Suppose that X_1 and X_2 are independent random variables with

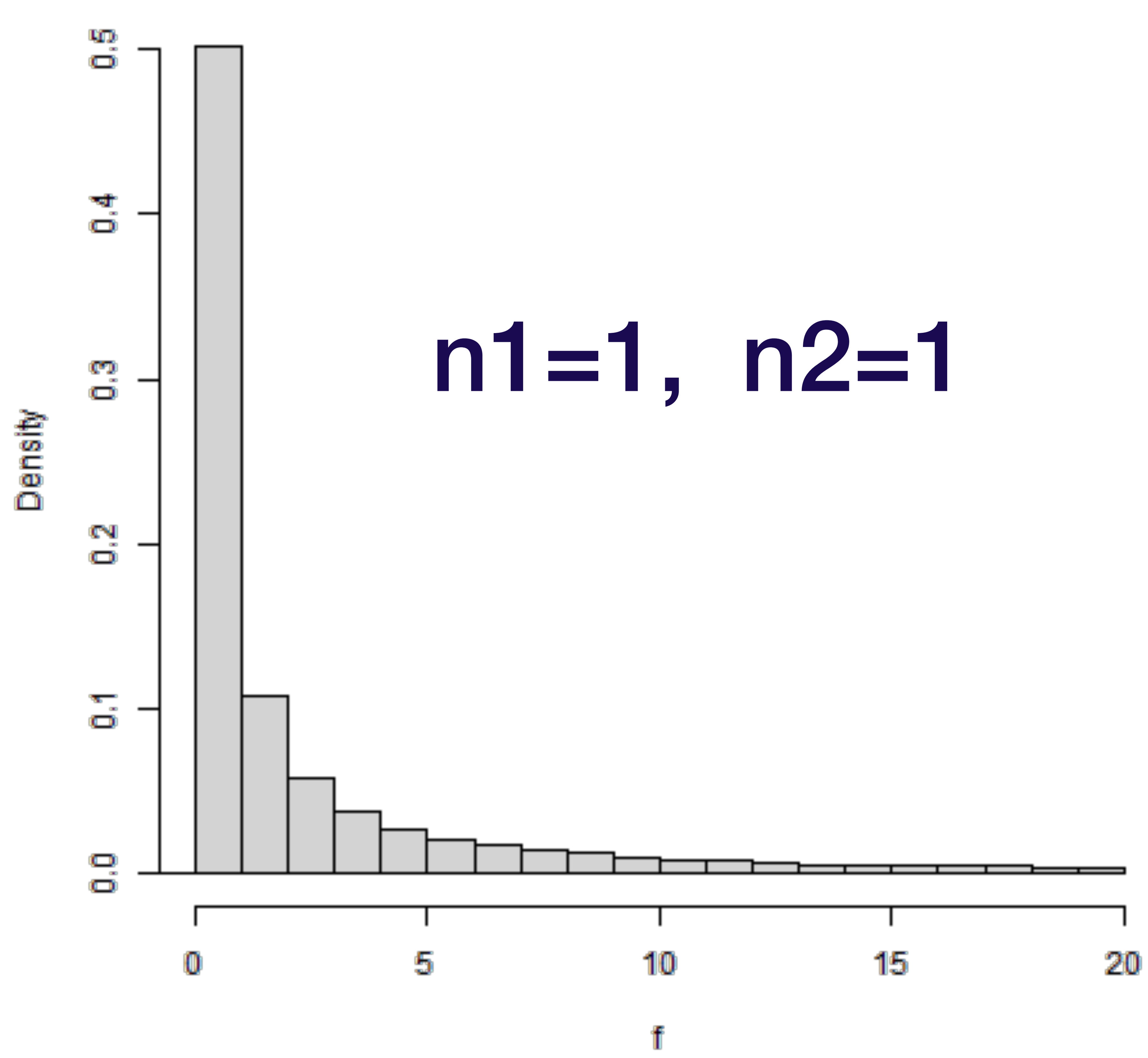
$$X_1 \sim \chi^2(n_1) \quad \text{and} \quad X_2 \sim \chi^2(n_2)$$

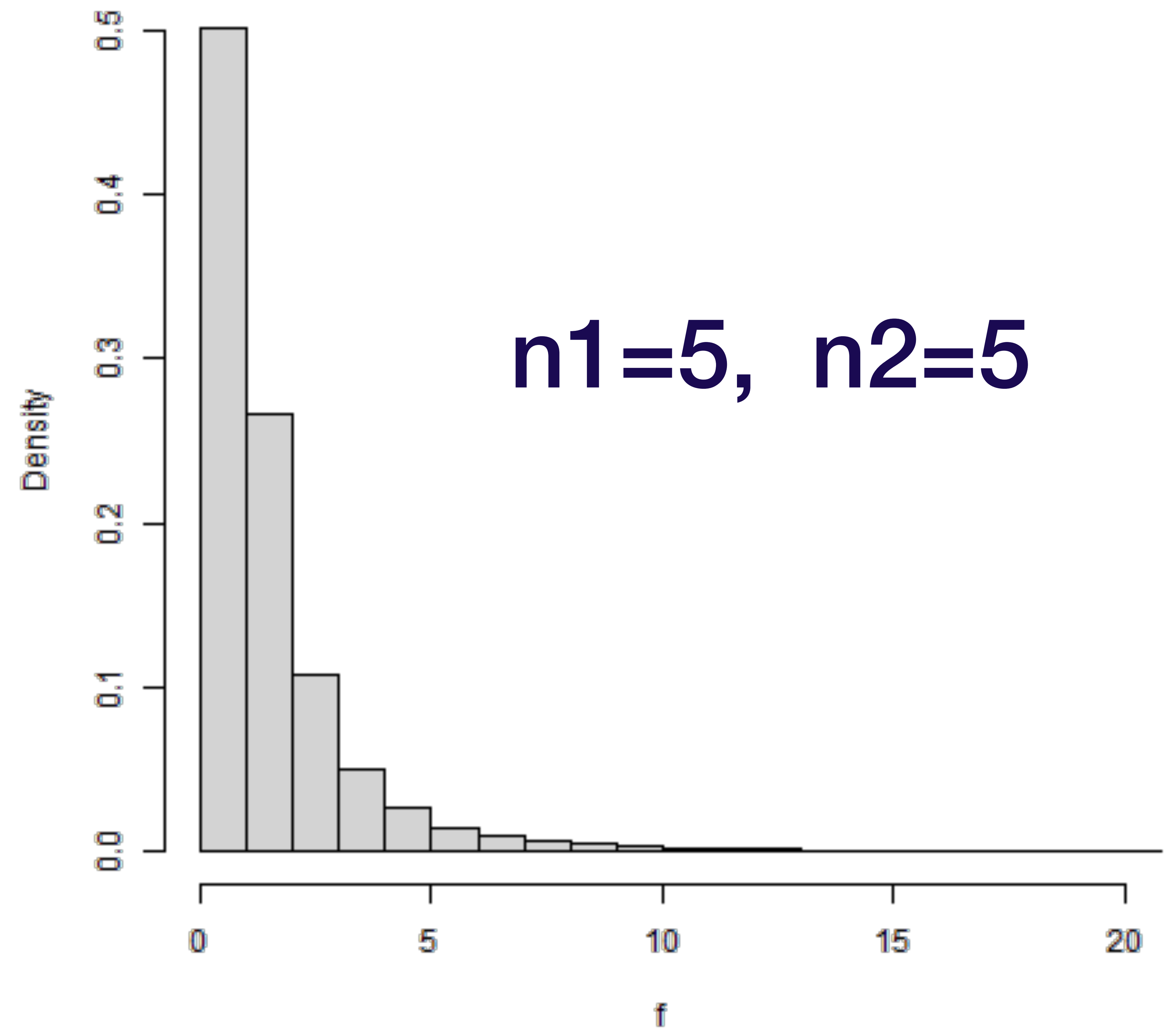
Define a new random variable

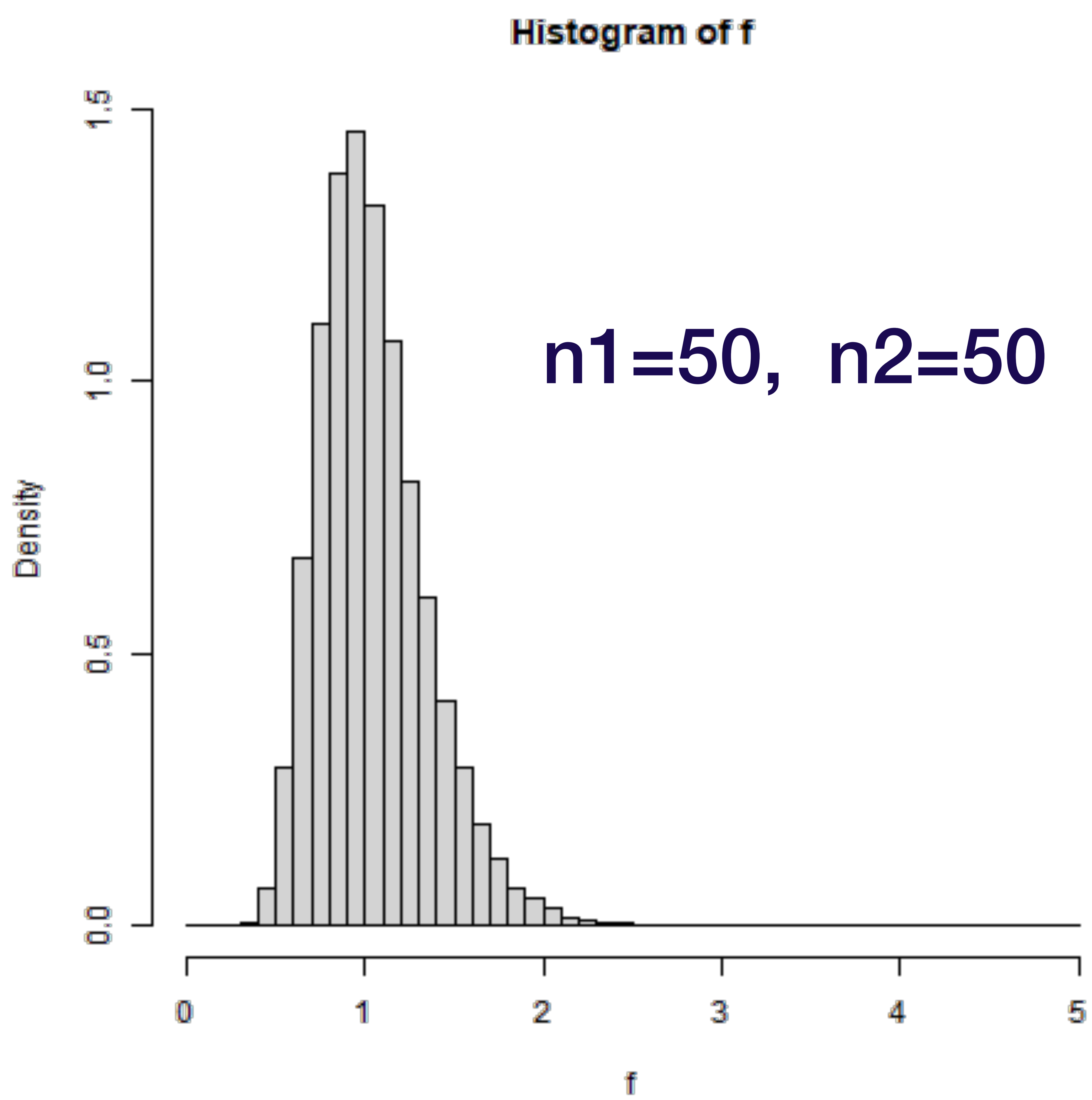
$$F = \frac{X_1/n_1}{X_2/n_2}$$

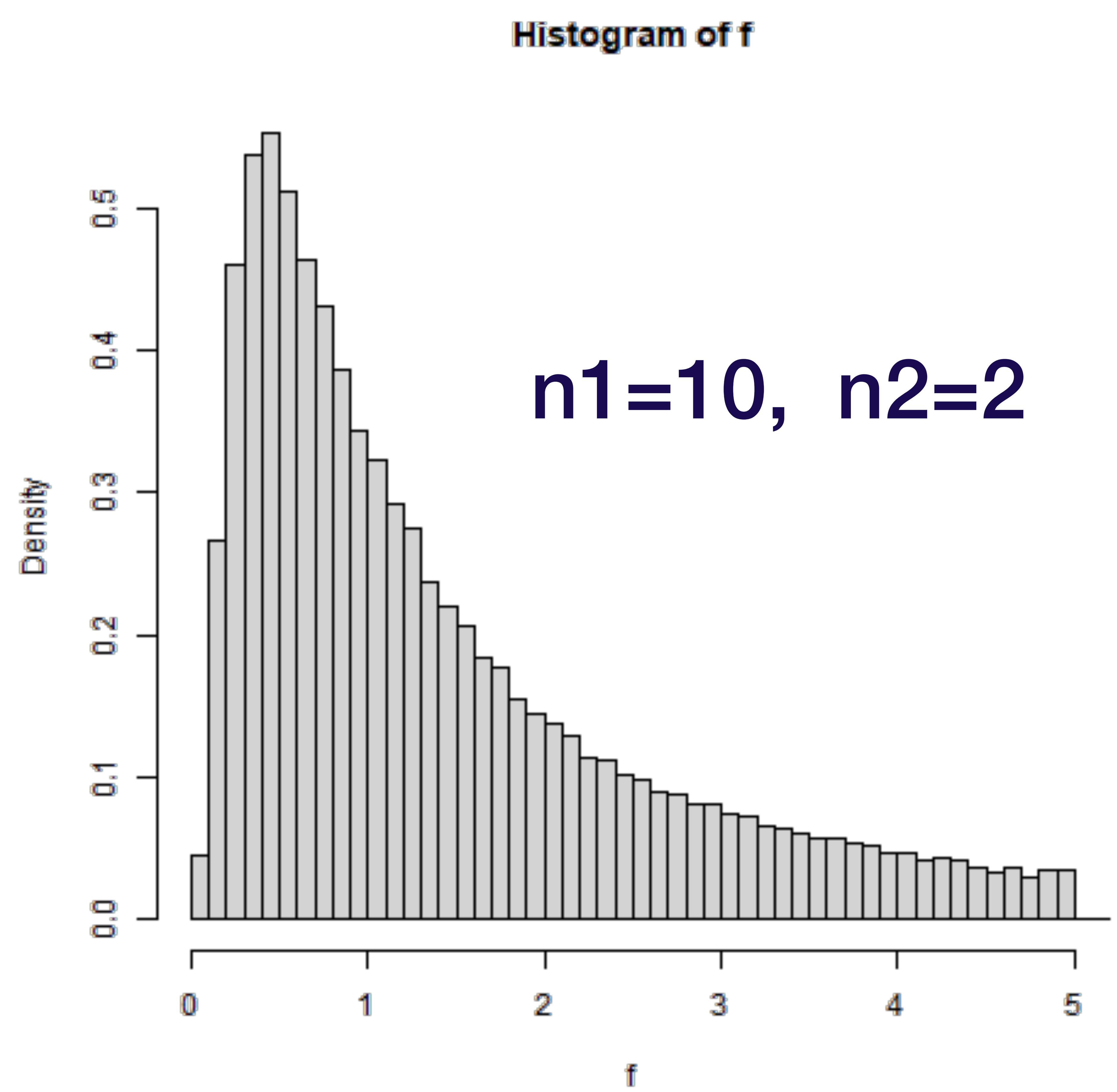
F has an “F distribution” with n_1 and n_2 degrees of freedom.

$$F \sim F(n_1, n_2)$$


n1=1, n2=1
Density
f
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20


n1=5, n2=5
Density
f
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20


Histogram of f
n1=50, n2=50
Density
f
0.0
0.5
1.0
1.5
0
1
2
3
4
5


Histogram of f
n1=10, n2=2
Density
f
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5

pdf:

$$f(x; n_1, n_2) =$$

$$\frac{1}{B(n_1/2, n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2-1} \left(1 + \frac{n_1}{n_2} x\right)^{-(n_1+n_2)/2}$$

for x>0.

---

mean: $\frac{n_2}{n_2 - 2}$ **if** $n_2 > 2$

variance: $\frac{2n_2^2 \, (n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$

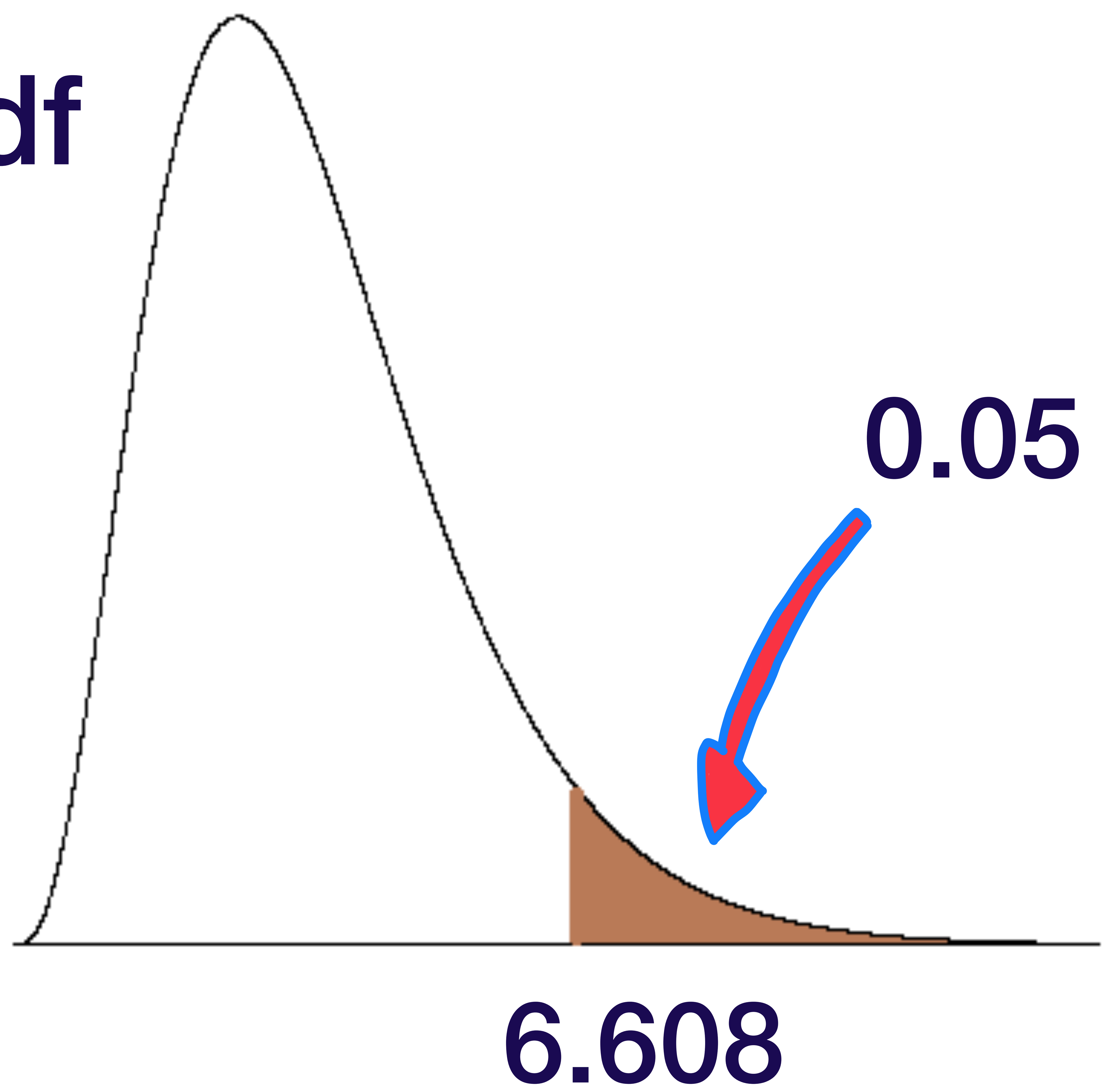
if $n_2 > 4$

In R:

qf(0.95,5,1) = 6.608

pf(6.608,5,1) = 0.9499824

F(5,1) pdf



The Mean:

$$E[F] = E\left[\frac{X_1/n_1}{X_2/n_2}\right] = \frac{n_2}{n_1} E\left[\frac{X_1}{X_2}\right]$$

$$\overset{\text{indep}}{=} \frac{n_2}{n_1} \underbrace{E[X_1]}_{n_1} \cdot E\left[\frac{1}{X_2}\right]$$

$$= n_2 E\left[\frac{1}{X_2}\right]$$

$$= n_2 \, E\left[\frac{1}{X_2}\right] = n_2 \int_{-\infty}^{\infty} \frac{1}{x} \, f_{X_2}(x)\, dx$$

$$= n_2 \int_0^{\infty} \frac{1}{x} \frac{1}{\Gamma(n_2/2)} \left(\frac{1}{2}\right)^{n_2/2} x^{n_2/2-1} \, e^{-x/2} \, dx$$

$$= n_2 \int_0^{\infty} \frac{1}{\Gamma(n_2/2)} \left(\frac{1}{2}\right)^{n_2/2} \underbrace{x^{n_2/2-2} \, e^{-x/2}}_{\text{like a } \Gamma(n_2/2 - 1,\, 1/2)} \, dx$$

$$= n_2 \, \frac{\Gamma(n_2/2 - 1)}{\Gamma(n_2/2)} \frac{1}{2} \, \cdot$$

$$\underbrace{\int_0^\infty \frac{1}{\Gamma(n_2/2 - 1)} \left(\frac{1}{2}\right)^{n_2/2-1} x^{n_2/2-2} \, e^{-x/2} \, dx}_{1}$$

$$= n_2 \, \frac{\Gamma(n_2/2 - 1)}{(n_2/2 - 1)\Gamma(n_2/2 - 1)} \frac{1}{2}$$

$$= \frac{n_2}{n_2 - 2}$$

And the point is... ?

- Suppose that $X_{11}, X_{12}, \ldots, X_{1,n_1}$ is a random sample of size n_1 from the $N(\mu_1, \sigma_1^2)$.

- Suppose that $X_{21}, X_{22}, \ldots, X_{2,n_2}$ is an independent random sample of size n_2 from the $N(\mu_2, \sigma_2^2)$.

Derive a test of size α for

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

And the point is... ?

- Suppose that $X_{11}, X_{12}, \ldots, X_{1,n_1}$ is a random sample of size n_1 from the $N(\mu_1, \sigma_1^2)$.

- Suppose that $X_{21}, X_{22}, \ldots, X_{2,n_2}$ is an independent random sample of size n_2 from the $N(\mu_2, \sigma_2^2)$.

Derive a test of size α for

$$H_0 : \sigma_1^2/\sigma_2^2 = 1 \qquad H_1 : \sigma_1^2/\sigma_2^2 \neq 1$$

Let S_1^2 and S_2^2 be the sample variances for the first and second samples, respectively.

We know that

$$\frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$$

and

$$\frac{(n_2 - 1)\, S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

are independent

So, define a test statistic F as

$$F := \frac{[(n_1 - 1)S_1^2/\sigma_1^2]/(n_1 - 1)}{[(n_2 - 1)S_2^2/\sigma_2^2]/(n_2 - 1)} = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2}$$

Then

$$F \sim F(n_1 - 1, n_2 - 1)$$

Similarly

$$\frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{S_2^2}{S_1^2} \sim F(n_2 - 1, n_1 - 1)$$

Under the assumption that H_0 is true, we have that

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

and that

$$\frac{S_2^2}{S_1^2} \sim F(n_2 - 1, n_1 - 1)$$

Derive a test of size α for

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

- We will reject H_0 if S_1^2/S_2^2 is too small or too large.
- Equivalently, we reject H_0 if S_2^2/S_1^2 is too large or too small.

Convention: Put the larger sample variance in the numerator and reject H_0 is above the appropriate upper $\alpha/2$ critical value.

Example:

Fifth grade students from two neighboring counties took a placement exam.

Group 1, from County A, consisted of 18 students. The sample mean score for these students was 77.2.

Group 2, from County B, consisted of 15 students and had a sample mean score of 75.3.

Example:

From previous years of data, it is believed that the scores for both counties are normally distributed, and that the variances of scores from Counties A and B, respectively, are 15.3 and 19.7.

You wish to create a confidence interval for $\mu_1 - \mu_2$, the difference between the true population means.

Example:

You are thinking of using a pooled-variance two-sample t-test, however this requires that the true population variances, σ_1^2 and σ_2^2 are the same.

Derive a test of size α for

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Example:

From previous years of data, it is believed that the scores for both counties are normally distributed.

The sample variances of scores from Counties A and B, respectively, are 15.3 and 19.7.

Derive a test of size α for

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Step One:

Choose a test statistic.

$$F := \frac{S_1^2}{S_2^2} \quad \text{or} \quad F := \frac{S_2^2}{S_1^2}$$

Use the one that is greater than 1.

Step Two:

Give the form of the test.

Reject H_0, in favor of the alternative if F is either too large or too small.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Step Two (continued):

Reject H_0, in favor of the alternative if F is either too large or too small.

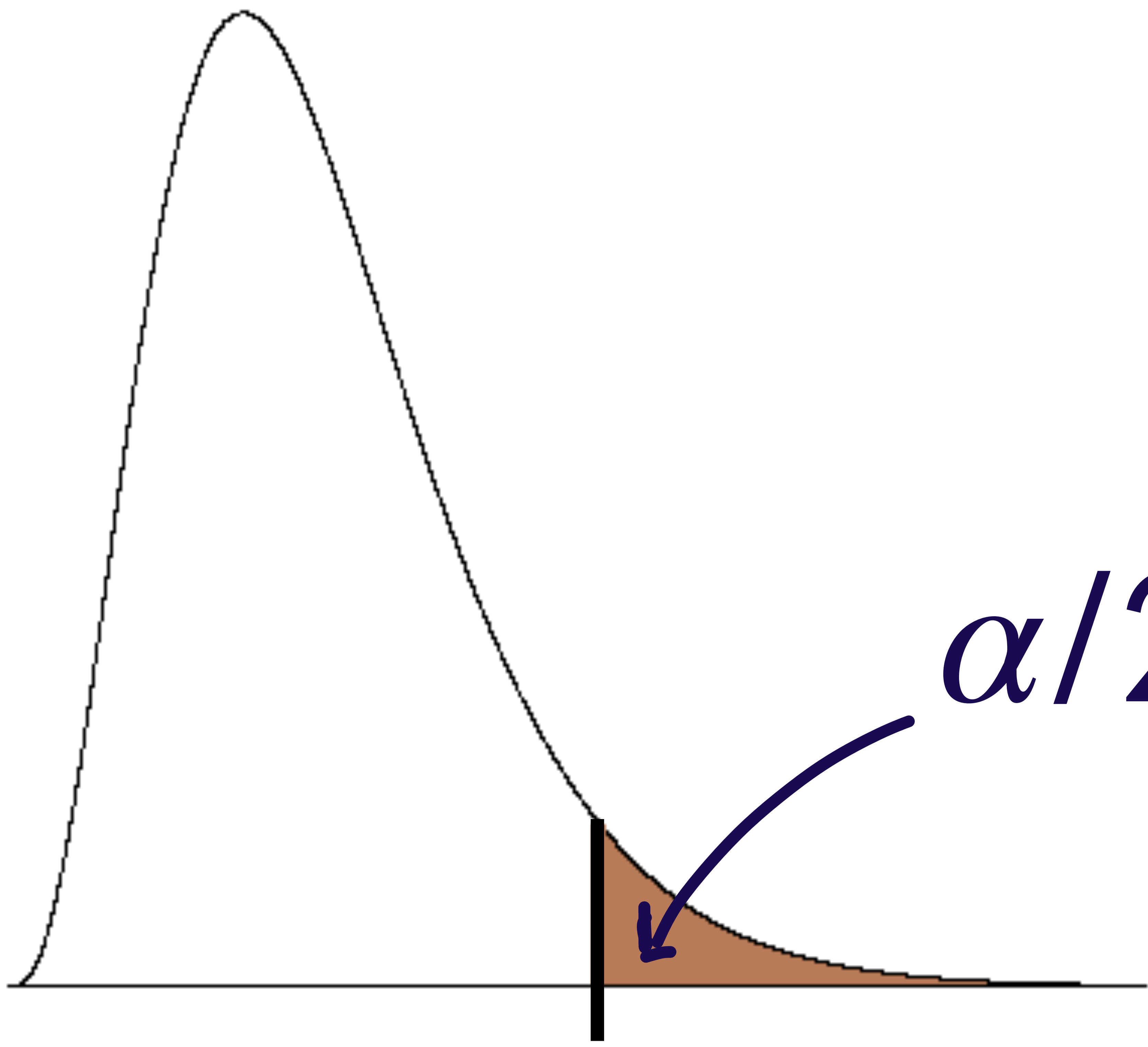
Since we chose to use the larger ratio for our test statistic, we will only reject if it is in the upper tail of the rejection region.

Note that this upper tail will have area $\alpha/2$.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Step Three:

Find the cutoff using $\alpha/2$.



$$F_{\alpha/2, n_i - 1, n_j - 1}$$

"numerator degrees of freedom"
vs
"denominator degrees of freedom"

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Step Four:

Conclusion.

Reject H_0, in favor of H_1, if

$$F > F_{\alpha/2, n_i - 1, n_j - 1}$$

Example:

$$n_1 = 18, \quad s_1^2 = 15.3 \qquad \alpha = 0.05$$

$$n_2 = 15, \quad s_2^2 = 19.7$$

$$F := \frac{S_2^2}{S_1^2} = \frac{19.7}{15.3} \approx 1.288$$

Critical value:

$$F_{0.025,17,14} = 2.900$$

qf(0.975,17,14)

Example:

$n_1 = 18, \quad s_1^2 = 15.3 \qquad \alpha = 0.05$

$n_2 = 15, \quad s_2^2 = 19.7$

The test statistic $F \approx 1.288$

does not fall above

$$F_{0.025,17,14} = 2.900$$

Thus we fail to reject H_0.