Hypothesis Testing: Tests Involving Variances

Lecture 7.1

1 Outline

1.1 Hypothesis Test: Ratio of Population Variances

Unlike the comparisons done in this chapter, comparing two population variances are usually based on the ratio of the variances (σ_1^2/σ_2^2).

> Note
>
> When $\sigma_1^2/\sigma_2^2 = 1$, the variances are equal. However, this ratio can not be negative, which is not consistent with the distributions we have considered (standard normal distribution and t-distribution)

1.2 F-distribution

The F-distribution is a distribution that has the following characteristics:

- Defined in the range $0 < \mathrm{X} < \infty$
- Defined by two degrees of freedom: the numerator degrees of freedom, df_{num} or df_1, and the denominator degrees of freedom, df_{den} or df_2.

In `R`, the following functions can be used to generate the properties of the *central* F-distribution.

- `df(x,df1,df2)`: PDF
- `pf(x,df1,df2)`: CDF
- `qf(p,df1,df2)`: Calculates F value for a given probability
- `rf(n,df1,df2)`: Generates n samples from the F-distribution

1.3 Hypothesis Test: Ratio of Population Variances

If we are interested in testing the null hypothesis $H_0 : \sigma_1/\sigma_2 = 1$, the alternative hypothesis can be one of the following:

- One-sided hypotheses: $H_a : \sigma_1/\sigma_2 < 1$ or $H_a : \sigma_1/\sigma_2 > 1$.
- Two-sided hypotheses: $H_a : \sigma_1/\sigma_2 \neq 1$

The problem dictates which alternative is appropriate.

1.4 Test Statistic

Assuming the following assumptions hold: s_1^2 and s_2^2 are computed from independent samples of size n_1 and n_2, respectively drawn from two normally distributed populations.

The quantity $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ follows the F-distribution with degrees of freedom $(df_1, df_2) = (n_1-1, n_2-1)$. Under the null hypothesis, $\sigma_1^2/\sigma_2^2 = 1$. Hence, our test statistic is expressed as:

$$F = S_1^2/S_2^2 \sim F(n_1 - 1, n_2 - 1)$$

1.5 Rejection Region and p-values

We will use the F-distribution to define the rejection region and the corresponding p-values.

1.5.1 Rejection Region

For one-sided tests, the rejection region is defined by the range:

$$F < F_{1-\alpha}(n_1 - 1, n_2 - 1)$$

As much as possible, we set the group with the expected larger variance to be in the numerator to avoid complications in calculating the rejection region.

For two-sided tests, the rejection region is defined by the ranges:

$$F < F_{\alpha/2}(n_1 - 1, n_2 - 1) \; or \; F > F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$$