to our sample $\bar{X}$, $\bar{Y}$ from the $N(\mu,\sigma^2/n)$ sampling distribution, we conclude that the random variable

$$S_b = \frac{\left(\bar{X}-\frac{\bar{X}+\bar{Y}}{2}\right)^2+\left(\bar{Y}-\frac{\bar{X}+\bar{Y}}{2}\right)^2}{2-1}$$

is an unbiased estimator of σ^2/n. (The b stands for "between sample.")

Using algebra,

$$S_b = \left(\frac{\bar{X}-\bar{Y}}{2}\right)^2+\left(\frac{\bar{X}-\bar{Y}}{2}\right)^2 = \frac{1}{2}\left(\bar{X}-\bar{Y}\right)^2.$$

Thus, if the null hypothesis is true, the numerator $\frac{n}{2}\left(\bar{X}-\bar{Y}\right)^2$ of T^2 is an unbiased estimator of σ^2, so we expect the quotient in T^2 to be close to 1. It can be proved that if the null hypothesis is false, then the ratio T^2 is *greater* than 1. So the F-test (equivalent to the t-test) can be interpreted as a test for the ratio of two estimates of σ^2.

This idea can be generalized to more than two samples: We form the sample variance for each sample, take the mean of these sample variances as one estimate of the common population variance σ^2, and compare with a "between sample" estimate of σ^2. With suitable modifications, this works, and is the idea behind the method of Analysis of Variance. However, we may, as above, multiply the numerator and denominator in the F-statistic by constants to make interpretations and/or formulas easier. In the notation used in the textbook, for the special case n = m considered here, we would express the F-statistic as

$$\frac{SST}{SSE/(2n-2)},$$

where SST (the *sum of squares for treatments* or *treatment sum of squares*) is

$$\text{SST} = \left(\bar{X} - \frac{\bar{X}+\bar{Y}}{2}\right)^2 + \left(\bar{Y} - \frac{\bar{X}+\bar{Y}}{2}\right)^2,$$

and SSE (the *sum of squares for error* or *error sum of squares*) is

$$\text{SSE} = \sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

($\frac{\bar{X}+\bar{Y}}{2}$ is sometimes called the *grand mean*, abbreviated GM.)