Welch's t-test and the Welch-Satterthwaite equation

Welch's t-test is probably the most commonly used hypothesis test for testing whether two populations have the same mean. Welch's t-test is generally preferred over Student's two-sample t-test: while both assume that the population of the two groups are normal, Student's t-test assumes that the two populations have the same variance while Welch's t-test does not make any assumption on the variances.

Assume we have n_1 samples from group 1 and n_2 samples from group 2. For $j = 1, 2$, let $\overline{X}_j$ and s_j^2 denote the sample mean and sample variance of group j respectively. **Welch's t-statistic** is defined by

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Under the null hypothesis, t is approximately distributed as the t-distribution with degrees of freedom

$$\nu = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2 \bigg/ \left(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)} \right).$$

The equation above is known as the **Welch-Satterthwaite equation**.

Welch-Satterthwaite equation

Where does the Welch-Satterthwaite equation come from? The commonly cited reference for this is F. E. Satterthwaite's 1946 paper (Reference 1). Satterthwaite tackles a more general problem: let $s_1^2, \ldots, s_J^2$ be J independent sample variances each having r_j degrees of freedom. Consider the combined estimate

$$\hat{V}_s = a_1 s_1^2 + \cdots + a_J s_J^2,$$

where $a_1, \ldots, a_J$ are constants. (Back then, this was called a *complex estimate of variance*.) The exact distribution of $\hat{V}_s$ is too difficult to compute (no simple closed form? I'm not sure), so **we approximate it with a chi-squared distribution that has the same variance as $\hat{V}_s$**. The question is, **what is the degrees of freedom for this approximating chi-squared distribution?**

The paper states that the number of degrees of freedom is

$$r_s = \left(\sum_{j=1}^{J} a_j \mathbb{E}[s_j^2] \right)^2 \bigg/ \left(\sum_{j=1}^{J} \frac{(a_j \mathbb{E}[s_j^2])^2}{r_j} \right).$$

In practice we don't know the value of the expectations, so we replace them with the observed values:

$$\hat{r}_s = \left(\sum_{j=1}^{J} a_j s_j^2 \right)^2 \bigg/ \left(\sum_{j=1}^{J} \frac{(a_j s_j^2)^2}{r_j} \right).$$

The degrees of freedom for Welch's t-test is this formula above with $J = 2$, $r_j = n_j - 1$ and $a_j = 1/n_j$.

Heuristic argument

Satterthwaite's 1946 paper is the commonly cited reference, but that paper actually contains just the formulas and not the heuristic argument. For that, we have to go to his 1941 paper (Reference 2).

Consider first the simple case where we have just two independent sample variances s_1^2 and s_2^2, each with degrees of freedom r_1 and r_2. Let's compute the degrees of freedom for the chi-squared distribution that approximates $V = s_1^2 + s_2^2$.

For $j = 1, 2$, let $\sigma_j^2 = \mathbb{E}[s_j^2]$. From our set-up, we have $\frac{r_j s_j^2}{\sigma_j^2} \sim \chi_{r_j}^2$, thus

$$\begin{aligned} \mathrm{Var}(s_j^2) &= \left(\frac{\sigma_j^2}{r_j} \right)^2 \mathbb{E}[\chi_{r_j}^2] = \left(\frac{\sigma_j^2}{r_j} \right)^2 \cdot (2r_j) \\ &= \frac{2\sigma_j^4}{r_j} \quad \text{for } j = 1, 2, \\ \mathrm{Var}(V) &= 2 \left(\frac{\sigma_1^4}{r_1} + \frac{\sigma_2^4}{r_2} \right), \end{aligned}$$

where the last equality holds because s_1^2 and s_2^2 are independent. On the other hand, to approximate V be a chi-squared distribution with r degrees of freedom means that $\frac{rV}{\mathbb{E}[V]} \sim \chi_r^2$. Under this approximation,

$$\mathrm{Var}(V) = \left(\frac{\mathbb{E}[V]}{r} \right)^2 \cdot (2r) = \frac{2(\sigma_1^2 + \sigma_2^2)^2}{r}.$$

For this approximation to be good, we want the variance obtained under the chi-squared approximation to be the same as the true variance. Hence, we have

$$\frac{2(\sigma_1^2 + \sigma_2^2)^2}{r} = 2 \left(\frac{\sigma_1^4}{r_1} + \frac{\sigma_2^4}{r_2} \right),$$

$$r = (\sigma_1^2 + \sigma_2^2) \bigg/ \left(\frac{\sigma_1^4}{r_1} + \frac{\sigma_2^4}{r_2} \right).$$

The argument above is perfectly general: we can rerun it to get the effective degrees of freedom for $V = a_1 s_1^2 + \cdots + a_J s_J^2$.

References:

1. Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components.
2. Satterthwaite, F. E. (1941). Synthesis of variance.