Chapter 7: Inferences about Population Variances and Standard Deviations

1. Some Perspective. Why Study Standard Deviations?

2. Studying σ for a single population.
   a. Hypothesis tests
   b. Confidence Intervals
   c. Sensitivity to assumptions

3. Comparing two standard deviations σ_1/σ_2 .
   a. Hypothesis tests
   b. Confidence interval
   c. Sensitivity to assumptions
   d. Levene's test

Example: Aphid Example

1. Some Perspective

CH5: Inference about a single <u>mean</u> (μ)
CH6: Inference about difference between two <u>means</u> (μ_1 - μ_2)
CH7: Inference about one or two standard deviations (σ) or
    variances (σ^2)
CH8: Inference about more than two <u>means</u> (μ_1, $\mu_{2,}$ …., μ_t)

CH5,6: Normal and t distributions

- Both symmetric
- Positive or Negative values

CH 7: χ^2 and F distributions

- Not symmetric
- Strictly positive

Why study standard deviations (or variances)?

A) Sometimes σ_1 and/or σ_2 are the parameters of interest.

Confidence intervals or tests about a single σ.
Compare σ_1, σ_2 using confidence intervals or tests for the ratio.

<u>Example:</u>
Two machines - both making a part that is <u>supposed to be</u> *x* units wide.
<u>machine 1</u> produces parts with mean μ_1 and std. dev. σ_1.
<u>machine 2</u> produces parts with mean μ_2 and std. dev. σ_2.
Say, we can adjust either machine (by turning adjustment screws) until $\mu_1 = x$ and $\mu_2 = x$. Then the machine with the <u>smaller</u> std.dev. will produce the <u>better</u> (more consistent) product.

B) When comparing population means, should we use the pooled (equal variance) t-test or the Welch-Satterthwaite (unequal variance) t-test? Need to know if we can demonstrate the need for using Welch-Satterthwaite test.

2. Studying σ for a single population

Example:

To study the precision of a lab instrument we do n = 10 serum cholesterol determinations on the same (well mixed) blood sample. A perfect instrument would yield the same result each time, but our imperfect instrument yields: $\bar{y} = 210 \qquad s = 10.2$

(We do not know the true mean, so we cannot evaluate whether the 210 is too high or low. That is an "accuracy" issue; we are interested in "precision", i.e. variance)

The manufacturer claims that the true standard deviation of results from this instrument is 5.0 mg/dl or less. Are the data consistent with this claim, or do they contradict this claim?

The claim suggests a one-sided alternative:

$$H_0 : \sigma \le 5.0 \quad (\sigma_0 = 5.0)$$
$$H_A : \sigma > 5.0$$

Rejection of H_0 in favor of H_A would be evidence against the claim.

χ2 Test for σ^2 (or σ)

Assumptions: Random sample, independent observations, normally distributed data.

Test Statistic: $$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

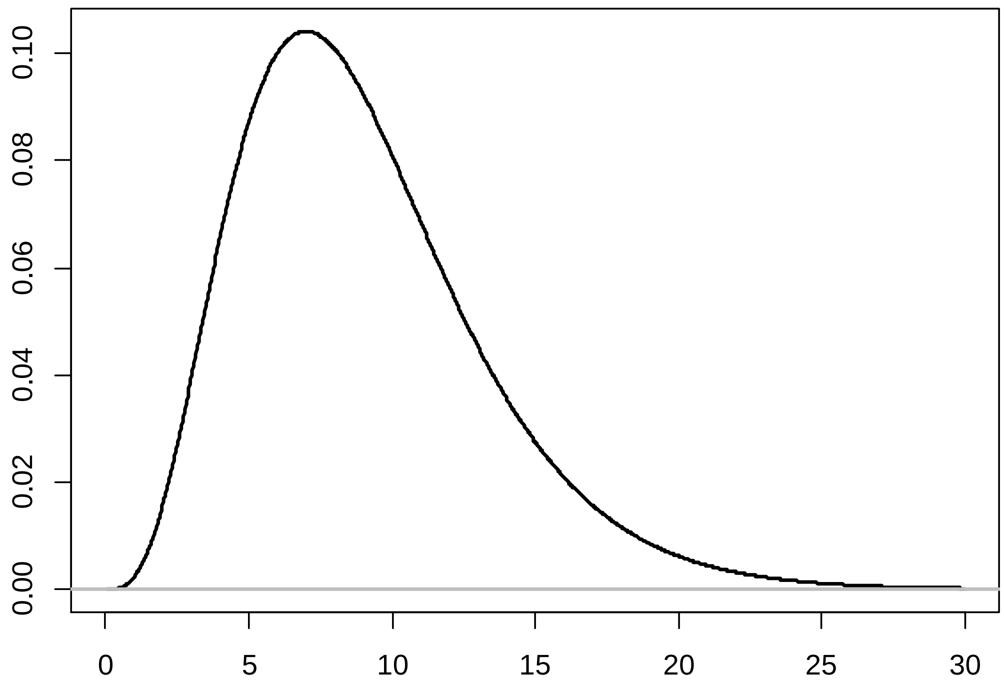
Note: df = n-1

Alternative Hypothesis:	**Rejection Region:**
(1) $H_A: \sigma^2 > \sigma^2_0$	$\chi^2 > \chi^2_{\alpha,n-1}$
(2) $H_A: \sigma^2 < \sigma^2_0$	$\chi^2 < \chi^2_{1-\alpha,n-1}$
(3) $H_A: \sigma^2 \neq \sigma^2_0$	$\chi^2 > \chi^2_{\alpha/2,n-1}$ or $\chi^2 < \chi^2_{1-\alpha/2,n-1}$

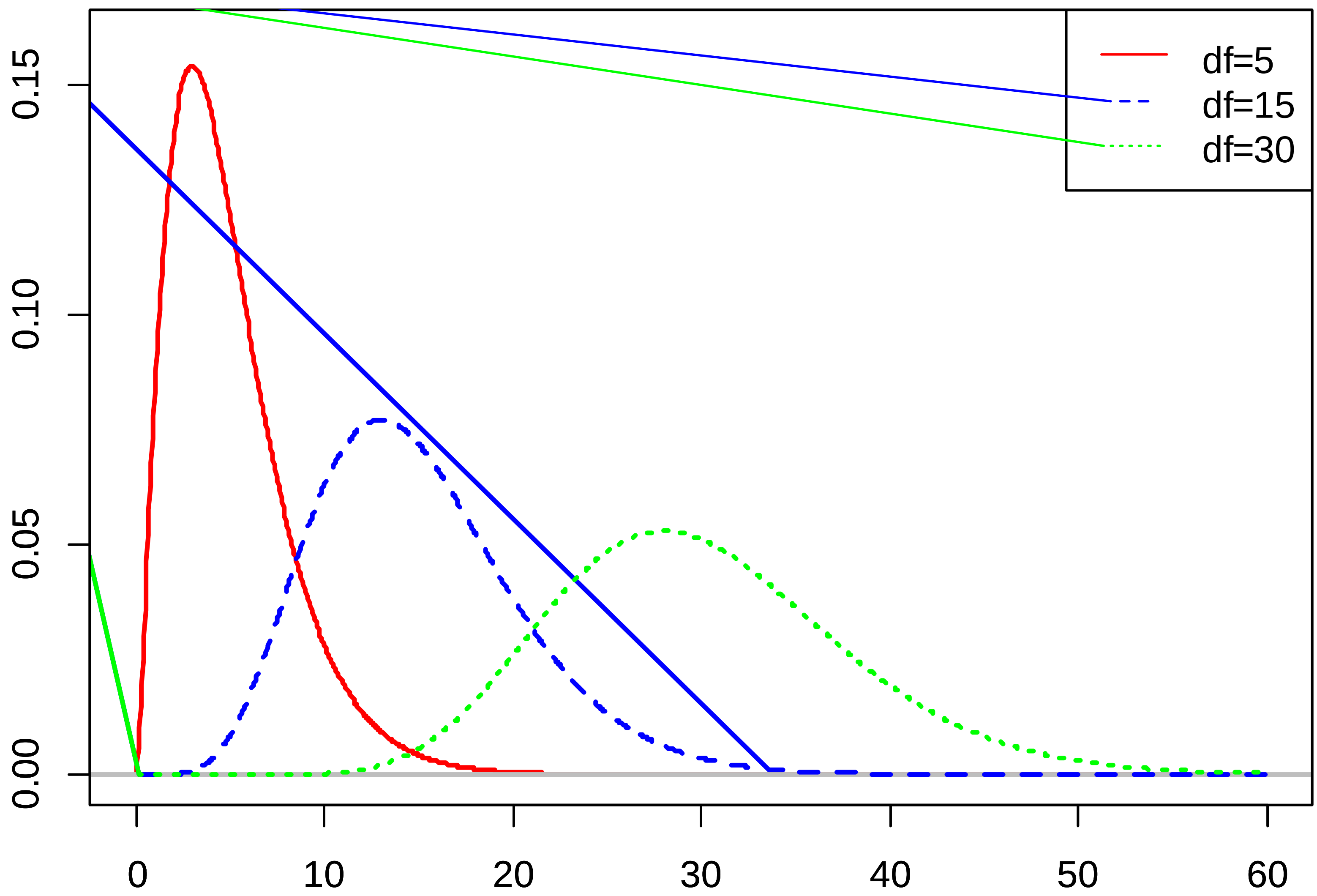
P-values: (1) area to the right of χ^2 (test statistic), (2) area to the left, (3) double the smaller of the first two areas.

Comments:

1. We use s^2 to test hypotheses about σ^2 rather than use s to test hypotheses about σ.
2. If H_0 is true, then the test statistic $\chi^2=(n-1)s^2/\sigma_0^2$ has a Chi-square distribution with degrees of freedom (df) = n-1 .
3. The mean and variance of the chi-square distribution are given by μ = df and σ^2 = 2df.

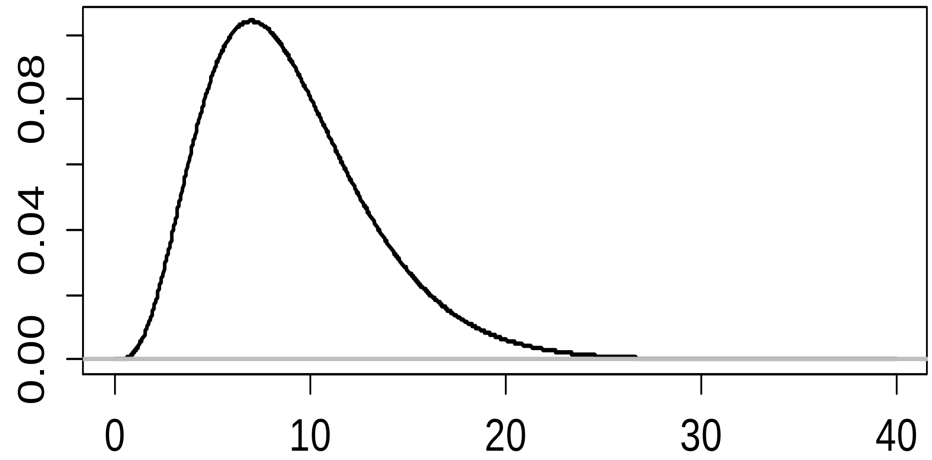


Example Chi-Square Distributions



Example: Right One-sided Alternative

1. $H_0: \sigma \leq 5.0$ vs. $H_A: \sigma > 5.0$



2. Test Statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)(10.2)^2}{5^2} = 37.45$$

3. Rejection Region:

df = n – 1= 10 – 1 = 9

From Table 7 (right side) $\chi^2_{0.05} = 16.92$

Using R: `critval = qchisq(0.95, df = 9)`

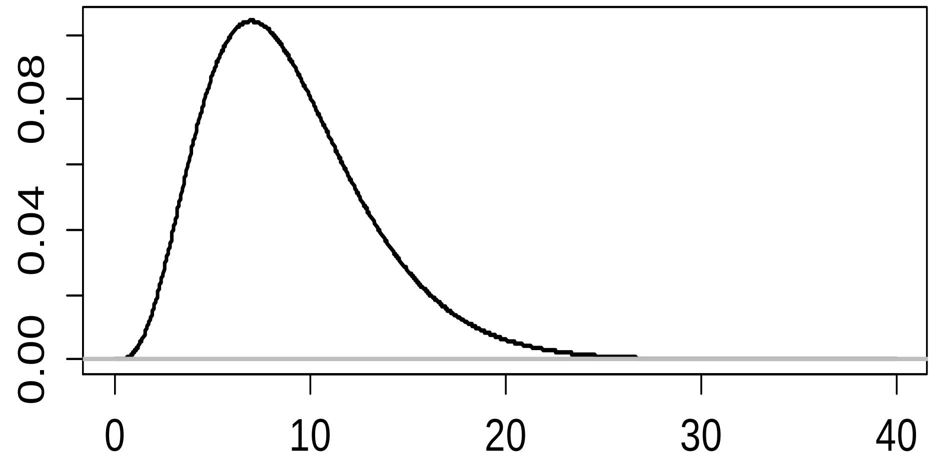
4. Since $\chi^2 = 37.45 > 16.92 = \chi^2_{\alpha,n-1}$ -> Reject H_0.

P-value (using R): `pvalue = 1-pchisq(37.45, df = 9)`

p-value=0.0000219

Example: Left One-sided Alternative

1. H_0: $\sigma \geq 5.0$ vs. H_A: $\sigma < 5.0$



2. Test Statistic (same as before):

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)(10.2)^2}{5^2} = 37.45$$

3. Rejection Region:

df = n – 1= 10 – 1 = 9

From Table 7 (left side) $\chi^2_{0.95} = 3.325$

Using R: `critval = qchisq(0.05,df = 9)`

4. Since $\chi^2 = 37.45$ which is not $< 3.325 = \chi^2_{1-\alpha,n-1}$ → Fail to Reject H_0.

---

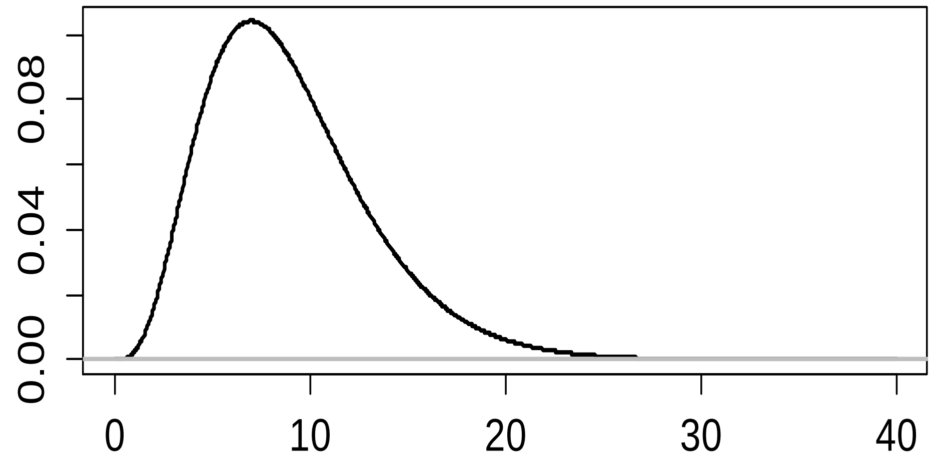
p-value (using R): `pvalue = pchisq(37.54, df = 9)`

p-value=0.99998

Note: The p-value for the right alternative and the p-value for the left alternative sum to 1.0.

Example: Two-sided Alternative

1. H_0: σ =5.0 vs. H_A: σ ≠ 5.0



2. Test Statistic (same as before):

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)(10.2)^2}{5^2} = 37.45$$

3. Rejection Region:

Reject H_0 if

$\chi^2 < \chi^2_{0.975} = 2.70$ or

$\chi^2 > \chi^2_{0.025} = 19.02$

4. Since $\chi^2 = 37.45 > 19.02$ -> Reject H_0.

Note: Two-sided P-value is calculated by doubling the smaller of the two one-sided p-values.

In this case, p-value = 2*0.0000219

Derivation of Confidence Interval for σ^2

A Confidence interval is derived by starting with a probability statement about the <u>test statistic</u>, then manipulating the inequalities so that the <u>parameter</u> is in the middle.

$$P\left(\chi^2_{0.975} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{0.025}\right) = 0.95$$

$$P\left(\frac{1}{\chi^2_{0.025}} < \frac{\sigma^2}{(n-1)s^2} < \frac{1}{\chi^2_{0.975}}\right) = 0.95$$

$$P\left(\frac{(n-1)s^2}{\chi^2_{0.025}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.975}}\right) = 0.95$$

Confidence interval for σ^2 (or σ)

Assumptions: Random sample, independent observations, normally distributed data.

Note: df = n -1

The 100(1-α)% confidence interval for $\boldsymbol{\sigma^2}$ is given by:

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right)$$

The 100(1-α)% confidence interval for $\boldsymbol{\sigma}$ is given by:

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}} \right)$$

For our example, the 95% CI for σ is:

$$\sqrt{\frac{(10-1)(10.2)^2}{19.02}} < \sigma < \sqrt{\frac{(10-1)(10.2)^2}{2.700}} \qquad \text{i.e., } 7.02 < \sigma < 18.62$$

Comments about Confidence Interval for σ

1) CI not symmetric about the point estimate. It is skewed to the right.

2) You will reject (in a two-sided test, α=0.05) any σ not in the 95% confidence interval.

3) O and L write $\chi^2_{\alpha/2}$ as χ^2_U and $\chi^2_{1-\alpha/2}$ as χ^2_L , referring to "upper" and "lower" table values.

4) Tests and Confidence Intervals assume that individual observations are from the normal distribution.

5) This hypothesis test and confidence interval are **very affected** by outlier-prone distributions (heavy-tailed distributions) and skewed distributions.

Simulation to study effect of Outliers and Skewness

Data simulated under H0. Test using α=0.05. Type 1 error rates recorded. For details see simulation in CH7 of O&L.

Scenario 1: H_0: $\sigma^2 \leq 100$ vs H_A: $\sigma^2 > 100$

	Normal	Uniform	t	Gamma(1)	Gamma(.1)
n=10	0.047	0.004	0.083	0.134	0.139
n=20	0.052	0.006	0.103	0.139	0.175
n=50	0.049	0.004	0.122	0.156	0.226

Scenario 2: H_0: $\sigma^2 \geq 100$ vs H_A: $\sigma^2 < 100$

	Normal	Uniform	t	Gamma(1)	Gamma(.1)
n=10	0.046	0.018	0.119	0.202	0.213
n=20	0.050	0.011	0.140	0.213	0.578
n=50	0.051	0.018	0.157	0.220	0.528

Conclusion: Far too many rejections with heavy tails (t) or skew (gamma). **Far too few** rejections with short tails (unif).

Tolerance Intervals (Optional):

A **tolerance interval** is a range that is likely to contain a specified proportion of the population. To generate tolerance intervals, you must specify both the proportion of the population (p) and a confidence level (1-α).

A commonly used approximate interval (from Howe 1969) is

$$\bar{y} \pm Z_{(0.5-\frac{p}{2})}\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha}}s^2\left(1+\frac{1}{n}\right)}$$

Note that this interval uses the one-sided upper confidence limit for σ^2.

3. Comparing two standard deviations or variances

Example with a right alternative: We now compare two instruments that measure serum cholesterol. Assume we have previously compared mean response, and are now interested in which machine gives more *consistent* (i.e. less variable) results.

We do $n_1=10$ determinations using machine 1, and $n_2=15$ determinations using machine 2, all on the same (well mixed) sample. We observe:

$s_1=15.4$ and $s_2=12.3$

Does this result support a claim that machine 2 is better than machine 1 (i.e. $\sigma_1 > \sigma_2$)?

$$H_0 : \sigma_1/\sigma_2 \leq 1$$

$$H_A : \sigma_1/\sigma_2 > 1 \text{ (a right alternative)}$$

Rejection of H_0 in favor of H_A would be evidence in favor of the claim. We will use $\alpha=0.05$.

F-test for Comparison of 2 Variances or Standard Deviations

Assumptions: Independent, random samples, normally distributed data.

Test Statistic:

$$F = s_1^2 / s_2^2$$

Alternative Hypothesis:	**Rejection Region:**
(1) $H_A: \sigma_1^2 > \sigma_2^2$	$F > F_{\alpha,df1,df2}$
(2) $H_A: \sigma_1^2 < \sigma_2^2$	$F < F_{1-\alpha,df1,df2}$
(3) $H_A: \sigma_1^2 \neq \sigma_2^2$	$F > F_{\alpha/2,df1,df2}$ or $F < F_{1-\alpha/2,df1,df2}$

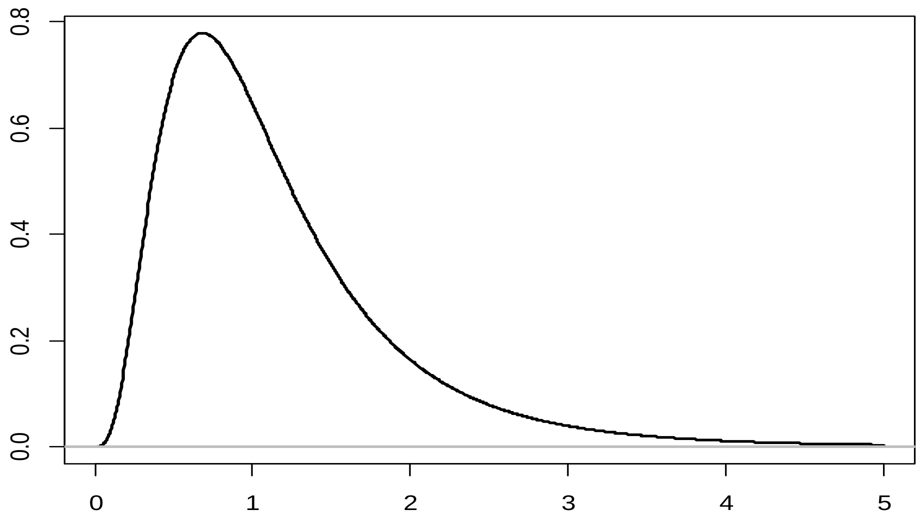
Where $df_1 = ndf = n_1 - 1$ and $df_2 = ddf = n_2 - 1$.

P-values: (1) area to the right of F (test statistic), (2) area to the left, (3) double the smaller of the first two areas.

Hypothesis test uses two facts:

1) We can use s_1^2/s_2^2 to test hypotheses about σ_1^2/σ_2^2 rather than use s_1/s_2 to test $H_0 : \sigma_1/\sigma_2$

2) $F = \dfrac{s_1{}^2/\sigma_1{}^2}{s_2{}^2/\sigma_2{}^2}$ has an F distribution with $\text{df}_1 = dfn = n_1 - 1$, $\text{df}_2 = dfd = n_2 - 1$. (If H_0 is true then $F = s_1{}^2 / s_2{}^2$, which can be calulated from the data.)



Notes: 1) F distribution looks a lot like a Chi-square.

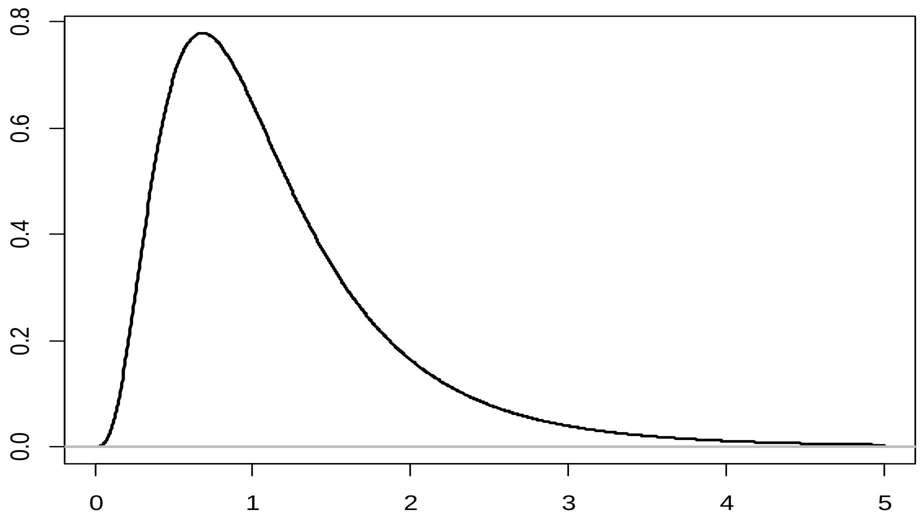
2) When df_2 is very large, it is a Chi-square (because the s_2^2/σ_2^2 in the denominator is nearly a constant = 1)

Example: Right One-sided Alternative

1. $H_0: \sigma_1/\sigma_2 \leq 1$ vs $H_A: \sigma_1/\sigma_2 > 1$
2. Test Statistic:

$$F = s_1^2 \big/ s_2^2 = (15.4)^2 /(12.3)^2 = 1.57$$

3. Rejection Rule:
   Reject H_0 if $F > F_{0.05}=2.65$ (From Table 8 using $df_1=9$ and $df_2=14$)
4. Since $F=1.57$ which is not $> F_{0.05}=2.65$, we Fail to Reject H_0.



<u>In R</u>:
```
fcrit = qf(0.95, df1=9, df2=14)
pvalue = 1-pf(1.57, df1=9, df2=14)
```
From R: pvalue=0.217

Problem computing the critical value for the left alternatives:

Very few books have lower F-percentiles.

Possible solutions:

1. Get it from R:
```
fcrit = qf(alpha, dfn, dfd)
pvalue = pf(F, dfn, dfd)
```

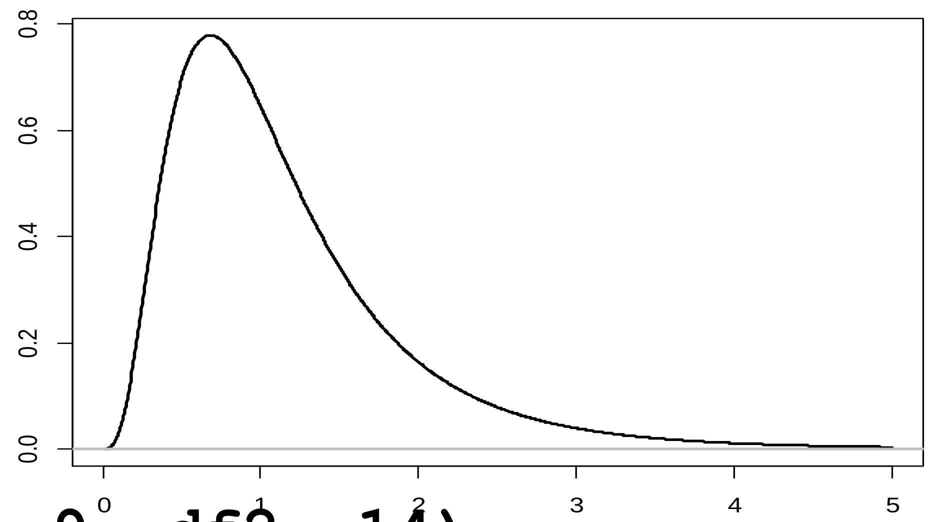
2. Get it by inverting upper values: $F_{1-\alpha:dfn,dfd} = \dfrac{1}{F_{\alpha:dfd,dfn}}$

   Note the switch of dfn and dfd.

3. An easy trick to avoid the problem: Reverse the labels on the groups so that the alternative is always a right alternative.

NOTE: Any of these will work. If you are using method (1), you could get the p-value at the same time.

Example: Left One-sided Alternative



1. $H_0: \sigma_1/\sigma_2 \geq 1$ vs. $H_A: \sigma_1/\sigma_2 < 1$
2. Test Statistic (same as before):

$$F = s_1^2 \big/ s_2^2 = (15.4)^2 / (12.3)^2 = 1.57$$

3. Rejection Rule:
   Reject H_0 if $F < F_{0.95}$=0.33
   In R: `fcrit = qf(0.05, df1=9, df2= 14)`
   Using Table 8:

$$F_{1-\alpha:dfn,dfd} = \frac{1}{F_{\alpha:dfd,dfn}}$$

$$F_{0.05:14,9} = 3.01 \text{ (Using } df_1 = 15 \text{ because there is no 14 column)}$$

(Note the switch of *dfn* and *dfd*.)

$$F_{0.95:9,14} = \frac{1}{3.01} = 0.332$$

4. Since F=1.57 which is not < $F_{0.95}$=0.33 , we Fail to Reject H_0.

Computing the critical values for the two-sided alternatives:

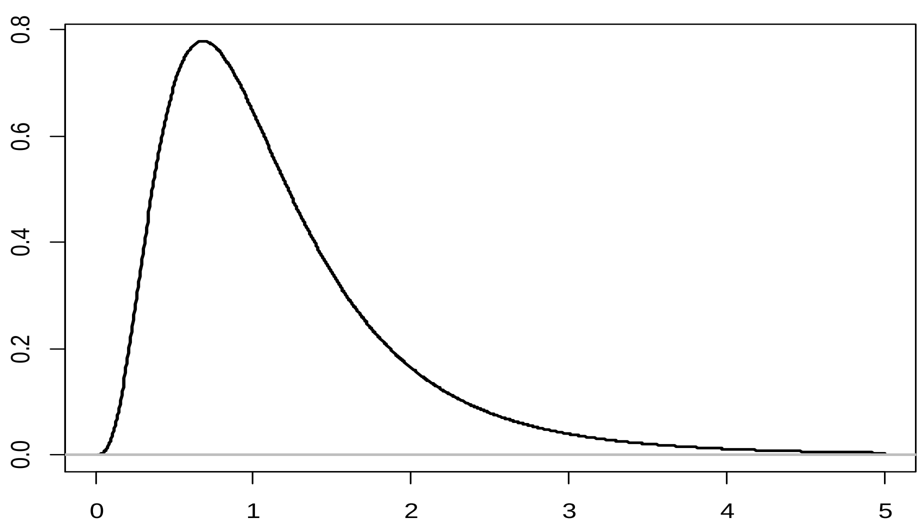
1. R Crit. values:
```
lower = qf(alpha/2, dfn, dfd)
upper = qf(1-alpha/2, dfn, dfd)
```

2. Critical values from Table 8:

$$F_{1-\alpha/2:dfn,dfd} = \frac{1}{F_{\alpha/2:dfd,dfn}}, \qquad F_{\alpha/2:dfn,dfd}$$

3. Or use another trick: Put the larger sample variance in the numerator, and test against a right alternative using:

Reject H_0 if: $F > F_{\alpha/2:dfn,dfd}$ (Note the $\alpha/2$)



Example: Two-sided Alternative

1. H_0: $\sigma_1/\sigma_2 = 1$ vs H_A: $\sigma_1/\sigma_2 \neq 1$
2. Test Statistic: F=1.57
3. Finding Critical Values:

   1. R Crit. vals: `lower = qf(0.025, df1=9, df2=14)`
   `upper = qf(0.975, df1=9, df2=14)`

   Results: Lower = 0.263 Upper = 3.21

   2. Critical values from Table 8:

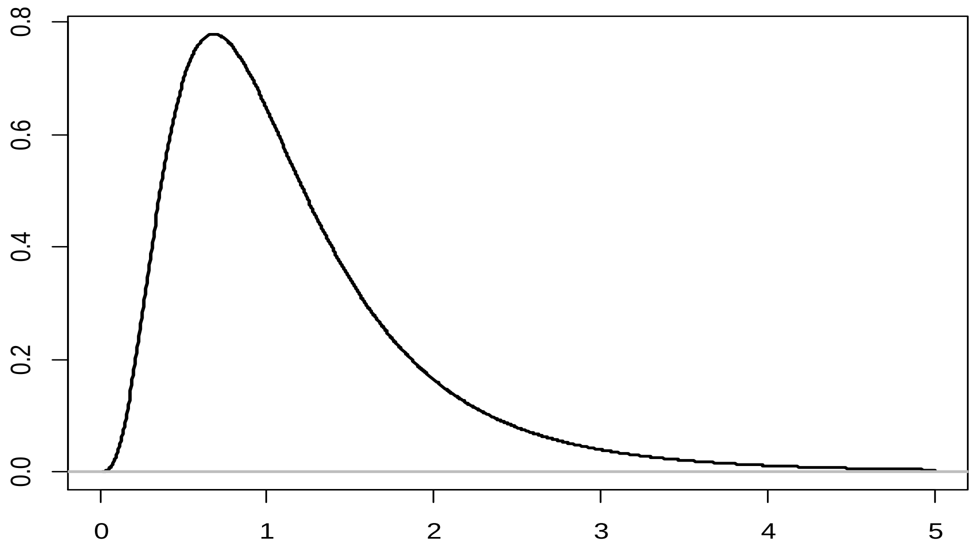
   $$F_{0.975,9,14} = \frac{1}{F_{0.025,14,9}}, \quad \frac{1}{3.77} = 0.266 \text{ (used dfn=15)}$$

   $$F_{0.025,9,14} = 3.21$$

   3. Since the <u>larger sample variance is already in the numerator,</u> compare the computed F to the <u>upper</u> critical value computed above.

4. We Fail to Reject H_0.

Example calculations: p-values for two-sided alternatives



For 2-sided p-values, double the smaller of the one-sided p-values

In R: `pvalue = 2*min(pf (1.57,9,14),1-pf (1.57,9,14))`

In our example: The one-sided p-values were 0.217 and 0.783.

Two-sided p-value is: 2(0.217) = 0.434

Confidence Interval for σ_1/σ_2

$$\sqrt{\frac{s_1^2}{s_2^2}F_L} < \frac{\sigma_1}{\sigma_2} < \sqrt{\frac{s_1^2}{s_2^2}F_U}$$

Assumptions: Independent, random samples, normally distributed data.

In the notation of the text,

$$F_L = \frac{1}{F_{\alpha/2,dfn,dfd}} \text{ and } F_U = \frac{1}{F_{1-\alpha/2,dfn,dfd}} = F_{\alpha/2,dfd,dfn}$$

For the example, we get a 95% CI for $\frac{\sigma_1}{\sigma_2}$ as

$$\sqrt{\frac{(15.4)^2}{(12.3)^2}(0.31)} < \sigma_1/\sigma_2 < \sqrt{\frac{(15.4)^2}{(12.3)^2}(3.80)} \text{ i.e., } 0.70 < \sigma_1/\sigma_2 < 2.44$$

$$\text{using } \frac{1}{F_{0.025,9,14}} = F_L = 0.31 \text{ and } \frac{1}{F_{0.975,9,14}} = F_{0.025,14,9} = F_U = 3.80.$$

A memory aid: When computing CI limits: either invert the upper $\alpha/2$ value, or reverse the df on the upper $\alpha/2$ value, not both.

F-test and CI for 2 Variances in R/Rcmdr

- In R, use the function var.test().
- In Rcmdr, choose Statistics -> Variances -> Two-variances F-test.

- For the Rat Lead Example: (See also "**Aphid Example**".)

```
> var.test(y ~ trt, data=RatLead)
     F test to compare two variances
data:  y by trt
F = 0.653, num df = 9, denom df = 9, p-value =
0.5356
alternative hypothesis: true ratio of variances
is not equal to 1
95 percent confidence interval:
 0.162208 2.629170
sample estimates:
ratio of variances
         0.6530487
```

Inference for Variances or Standard Deviations
(sensitivity to assumptions)

F-tests for comparing two variances, and CIs based on those F-tests, assume <u>normality of the individual observations</u>. Failure of this assumption causes far too many rejections in the tests and far too low coverage in the CIs. This is a similar result to the single variance case.

Levene's test

Levene's test is a <u>commonly used</u>, robust test for equality of variances. Can be used to test equality of two (or more) variances without the assumption of normality.

Can think of testing H_0: $\sigma_1^2=\sigma_2^2$ versus H_A: $\sigma_1^2\neq\sigma_2^2$

Idea: Compute the absolute values of the deviations from the sample means ($|y_{ij} - \bar{y}_i|$), then do a t-test to compare groups.

Alternative: When the deviations are calculated using the median (instead of the mean), the test is called the Brown-Forsythe test. This is the default in R.

In practice, people use Levene's test more often than the F-test!

Levene’s Test in R/Rcmdr

- In R, use the function leveneTest(). Need to load the car package first!
- In Rcmdr, choose Statistics -> Variances -> Levene’s Test.
- Note: Can choose to calculate the center using mean (classic) or median (R default, more robust) .
- For the Aphid Example: (See “**Levene’s Test**”)

```
> library(car)
> leveneTest(Aphids~Trt, data=Aphids)
Levene's Test for Homogeneity of Variance
(center = median)
      Df F value Pr(>F)
group  1  0.7575 0.3935
      22
```