F-tests vs. t-tests

BIO210 Biostatistics

Extra Reading Material for Lecture 32

Xi Chen

School of Life Sciences

Southern University of Science and Technology

Fall 2024

During the lecture, we talked about investigating if the means of more than two groups are different or not by analysing the source of the variation. We see that the total variation in our data comes from within each group and between different groups. If most variation comes from between groups, we say the means of those groups might be different.

This reasoning is very intuitive, and the same logic still holds when there are only two groups. Therefore, intuitively, we should be able to perform an $\boldsymbol{F}$-test even when there are only two groups. In general, we use a $\boldsymbol{t}$-test for two groups. Will we reach the same conclusion if we use an $\boldsymbol{F}$-test? Let's have a look at the test statistic when we perform a hypothesis testing. In this case, we have $k = 2$ groups:

$$F_{k-1, n_1+n_2-k} = F_{1, n_1+n_2-2} = \frac{\frac{\text{SSB}}{1}}{\frac{\text{SSW}}{n_1+n_2-2}} = \frac{\text{SSB}}{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Note the denominator $\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is basically the pooled variance (s_p^2) from a $\boldsymbol{t}$-test, so we have:

$$
\begin{aligned}
F_{1, n_1+n_2-2} &= \frac{\text{SSB}}{s_p^2} = \frac{n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_1(\bar{x}_1 - \bar{\bar{x}})^2}{s_p^2} \\[2em]
&= \frac{n_1\left[\bar{x}_1 - \dfrac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2}\right]^2 + n_2\left[\bar{x}_2 - \dfrac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2}\right]^2}{s_p^2} \\[2em]
&= \frac{n_1\left[\dfrac{n_2(\bar{x}_1 - \bar{x}_2)}{n_1 + n_2}\right]^2 + n_2\left[\dfrac{n_1(\bar{x}_1 - \bar{x}_2)}{n_1 + n_2}\right]^2}{s_p^2} \\[2em]
&= \frac{\left[\dfrac{n_1 n_2^2}{(n_1 + n_2)^2} + \dfrac{n_1^2 n_2}{(n_1 + n_2)^2}\right](\bar{x}_1 - \bar{x}_2)^2}{s_p^2} = \frac{\dfrac{n_1 n_2}{n_1 + n_2} \cdot (\bar{x}_1 - \bar{x}_2)^2}{s_p^2} \\[2em]
&= \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)} = t_{n-2}^2
\end{aligned}
$$

Now, we see that when we only have two groups, the $\boldsymbol{F}$ score is basically the $\boldsymbol{t}$ score squared. If we check the p-value using R, you will see that their p-value is the same, which is good, because we have consistency.