Hypothesis Testing

This note briefly describes two sample and one sample "t" tests.

Welch's Two Sample t Test

We can use Welch's "t" test to test the null hypothesis of no difference between the mean values for two distributions. The test statistic is the ratio of the difference of two independent sample means to the standard error of the difference:

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X{}^2}{m} + \frac{s_Y{}^2}{n}}}$$

... where ...

$\overline{X} = \frac{\sum_{i=1}^{m} X_i}{m}$ is the estimated mean of the first independent sample

$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$ is the estimated mean of the second independent sample

$s_X{}^2 = \frac{\sum_{i=1}^{m} \left(X - \overline{X}\right)^2}{m - 1}$ is the estimated variance of the first independent sample

$s_Y{}^2 = \frac{\sum_{i=1}^{n} \left(Y - \overline{Y}\right)^2}{n - 1}$ is the estimated variance of the second independent sample

$$Var\left(\overline{X}\right) = Var\left(\frac{\sum_{i=1}^{m} X_i}{m}\right) = \frac{1}{m^2} Var\left(\sum_{i=1}^{m} X_i\right) = \frac{1}{m^2} \sum_{i=1}^{m} Var\left(X_i\right) = \frac{m}{m^2} Var\left(X\right) = \frac{Var\left(X\right)}{m}$$

$$Var\left(\overline{X} - \overline{Y}\right) = Var\left(\overline{X}\right) + Var\left(\overline{Y}\right) - 2\, Cov\left(\overline{X}, \overline{Y}\right)$$

$Cov\left(\overline{X}, \overline{Y}\right) = 0$ for independent samples

This test statistic comes from a "t" distribution, which is parameterized by the degrees of freedom:

$$\nu = \frac{\left(\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X{}^2}{m} + \frac{s_Y{}^2}{n}}} \right)^2}{\frac{1}{m-1}\left(\frac{s_X{}^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{s_Y{}^2}{n}\right)^2}$$

If the probability of observing a "t" value at least as far away from zero as the observed test statistic is less than 0.05, then we can reject the null hypothesis (and claim there is a statistically significant difference).

The following describes how to derive the estimated degrees of freedom.

If $Z \sim N(0, 1)$ and $V \sim \chi_\nu{}^2$ are independent random variables, then $\frac{Z}{\sqrt{V/\nu}} \sim T_\nu$.

We can express the test statistic as ...

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X{}^2}{m} + \frac{s_Y{}^2}{n}}} = \frac{\left(\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X{}^2}{m} + \frac{\sigma_Y{}^2}{n}}} \right)}{\sqrt{\frac{\frac{s_X{}^2}{m} + \frac{s_Y{}^2}{n}}{\frac{\sigma_X{}^2}{m} + \frac{\sigma_Y{}^2}{n}}}}$$

... where ...

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X{}^2}{m} + \frac{\sigma_Y{}^2}{n}}}$$

... has a standard normal distribution and ...

$$\nu \frac{\left(\frac{s_X{}^2}{m} + \frac{s_Y{}^2}{n}\right)}{\left(\frac{\sigma_X{}^2}{m} + \frac{\sigma_Y{}^2}{n}\right)}$$

... has a chi-squared distribution with ν degrees of freedom.

$$Var\left[\nu\,\frac{\left(\frac{s_X{}^2}{m}+\frac{s_Y{}^2}{n}\right)}{\left(\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}\right)}\right]$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2 Var\left(\frac{s_X{}^2}{m}+\frac{s_Y{}^2}{n}\right)$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2\left[\frac{1}{m^2}\,Var\left(s_X{}^2\right)+\frac{1}{n^2}\,Var\left(s_Y{}^2\right)\right]$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2\left[\frac{1}{m^2}\,Var\left(\frac{(m-1)\,\sigma_X{}^2}{(m-1)\,\sigma_X{}^2}\,s_X{}^2\right)+\frac{1}{n^2}\,Var\left(\frac{(n-1)\,\sigma_Y{}^2}{(n-1)\,\sigma_Y{}^2}\,s_Y{}^2\right)\right]$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2\left[\frac{\sigma_X{}^4}{m^2\,(m-1)^2}\,Var\left(\frac{(m-1)}{\sigma_X{}^2}\,s_X{}^2\right)+\frac{\sigma_Y{}^4}{n^2\,(n-1)^2}\,Var\left(\frac{(n-1)}{\sigma_Y{}^2}\,s_Y{}^2\right)\right]$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2\left[\frac{\sigma_X{}^4}{m^2\,(m-1)^2}\,Var\left((m-1)\,\frac{s_X{}^2}{\sigma_X{}^2}\right)+\frac{\sigma_Y{}^4}{n^2\,(n-1)^2}\,Var\left((n-1)\,\frac{s_Y{}^2}{\sigma_Y{}^2}\right)\right]$$

$$=\nu^2\left(\frac{1}{\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}}\right)^2\left[\frac{\sigma_X{}^4}{m^2\,(m-1)^2}\,2\,(m-1)+\frac{\sigma_Y{}^4}{n^2\,(n-1)^2}\,2\,(n-1)\right]$$

$$=2\nu^2\left[\frac{\left(\frac{\sigma_X{}^4}{m^2\,(m-1)}+\frac{\sigma_Y{}^4}{n^2\,(n-1)}\right)}{\left(\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}\right)^2}\right]$$

$$=2\nu$$

... so ...

$$\nu=\frac{\left(\frac{\sigma_X{}^2}{m}+\frac{\sigma_Y{}^2}{n}\right)^2}{\left(\frac{1}{m-1}\left(\frac{\sigma_X{}^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{\sigma_Y{}^2}{n}\right)^2\right)}$$

Student's One Sample t Test

We can also use Student's "t" test to test the null hypothesis that the mean difference of two paired distributions is zero. The test statistic is the mean difference of the two paired (dependent) samples to the standard error of the mean difference:

$$t = \frac{\bar{\delta}}{\sqrt{\frac{{s_\delta}^2}{m}}}$$

... where ...

$\delta_i = X_i - Y_i$ is the difference of a pair of values (e.g. error for two models on the same test set)

$\bar{\delta} = \frac{\sum_{i=1}^{m} \delta_i}{m}$ is the estimated mean difference

${s_\delta}^2 = \frac{\sum_{i=1}^{m} \left(\delta_i - \bar{\delta}\right)^2}{m - 1}$ is the estimated variance of the mean difference

This test statistic comes from a "t" distribution, with "m-1" degrees of freedom.

If the probability of observing a "t" value at least as far away from zero as the observed test statistic is less than 0.05, then we can reject the null hypothesis (and claim there is a statistically significant difference between the two paired samples).

P.S. William Sealy Gosset published under the name "Student", as Guinness prohibited employees from publishing papers. He defined the "t" distribution.