MS04 Support Material

F distribution

For use with AQA A-level Mathematics Specification (6360)

1 Background

Consider two **independent normal** random variables, X_1 and X_2, such that:

X_1 has mean μ_1 and variance σ_1^2;
X_2 has mean μ_2 and variance σ_2^2.

Let $\bar{x}_1$ and s_1^2 denote **unbiased** estimates of μ_1 and σ_1^2 respectively, calculated from a random sample of size n_1.

Let $\bar{x}_2$ and s_2^2 denote **unbiased** estimates of μ_2 and σ_2^2 respectively, calculated from a random sample of size n_2.

Then:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{\nu_2}^{\nu_1}$$ where $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ are called the **degrees of freedom**

Note that sketches of the F distribution curve (see TABLE 7 in the blue AQA booklet of formulae and statistical tables) may assist candidates' understanding of the material on the following pages.

2 Tables of the F distribution

Tables (see TABLE 7 in the **blue** AQA booklet of formulae and statistical tables) of the F distribution only provide **upper percentage values:**

$$\text{P}(F < f(\text{U})) = p \quad \text{for sets of values of} \quad \nu_1 \quad \text{and} \quad \nu_2 \qquad ①$$

For example, from Table 7: $\text{P}(F < 4.70) = 0.95$ for $\nu_1 = 11$ and $\nu_2 = 5$.

As the shape of the F distribution is not symmetrical (see the χ^2 distribution), a method is required to find **lower percentage values**:

$$\text{P}(F < f(\text{L})) = 1 - p \quad \text{with degrees of freedom} \quad \nu_1 \quad \text{and} \quad \nu_2$$

This probability statement can be rearranged as:

$$\text{P}\left(\frac{1}{F} > \frac{1}{f(\text{L})}\right) = 1 - p \quad \text{with degrees of freedom} \quad \nu_1 \quad \text{and} \quad \nu_2$$

However, from the previous 'definition' of the F distribution: $\dfrac{1}{F_{\nu_2}^{\nu_1}} = F_{\nu_1}^{\nu_2}$

Thus $$\text{P}\left(F > \frac{1}{f(\text{L})}\right) = 1 - p \quad \text{with degrees of freedom} \quad \nu_2 \quad \text{and} \quad \nu_1$$

or $$\text{P}\left(F < \frac{1}{f(\text{L})}\right) = p \quad \text{with degrees of freedom} \quad \nu_2 \quad \text{and} \quad \nu_1 \qquad ②$$

Comparing ① and ② it can be seen that $\dfrac{1}{f(\text{L})}$ is the upper percentage value with the degrees of freedom interchanged or, in other words:

The **lower** percentage value for $F_{\nu_2}^{\nu_1}$ is the **reciprocal** of the corresponding **upper** percentage value for $F_{\nu_1}^{\nu_2}$.

For example, from TABLE 7, the **lower** 0.025 (2.5%) value for $\nu_1 = 7$ and $\nu_2 = 10$ is:

$$\frac{1}{4.761} = 0.210$$

3 Confidence limits for the ratio of two independent normal population variances

Consider the following:

$$P(f(L) < F < f(U)) = 2p - 1$$

Hence, substituting for F gives:

$$P\left(f_{\nu_2}^{\nu_1}(L) < \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} < f_{\nu_2}^{\nu_1}(U) \right) = 2p - 1$$

Rearranging gives:

$$P\left(\frac{s_1^2 / s_2^2}{f_{\nu_2}^{\nu_1}(U)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2 / s_2^2}{f_{\nu_2}^{\nu_1}(L)} \right) = 2p - 1$$

However, using the result from 2 , this becomes:

$$P\left(\frac{s_1^2 / s_2^2}{f_{\nu_2}^{\nu_1}(U)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \times f_{\nu_1}^{\nu_2}(U) \right) = 2p - 1$$

Thus, for example, a 95% confidence interval for $\dfrac{\sigma_1^2}{\sigma_2^2}$ is given by:

$$\left(\frac{s_1^2 / s_2^2}{f_{\nu_2}^{\nu_1}(U)},\ \frac{s_1^2}{s_2^2} \times f_{\nu_1}^{\nu_2}(U) \right) \quad \text{with} \quad p = 0.975$$

Note that if $n_1 = n_2$, then the two f-values are the same.

{The CI can be remembered as:

(ratio of sample variances) ÷ ('correct' f-value), (ratio of sample variances) × ('incorrect' f-value)}

Example 3.1

A random sample of size 10 from a normal population gives an unbiased estimate of 16.7 for the population's variance. A random sample of size 12 from an independent normal population gives an unbiased estimate of 33.7 for the population's variance.

Calculate a 98% confidence interval for the ratio of the two population standard deviations, giving the limits to two decimal places. Comment on the interval obtained.

Here $n_1 = 10$ with $s_1^2 = 16.7$ and $n_2 = 12$ with $s_2^2 = 33.7$

Thus $\nu_1 = 10 - 1 = 9$ and $\nu_2 = 12 - 1 = 11$

Also $2p - 1 = 0.98$ so $p = 0.99$

The two f-values are thus: 4.632 ('correct') and 5.178 ('incorrect')

The 98% confidence interval for the ratio of the population variances is thus:

$$\left(\frac{16.7 / 33.7}{4.632},\ \frac{16.7}{33.7} \times 5.178 \right) \quad \text{or} \quad (0.1070,\ 2.5660)$$

The 98% confidence interval for the ratio of the population standard deviations is thus:

$$(0.33,\ 1.60)$$

Since the interval(s) include(s) unity, there is no evidence, at the 2% significance level, of a difference in the two population variances.

[Since the latter interval includes 1.5, there is evidence, at the 2% significance level, that one of the population standard deviations is 1½ times the other.]

4 Hypothesis tests for the ratio of two independent normal population variances

Tests of:

H_0: $\sigma_1^2 = \sigma_2^2$ against H_1: $\sigma_1^2 \neq \sigma_2^2$ or H_1: $\sigma_1^2 > \sigma_2^2$ or H_1: $\sigma_1^2 < \sigma_2^2$ are based on:

$$F = \frac{s_1^2}{s_2^2} \sim F_{n_2-1}^{n_1-1} \quad \text{if } H_0 \text{ is true}$$

Similarly, tests of:

H_0: $\sigma_1^2 = a\sigma_2^2$ against H_1: $\sigma_1^2 \neq a\sigma_2^2$ or H_1: $\sigma_1^2 > a\sigma_2^2$ or H_1: $\sigma_1^2 < a\sigma_2^2$ where

$a > 0$ is a known constant are based on:

$$F = \frac{s_1^2}{as_2^2} \sim F_{n_2-1}^{n_1-1} \quad \text{if } H_0 \text{ is true}$$

Note that, because tables only provide **upper** percentage points directly, it is normal convention to always ensure that the **larger** of s_1^2 and s_2^2 (or of s_1^2 and as_2^2) is in the **numerator** of the F-statistic (so that the latter is greater than unity).

Example 4.1

Prior to a machine's overhaul, a random sample of 10 items from the machine' output had diameters that had a standard deviation of 13.6 mm. Following the machine's overhaul, a random sample of 15 items from the machine had diameters that had a standard deviation of 6.2 mm.

Stating any necessary assumptions, investigate, at the 1% level of significance, the claim that the overhaul has reduced the variability in item diameters.

Assumption is that diameters are normally distributed.

Here, since a reduction is to be investigated: 1 ≡ Before and 2 ≡ After

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 > \sigma_2^2$ (1-tailed)

SL $\alpha = 0.01$

DF $\nu_1 = 10 - 1 = 9$ $\nu_2 = 15 - 1 = 14$

CV $F = 4.030$

$$F = \frac{13.6^2}{6.2^2} = 4.81$$

Since $F >$ CV $\Rightarrow$ Reject H_0

There is evidence, at the 1% level of significance, that the overhaul has resulted in a reduction in the variability in item diameters.

Example 4.2

The standard deviation of a random sample of 12 of John's journey times from home to work is 4.23 minutes. The standard deviation of a random sample of 16 of John's journey times from work to home is 8.36 minutes.

Assuming journey times are normally distributed, test, at the 5% level of significance, the claim that the variance of John's journey times from work to home is twice that of his journey times from home to work.

Here, since $2 \times 4.23^2 = 35.7858$ and $8.36^2 = 69.8896$:

$1 \equiv$ work to home and $2 \equiv$ home to work

H_0: $\sigma_1^2 = 2\sigma_2^2$
H_1: $\sigma_1^2 \neq 2\sigma_2^2$ (2-tailed)

SL $\alpha = 0.05$
DF $\nu_1 = 16 - 1 = 15$ $\nu_2 = 12 - 1 = 11$
CV $F = 3.330$

$$F = \frac{8.36^2}{2 \times 4.23^2} = 1.95$$

Since $F <$ CV $\Rightarrow$ Accept H_0

There is no evidence, at the 5% level of significance, to reject the claim that that the variance of John's journey times from work to home is twice that of his journey times from home to work.

[Note that it is equally valid, though more prone to errors by candidates, to compare:

$$1.95^{-1} = 0.513 \quad \text{with} \quad \frac{1}{F_{15}^{11}} = \frac{1}{3.008} = 0.332$$

and then come to the same conclusion, since here $F >$ CV.]

5 Exemplar examination questions

These can be found on the following past papers:

MAS3 (CIs and HTs)

MBS7 (HTs only)

6 Special cases of the F distribution *(Not Examined)*

It is possible, though not at all easy, to prove the following results from the pdf of the F distribution. However the results can be illustrated by simply using percentage points from tables in the blue AQA booklet of formulae and statistical tables.

1 $F_{\infty}^{\nu} = \dfrac{\chi_{\nu}^{2}}{\nu}$

$F_{\infty}^{10}(0.95) = 1.831$ $\dfrac{\chi_{10}^{2}}{10} = \dfrac{18.307}{10} \approx 1.831$

2 $F_{\nu}^{1} = t_{\nu}^{2}$

$F_{10}^{1}(0.95) = 4.965$ $t_{10}^{2}(0.975) = 2.228^{2} \approx 4.964$

3 $F_{\infty}^{1} = z^{2}$

$F_{\infty}^{1}(0.95) = 3.841$ $z^{2}(0.975) = 1.96^{2} \approx 3.842$

4 Other results, perhaps better known, follow from these.

$t_{\infty} = z$ $\chi_{1}^{2} = z^{2}$