

STA258H5

Al Nosedal
and Alison Weir

Winter 2017

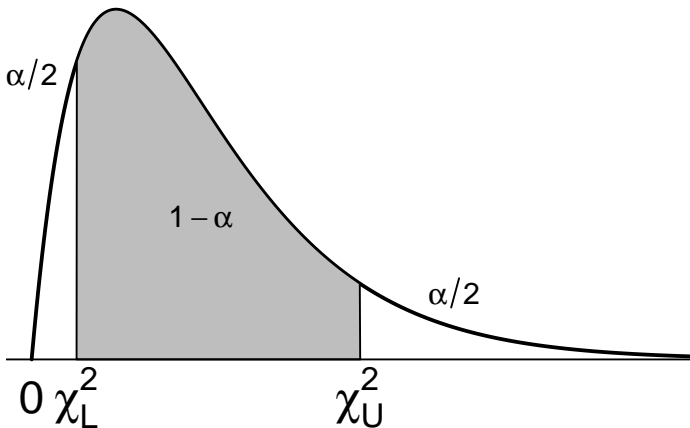
CONFIDENCE INTERVALS FOR σ^2

We know from Theorem 7.3 that $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2 distribution with $(n-1)$ df.

We can then proceed by the **pivotal method** to find two numbers χ_L^2 and χ_U^2 such that

$$P \left[\chi_L^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_U^2 \right] = 1 - \alpha$$

for any confidence coefficient $(1 - \alpha)$. (The subscripts L and U stand for **lower** and **upper**, respectively.)



Confidence interval for σ^2

By choosing points that cut off equal tail areas and reordering the inequality in the probability statement, we obtain

$$P \left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} \right] = 1 - \alpha.$$

Confidence interval for σ^2

A $(1 - \alpha)100\%$ confidence interval for the population variance σ^2 (where the population is assumed Normal) is given by:

$$\left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} \right].$$

Where $\chi_{\alpha/2}^2$ is the value of the chi-square distribution with $n - 1$ degrees of freedom that cuts off an area of $\alpha/2$ to its right and $\chi_{1-\alpha/2}^2$ is the value of the distribution that cuts off an area of $\alpha/2$ to its left.

Example

In an automated process, a machine fills cans of coffee. If the average amount filled is different from what it should be, the machine may be adjusted to correct the mean. If the **variance** of the filling process is too high, however, the machine is out of control and needs to be repaired. Therefore, from time to time regular checks of the variance of the filling process are made. This is done by randomly sampling filled cans, measuring their amounts, and computing the sample variance. A random sample of 30 cans gives an estimate $S^2 = 18.540$. Give a 95% confidence interval for the population variance σ^2 .

From our table we get, for $df = 29$, $\chi_{0.025}^2 = 45.7222$ and $\chi_{0.975}^2 = 16.0471$. Using those values, we compute the confidence interval as follows:

$$\left[\frac{29(18.540)}{45.7222}, \frac{29(18.540)}{16.0471} \right] \approx [11.7593, 33.5051]$$

We can be 95% confident that the population variance is between 11.7593 and 33.5051.

Consistency of a container-filling machine

Container-filling machines are used to package a variety of liquids, including milk, soft drinks, and paint. Ideally, the amount of liquid should vary only slightly because large variations will cause some containers to be underfilled (cheating the customer) and some to be overfilled (resulting in costly waste). The president of a company that developed a new type of machine boasts that this machine can fill 1-liter (1,000 cubic centimeters) containers so consistently that the variance of the fills will be less than 1 cubic centimeter². To examine the veracity of the claim, a random sample of 25 1-liter fills was taken and the results (cubic centimeters) recorded (See fills.csv).

Estimate with 99% confidence the variance of fills.

Reading our data

```
# Step 1. Reading data;  
  
fills_file = read.csv(file="fills.csv",header=TRUE)  
  
names(fills_file);  
  
## [1] "Fills"  
  
fills = fills_file$Fills;
```

Finding CI for σ^2

```
# Step 2. Construction of CI;  
  
n= length(fillls)  
  
df = n-1;  
  
alpha = 0.01;  
  
xL = qchisq(alpha/2,df);  
  
xU = qchisq(1 - (alpha/2),df);
```

Finding CI for σ^2

```
# Step 2. Construction of CI (cont);  
  
#LCL = Lower Confidence Limit  
  
LCL= df*var(fillS)/xU;  
  
# UCL = Upper Confidence Limit  
  
UCL =df*var(fillS)/xL;  
  
c(LCL, UCL);  
  
## [1] 0.3336369 1.5374915
```

Here we see that σ^2 is estimated to lie between 0.3336369 and 1.5374915. Part of this interval is above 1, which tells us that the variance may be larger than 1.

Example

- 1 Estimate σ^2 with 90% confidence given that $n = 15$ and $S^2 = 12$.
- 2 Repeat part 1) with $n = 30$
- 3 What is the effect of increasing the sample size?

Solution

From our table we get, for $df = 14$, $\chi_{0.95}^2 = 6.57063$ and $\chi_{0.05}^2 = 23.6848$. Using those values, we compute the confidence interval as follows:

$$\left[\frac{14(12)}{23.6848}, \frac{14(12)}{6.57063} \right] \approx [7.0932, 25.5683]$$

We can be 90% confident that the population variance is between 7.0932 and 25.5683.

TESTING HYPOTHESES CONCERNING VARIANCES

Test of Hypotheses concerning a Population Variance

Assumptions: Y_1, Y_2, \dots, Y_n constitute a random sample from a Normal distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$.

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2 & \text{upper-tailed alternative} \\ \sigma^2 < \sigma_0^2 & \text{lower-tailed alternative} \\ \sigma^2 \neq \sigma_0^2 & \text{two-tailed alternative} \end{cases}$$

Test of Hypotheses concerning a Population Variance

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

RejectionRegion : $\begin{cases} \chi^2 > \chi_\alpha^2 & \text{upper-tailed RR} \\ \chi^2 < \chi_\alpha^2 & \text{lower-tailed RR} \\ \chi^2 > \chi_\alpha^2 \text{ or } \chi^2 < \chi_\alpha^2 & \text{two-tailed RR} \end{cases}$

Notice that χ_α^2 is chosen so that, for $\nu = n - 1$ df, $P(\chi^2 > \chi_\alpha^2) = \alpha$.

Example

A manufacturer of car batteries claims that the life of his batteries is approximately Normally distributed with a standard deviation equal to 0.9 year. If a random sample of 10 of these batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ year? Use a 0.05 level of significance.

Step 1. State hypotheses.

$$H_0 : \sigma^2 = 0.81$$

$$H_a : \sigma^2 > 0.81$$

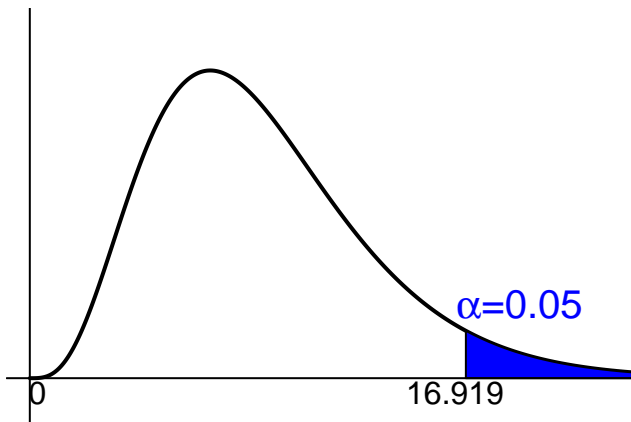
Step 2. Compute test statistic.

$S^2 = 1.44$, $n = 10$, and

$$\chi^2 = \frac{(9)(1.44)}{0.81} = 16$$

Step 3. Find Rejection Region.

From Figure and our table we see that the null hypothesis is rejected when $\chi^2 > 16.919$, where $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ with $\nu = 9$ degrees of freedom.



Step 4. Conclusion.

The χ^2 statistic is not significant at the 0.05 level. We conclude that there is insufficient evidence to claim that $\sigma > 0.9$ year.

Test of Hypotheses concerning Two Population Variances

Assumptions: Independent samples from a Normal populations.

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

$$H_a : \sigma_1^2 > \sigma_2^2$$

Test of Hypotheses concerning Two Population Variances

Test statistic: $F = \frac{S_1^2}{S_2^2}$

Rejection region: $F > F_\alpha$, where F_α is chosen so that $P(F > F_\alpha) = \alpha$ when F has $\nu_1 = n_1 - 1$ numerator degrees of freedom and $\nu_2 = n_2 - 1$ denominator degrees of freedom.

Example

One of the problems that insider trading supposedly causes is unnaturally high stock price volatility. When insiders rush to buy a stock they believe will increase in price, the buying pressure causes the stock price to rise faster than under usual conditions. Then, when insiders dump their holdings to realize quick gains, the stock price dips fast. Price volatility can be measured as the variance of prices.

Example (cont.)

An economist wants to study the effect of the insider trading scandal and ensuing legislation on the volatility of the price of a certain stock. The economist collects price data for the stock during the period before the event (interception and prosecution of insider traders) and after the event. The economist makes the assumptions that prices are approximately Normally distributed and that the two price data sets may be considered independent random samples from the populations of prices before and after the event.

Example (cont.)

Suppose that the economist wants to test whether or not the event has decreased the variance of prices of the stock. The 25 daily stock prices before the event give $S_1^2 = 9.3$ (dollars squared) and the 24 stock prices after the event give $S_2^2 = 3.0$ (dollars squared). Conduct the test at the $\alpha = 0.05$.

Step 1. State hypotheses.

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

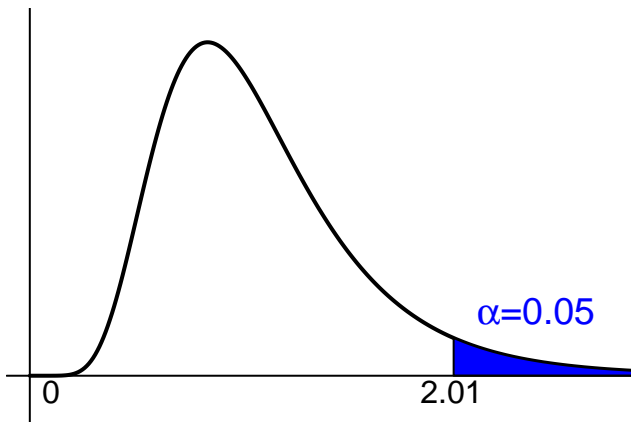
$$H_a : \sigma_1^2 > \sigma_2^2$$

Step 2. Compute test statistic.

$$F_{(n_1-1, n_2-1)} = F_{(24, 23)} = \frac{S_1^2}{S_2^2} = \frac{9.3}{3.0} = 3.1$$

Step 3. Find Rejection Region.

The critical point for $\alpha = 0.05$, from our Table, is equal to 2.01 (see 24 degrees of freedom for the numerator and 23 degrees of freedom for the denominator).



Step 4. Conclusion.

As can be seen from our Figure, this value of the test statistic ($F_{24, 23} = 3.1$) falls in the rejection region for $\alpha = 0.05$.

The economist may conclude (subject to the validity of the assumptions) the data present significant evidence that the event in question has reduced the variance of the stock's price.

Testing Population Variances

The hypotheses to be tested are

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

The test statistic is the ratio of the sample variances $\frac{S_1^2}{S_2^2}$, which is F-distributed with degrees of freedom $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. The required condition is the same as that for the t-test of $\mu_1 - \mu_2$, which is that both populations are Normally distributed.

This is a two-tail test so that the rejection region is

$$F > F_{\alpha/2, \nu_1, \nu_2} \text{ OR } F < F_{1-\alpha/2, \nu_1, \nu_2}$$

Example: Direct and Broker-Purchased Mutual Funds

Millions of investors buy mutual funds, choosing from thousands of possibilities. Some funds can be purchased directly from banks or other financial institutions whereas others must be purchased through brokers, who charge a fee for this service. This raises the question, Can investors do better by buying mutual funds directly than by purchasing mutual funds through brokers? To help answer this question, a group of researchers randomly sampled the annual returns from mutual funds that can be acquired directly and mutual funds that are bought through brokers and recorded the net annual returns, which are the returns on investment after deducting all relevant fees.

Example: Direct and Broker-Purchased Mutual Funds (cont.)

From the data, the following statistics were calculated:

$$n_1 = 50$$

$$n_2 = 50$$

$$\bar{x}_1 = 6.63$$

$$\bar{x}_2 = 3.72$$

$$s_1^2 = 37.49$$

$$s_2^2 = 43.34$$

Can we conclude at the 5% significance level that directly purchased mutual funds outperform mutual funds bought through brokers?

The hypothesis to be tested is that the mean net annual return from directly purchased mutual funds (μ_1) is larger than the mean of broker-purchased funds (μ_2). To decide which of the t-tests of $\mu_1 - \mu_2$ to apply, we conduct the F-test of $\frac{\sigma_1^2}{\sigma_2^2}$.

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

Test statistic:

$$F = \frac{S_1^2}{S_2^2} = \frac{37.49}{43.34} = 0.86$$

Rejection Region:

$$F > F_{\alpha/2, \nu_1, \nu_2} = F_{0.025, 49, 49} \approx F_{0.025, 50, 50} = 1.75$$

OR

$$F < F_{1-\alpha/2, \nu_1, \nu_2} = F_{0.975, 49, 49} = 1/F_{0.025, 49, 49} \\ \approx 1/F_{0.025, 50, 50} = 1/1.75 = 0.57$$

(See page 536 for more details about this "trick")

Conclusion:

Because $F = 0.86$ is not greater than 1.75 or smaller than 0.57, we cannot reject the null hypothesis. There is not enough evidence to infer that the population variances differ. It follows that we must apply the equal-variances t-test of $\mu_1 - \mu_2$.