Inference on Population Variances

553.211 Intro to Prob. & Stats.

Johns Hopkins University

EN.553.211

Why are we interested in population variance

When we study the quantitative data, for the past few weeks we've learned theory on estimating and testing hypothesis on the population mean, μ. However, it may be the case that what we're actually interested in is the population variance σ^2. For example:

- An altimeter on an aircraft might give a pilot the correct altitude *on average*... but if that measurement could vary by as much as ± 1000 ft. above/below the correct altitude, then you probably wouldn't be on that plane.
- Most exams must be designed in a way in which scores *will* exhibit a reasonable amount of variability. E.g. SAT math scores would not very useful for college admissions if everyone gets close to a perfect score.

χ^2 distribution

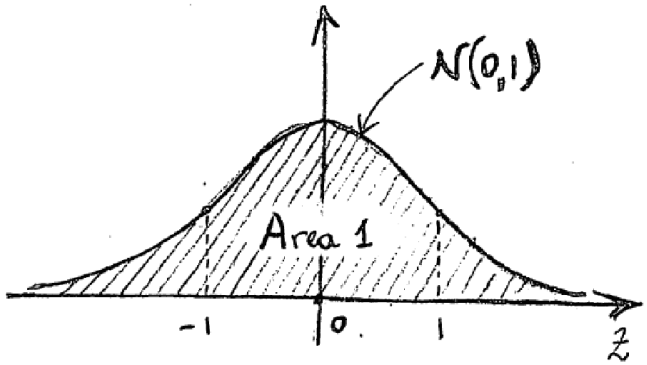
To estimate or do hypothesis testing on the population variance, we need to study the sampling distribution of its estimator, the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

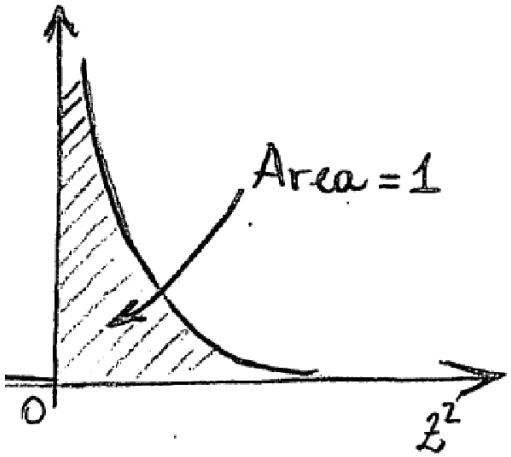
Let's first study a new family of distributions, the χ^2 (chi-squared) distribution, to prepare for it.

χ^2 distribution

- Suppose we have a standard normal random variable $Z \sim N(0, 1)$.

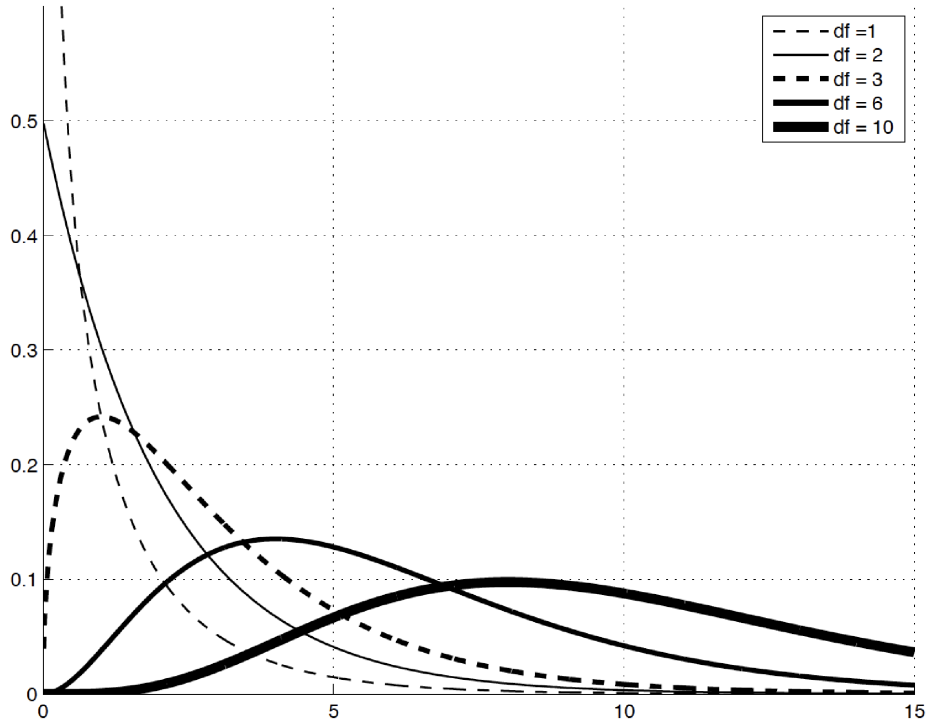


- Now let's square it. It can be shown that Z^2 follows χ^2 **(chi-squared) distribution with** $df = 1$ or we simply write $Z^2 \sim \chi^2_1$. Notice $Z^2 \geq 0$, so $P(Z^2 \geq 0) = 1$.



χ^2 distribution

- Now let $Z_1, Z_2, \ldots, Z_n$ are n independent R.V. following $N(0,1)$. Then the sum of their squares $Z_1^2 + Z_2^2 + \cdots + Z_n^2$ follows χ^2 distribution with $df = n$. We can write $Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$.

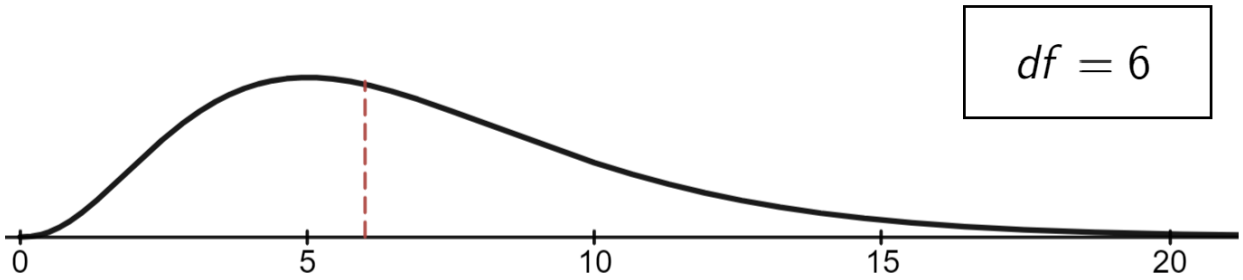


Properties of χ^2 distribution

- χ^2_{df} is not symmetric.
- The range of χ^2_{df} is $[0, \infty)$.
- The expectation is the degrees of freedom:

$$E(\chi^2_{df}) = df.$$

E.g.



- The variance is 2×the degrees of freedom:

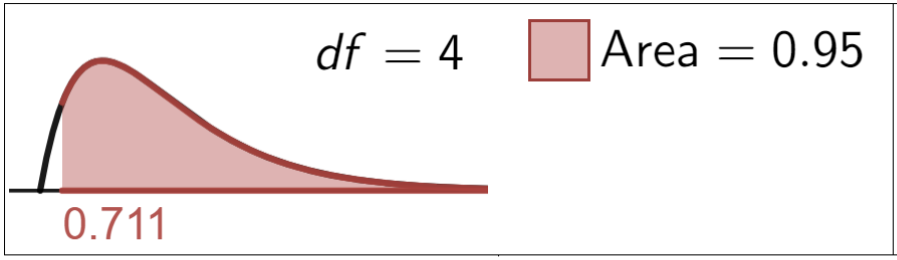
$$Var(\chi^2_{df}) = 2df.$$

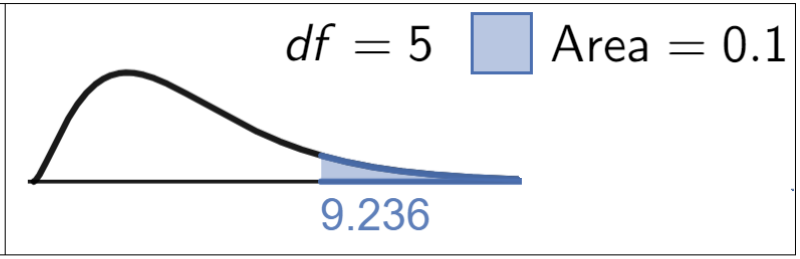
χ^2 table

Suppose the R.V. χ^2 follows χ^2_{df} distribution. We define $\chi^2_{df,\alpha}$ such that

$$P(\chi^2 > \chi^2_{df,\alpha}) = \alpha$$

We have the χ^2 table to find $\chi^2_{df,\alpha}$:





df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

The loss of a degree of freedom

Theorem

Let $Z_1, \ldots, Z_n$ be independent standard normal RV's. Then

$$\sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2 \sim \chi^2_{n-1}.$$

So, using $\bar{Z}$ to estimate $\mu = 0$ causes us to lose a *df*.

Sampling distribution of $\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2$

Recall that the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Theorem

If $X_1, \ldots, X_n$ are independent and follow $\mathcal{N}(\mu, \sigma)$ with sample variance S^2.
Then we have that $\frac{n-1}{\sigma^2} \cdot S^2$ is a random variable following χ^2_{n-1} (where n is the sample size). We write

$$\chi^2 := \frac{n-1}{\sigma^2} \cdot S^2 \sim \chi^2_{n-1}$$

Proof

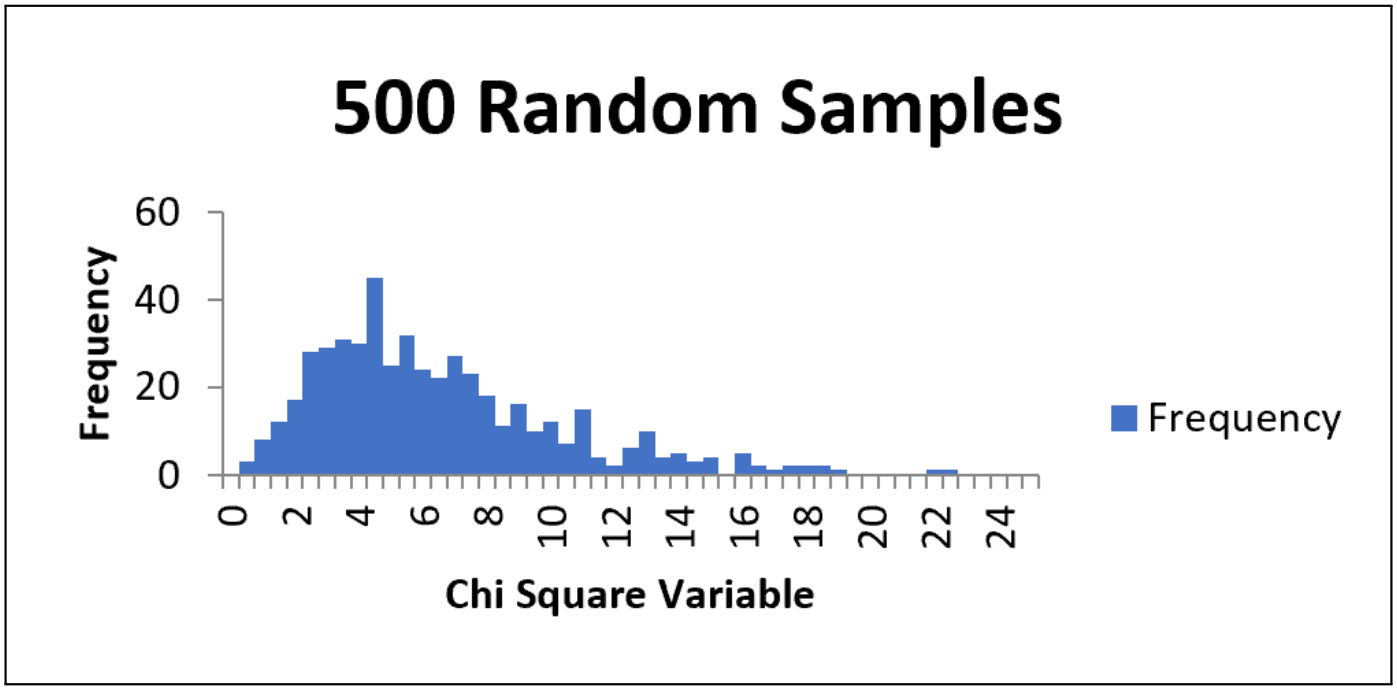
$$\frac{n-1}{\sigma^2} \cdot S^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{X_i - \mu + \mu - \bar{X}}{\sigma} \right)^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} - \frac{\bar{X} - \mu}{\sigma} \right)^2.$$

Now, the result follows since $\frac{X_i - \mu}{\sigma} \sim Z \sim \mathcal{N}(0,1)$ and $\frac{\bar{X} - \mu}{\sigma} \sim \bar{Z} \sim \mathcal{N}\left(0, 1/\sqrt{n}\right)$.

Simulation for $\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2$

In the Excel file `Chi_Square_Simulation`, we take 500 random samples ($n = 7$) from a $\mathcal{N}(0, 1)$ distributed population and plot the histogram of $\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2$:



χ^2 follows χ^2_6

Sampling distribution of $\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2$

If we know the population distribution is approximately normal, we can apply the above theorem to have the following sampling distribution of $\chi^2_{n-1} = \frac{n-1}{\sigma^2} \cdot S^2$.

Theorem (Sampling distribution of $\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2$)

If the population is approximately normal.
Then we have

$$\chi^2 = \frac{n-1}{\sigma^2} \cdot S^2 \sim \chi^2_{n-1}$$

Note: when we study the population variance, if the question doesn't mention the population distribution, we always suppose it is approximately normal.

Confidence interval of σ^2

Theorem (Confidence interval of σ^2)

Suppose we get a random sample of size n with sample variance s^2

- *If the population distribution is approximately normal*

We have with $(1-\alpha)\times 100\%$ confidence that

$$\sigma^2 \in \mathrm{CI}_{(1-\alpha)\times 100\%} = \left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

and

$$\sigma \in \mathrm{CI}_{(1-\alpha)\times 100\%} = \left(\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}} \right)$$

Note: Unlike the normal and *t* distributions, the χ^2 distribution is *not* symmetric, therefore we need to get the critical values for the two tails, respectively, from the χ^2 table.

Example 1

Example (A χ^2 CI example)

Suppose the population distribution is approximately normal. An experimenter recorded the measurements 4.1, 5.2, 6.7 and 10.2. Let's construct a 90% confidence interval to estimate the population variance.

Solution

The population distribution is approximately normal. Thus we have the above formula for CI

We can compute $s^2 = 7.057$.

With $df = n - 1 = 4 - 1 = 3$ and $\alpha = 0.1$, the critical values are

$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{3,0.95} = 0.352$$

and

$$\chi^2_{n-1,\alpha/2} = \chi^2_{3,0.05} = 7.815$$

So the 90% confidence interval for σ^2 is

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right] = \left[\frac{3 \cdot 7.057}{7.815}, \frac{3 \cdot 7.057}{0.352}\right] = [2.709, 60.145]$$

Hypothesis testing for population variance

Hypothesis testing for σ^2

Null Hypothesis

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

Choice of Alternative Hypothesis:

- $H_a : \sigma^2 \neq \sigma_0^2$ (two-sided)
- $H_a : \sigma^2 < \sigma_0^2$ (one-sided)
- $H_a : \sigma^2 > \sigma_0^2$ (one-sided)

Hypothesis testing for population variance

Hypothesis testing for σ^2

Suppose we get a random sample of size n with sample variance s^2 to do hypothesis testing for σ^2

- If the population distribution is approximately normal

Then the null distribution for the test statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

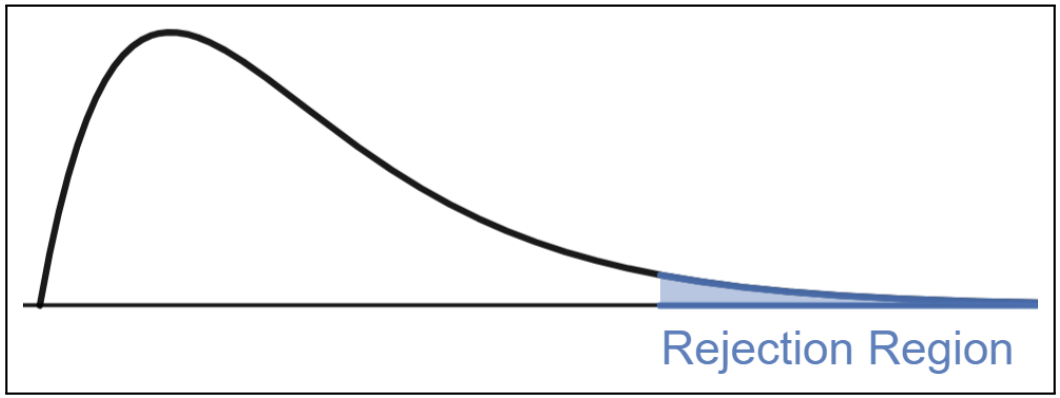
We compute the test statistic for the given sample as

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

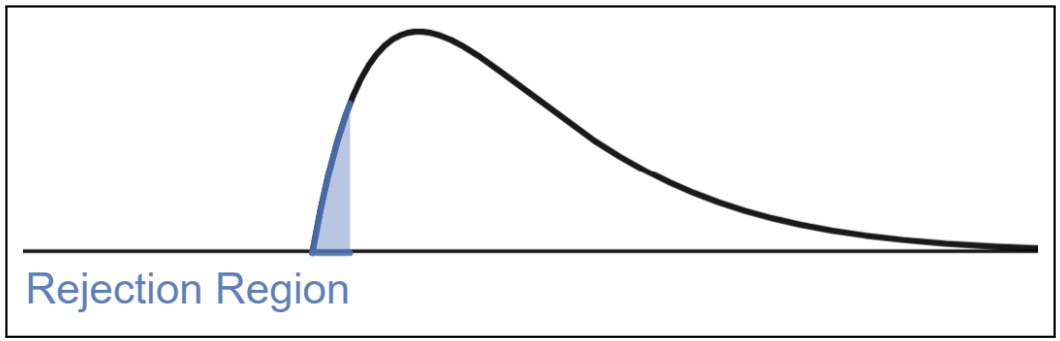
Note: We cannot get p-value from the χ^2 table. We always get the rejection region to do the test.

Rejection region (one tailed test)

If $H_a : \sigma^2 > \sigma_0^2$, $R_0 = [\chi^2_{n-1,\alpha}, \infty)$

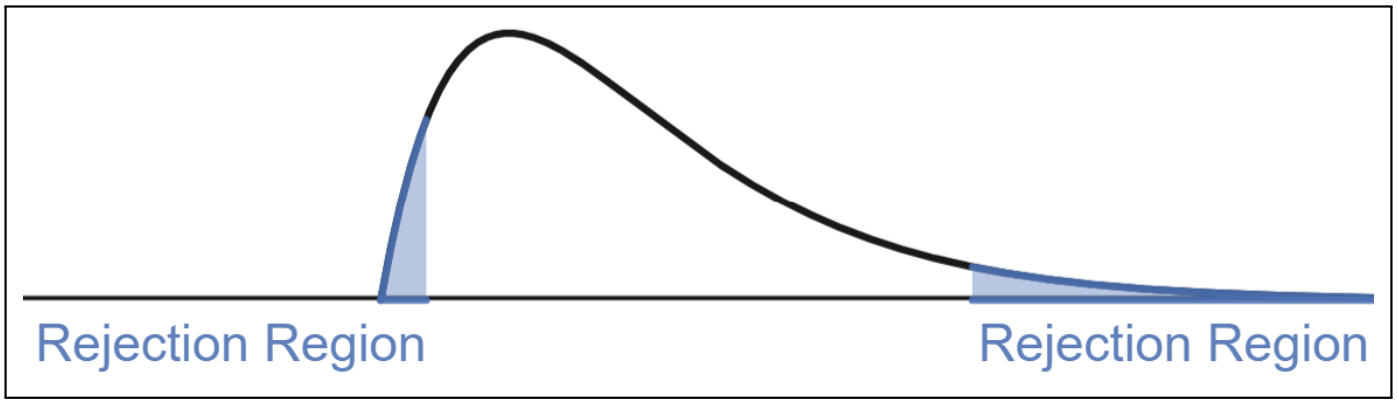


If $H_a : \sigma^2 < \sigma_0^2$, $R_0 = (-\infty, \chi^2_{n-1,1-\alpha}]$



Reject region (two tailed test)

If $H_a : \sigma^2 \neq \sigma_0$, $R_0 = (-\infty, \chi^2_{n-1,1-\alpha/2}] \cup [\chi^2_{n-1,\alpha/2}, \infty)$



Example 2

Example (Concrete)

A cement manufacturer claims that concrete prepared from his product has a relatively stable compressive strength and that the strength lies within a range of 40 kg/cm^2.
A sample of $n = 10$ measurements produces a sample variance $s^2 = 195$.
Does the data present evidence to show that the manufacturer's claim is *not* correct at significance level $\alpha = 0.05$? (Suppose that range $\approx 4\sigma$)

Solution

1. Establish null and alternative hypothesis
The claim is about the range, we are not given the claim about σ^2, however.

As before, we have an approximate relationship between the range and the standard deviation, range $\approx 4\sigma$. For range$= 40$, the corresponding $\sigma = 40/4 = 10$ and then $\sigma^2 = 10^2 = 100$.

With this relationship, the hypothesis we'll test is then

- $H_0 : \sigma^2 = 100$
- $H_a : \sigma^2 > 100$; the manufacturer's claim is *not* correct.

The null value $\sigma_0^2 = 100$

Solution

2. Check the conditions and specify the null distribution of the test statistic

When we study the sample variance, we always suppose the population distribution is approximately normal. Thus

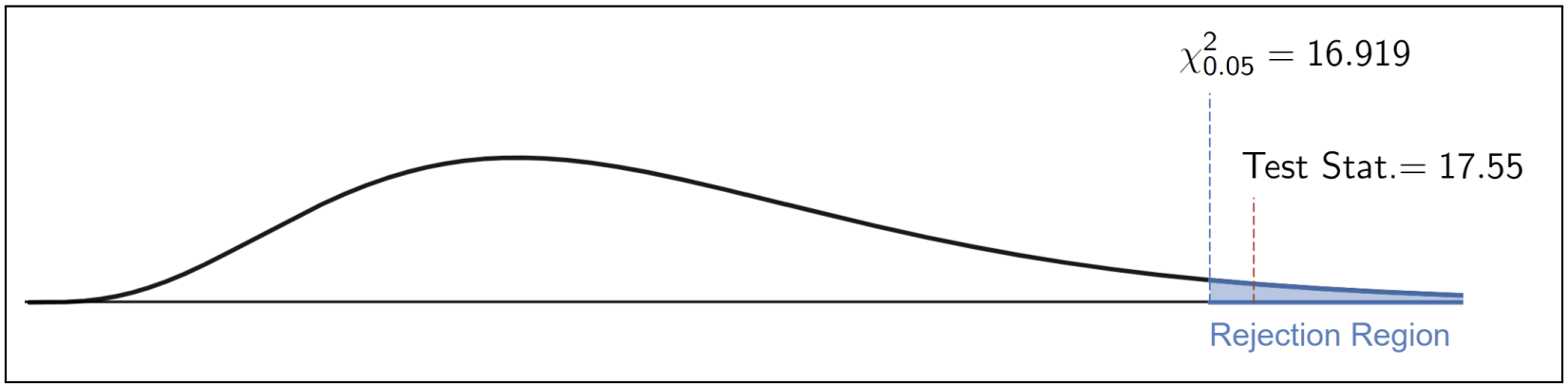
$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

3. Compute the test statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \cdot 195}{100} = 17.55.$$

Solution

4. Find rejection region R_0 with respect to the direction of H_a
The critical $\chi^2_{n-1,\alpha}$ with $df = n - 1 = 9$ and $\alpha = 0.05$ is $\chi^2_{9,0.05} = 16.919$.



$$R_0 = [\chi^2_{n-1,\alpha}, \infty) = [\chi^2_{9,0.05}, \infty) = [16.919, \infty)$$

5. Decide whether the test statistic belongs to the rejection region $\chi^2 = 17.55 \in R_0$. Thus we reject H_0

Solution

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.05$, and we have evidence to conclude that the manufacturer's claim is not correct.

Compare two population variances

- Sometimes we need to compare two population variances.
- When we study the relationship between σ_1^2 and σ_2^2, we always study the ratio $\frac{\sigma_1^2}{\sigma_2^2}$.

$$\sigma_1^2 \neq \sigma_2^2 \iff \frac{\sigma_1^2}{\sigma_2^2} \neq 1, \quad \sigma_1^2 > \sigma_2^2 \iff \frac{\sigma_1^2}{\sigma_2^2} > 1, \quad \ldots$$

- The point estimator of $\frac{\sigma_1^2}{\sigma_2^2}$ is $\frac{S_1^2}{S_2^2}$.
- We need to know the sampling distribution of $\frac{S_1^2}{S_2^2}$ for the inference on $\frac{\sigma_1^2}{\sigma_2^2}$.
- Let's first study a new family of distributions, the F distribution, to prepare for it.

F distribution

Definition

For two independent R.V.s $X \sim \chi^2_{df_1}$ and $Y \sim \chi^2_{df_2}$.

The ratio $\frac{X/df_1}{Y/df_2}$ follows **F distribution** with df_1 as **degrees of freedom in the numerator** and df_2 as **degrees of freedom in the denominator**. We write

$$\frac{X/df_1}{Y/df_2} \sim F_{df_1, df_2}$$

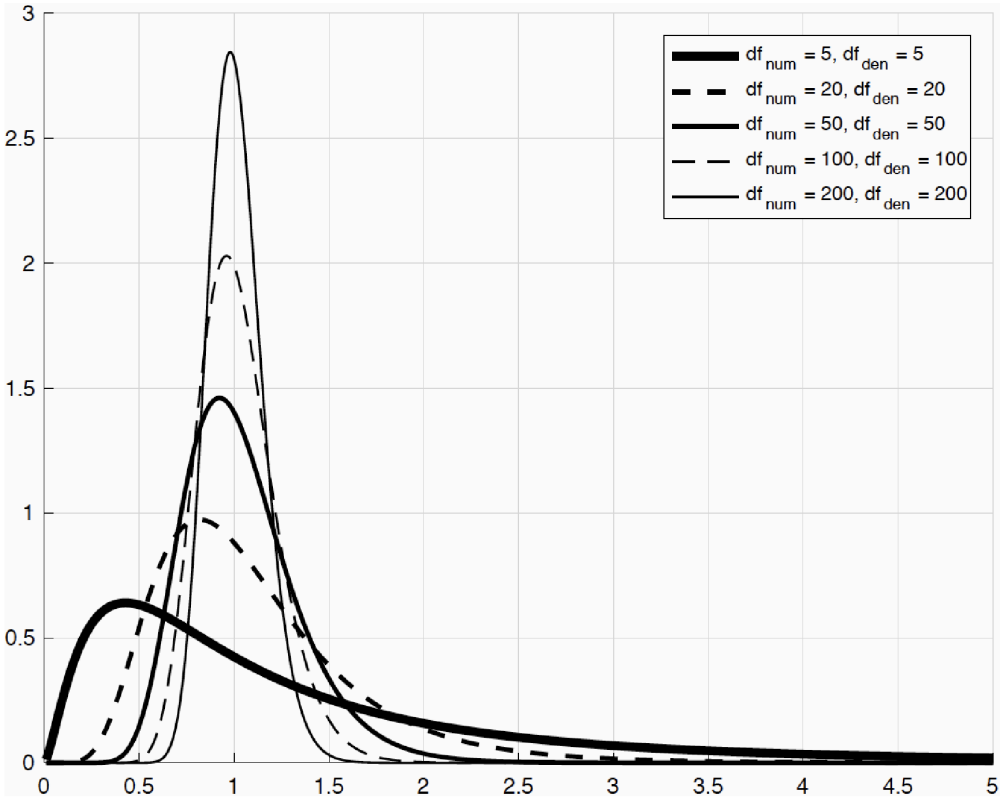
Fact

If S_1^2 and S_2^2 are the sample variances from two independent normal populations with sample sizes n_1, n_2 and variances σ_1^2, σ_2^2, then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}.$$

Properties of F distribution

- The randomness from the $F = \frac{X/df_1}{Y/df_2}$ R.V. comes from the randomness of *both* X and Y.
- F_{df_1,df_2} is not symmetric.
- The pdf of the F_{df_1,df_2}-distribution is shaped like the χ^2 distribution:



Properties of F distribution

- When $df_2 \to \infty$, we have $Y/df_2 \to 1$, then $F \to X/df_1$.

  Recall that for $Y \sim \chi^2_{df_2}$, we have $E[Y] = df_2$ and $V[Y] = 2df_2$. Thus $E[Y/df_2] = 1$ and $V[Y/df_2] = 2/df_2$. When $df_2 \to \infty$, the variability of $Y/df_2 = 2/df_2 \to 0$, and Y/df_2 is concentrated around the expectation 1. So we have $Y/df_2 \to 1$.

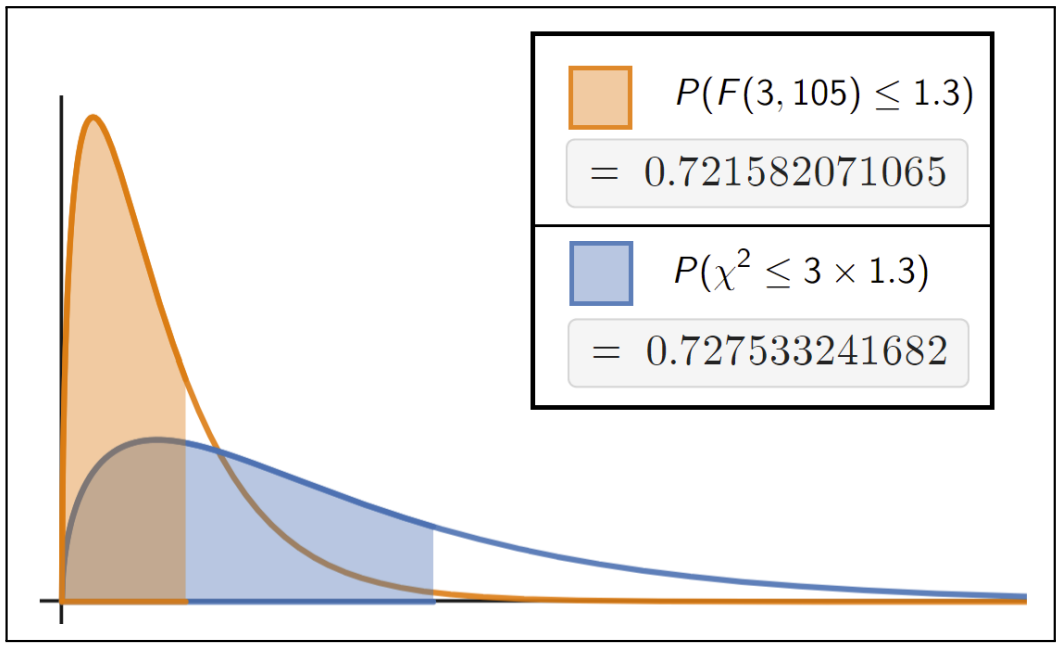
  With the identical analysis, we can prove if $n \to \infty$, we have $S^2 \to \sigma^2$ (Recall that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$). Intuitively, the large sample reduces the variability of the estimator and we get the precise estimator.

Properties of F distribution

- When df_2 is large enough, $F \approx X/df_1$. Then for any constant C, the cdf of F_{df_1,df_2} for C is approximately equal to the cdf of $\chi^2_{df_1}$ for $df_1 \cdot C$. Because

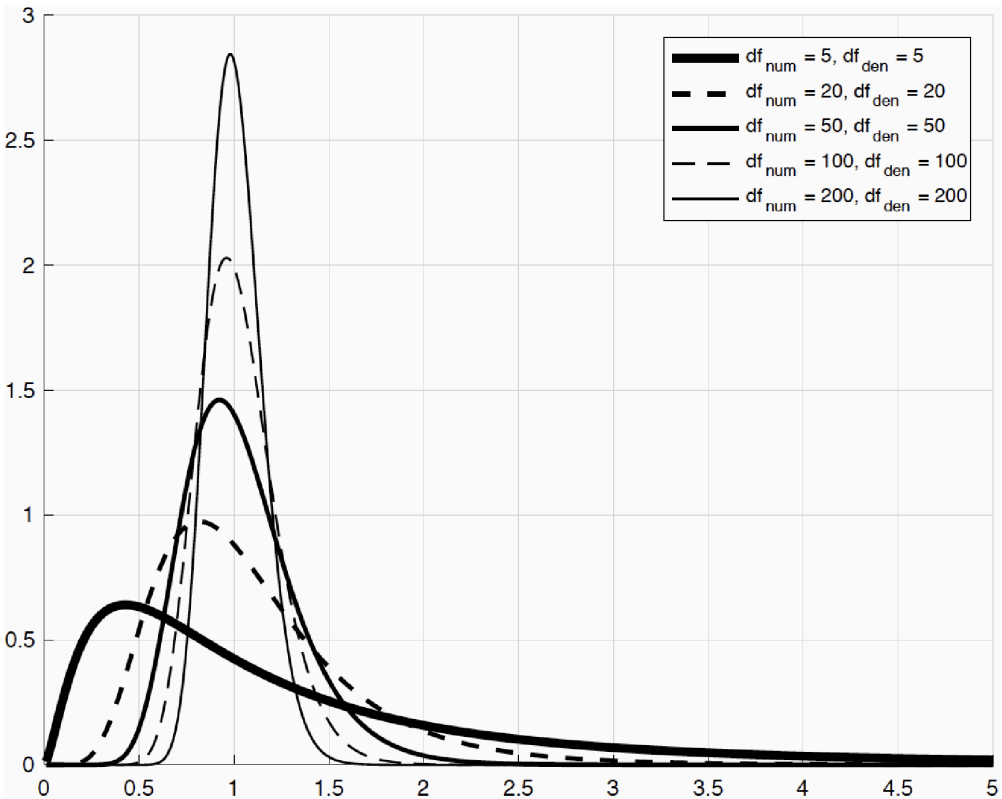
$$P(F_{df_1,df_2} \leq C) \approx P(\chi^2_{df_1}/df_1 \leq C) = P(\chi^2_{df_1} \leq df_1 \cdot C)$$

E.g.



Properties of F distribution

- If df_1 and df_2 are both large, we have $X/df_1 \approx 1$ $Y/df_2 \approx 1$, then $F \approx 1$.
- This is why, for the large values of df_1 and df_2 the pdf of F distribution has a tall and narrow peak near $F = 1$:



F-Distribution tables

Suppose the R.V. F follows F_{df_1,df_2} distribution. We define $F_{df_1,df_2,\alpha}$ such that

$$P(F > F_{df_1,df_2,\alpha}) = \alpha$$

We have the F distribution table to find $F_{df_1,df_2,\alpha}$:

F-Distribution Table

Numerator Degrees of Freedom		Alpha	Denominator Degrees of Freedom 1	2	3	4	5	6
	1	0.005	16210.7227	198.5013	55.5520	31.3328	22.7848	18.6350
		0.01	4052.1807	98.5025	34.1162	21.1977	16.2582	13.7450
		0.02	1012.5452	48.5051	20.6180	14.0396	11.3228	9.8764
		0.025	647.7890	38.5063	17.4434	12.2179	10.0070	8.8131
		0.05	161.4476	18.5128	10.1280	7.7086	6.6079	5.9874
		0.1	39.8635	8.5263	5.5383	4.5448	4.0604	3.7759
	2	0.005	19999.5000	199.0000	49.7993	26.2843	18.3138	14.5441
		0.01	4999.5000	99.0000	30.8165	18.0000	13.2739	10.9248
		0.02	1249.5000	49.0000	18.8581	12.1421	9.4544	8.0521
		0.025	799.5000	39.0000	16.0441	10.6491	8.4336	7.2599
		0.05	199.5000	19.0000	9.5521	6.9443	5.7861	5.1433
		0.1	49.5000	9.0000	5.4624	4.3246	3.7797	3.4633
	3	0.005	21614.7414	199.1664	47.4672	24.2591	16.5298	12.9166
		0.01	5403.3520	99.1662	29.4567	16.6944	12.0600	9.7795
		0.02	1350.5047	49.1657	18.1097	11.3435	8.6702	7.2870
		0.025	864.1630	39.1655	15.4392	9.9792	7.7636	6.5988
		0.05	215.7073	19.1643	9.2766	6.5914	5.4095	4.7571
		0.1	53.5932	9.1618	5.3908	4.1909	3.6195	3.2888

E.g. $F_{3,6,0.05} \approx 4.7571$

Sampling distribution of $F = \frac{S_1^2}{S_2^2}$

Let $X := \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ and $Y := \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$. We have the following theorem.

Theorem (Sampling distribution of $F = \frac{S_1^2}{S_2^2}$)

We get two independent random samples from two populations with sample sizes n_1, n_2 and sample variances S_1^2, S_2^2.

- *If the two populations are approximately normal.*
- *And we know the populations have the equal variance, $\sigma_1^2 = \sigma_2^2$*

Then we have

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

Hypothesis testing for two population variances

Hypothesis testing for $\frac{\sigma_1^2}{\sigma_2^2}$

Null Hypothesis

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

Choice of Alternative Hypothesis:

- $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ (two-sided)
- $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$ (one-sided)
- $H_a : \frac{\sigma_1^2}{\sigma_2^2} > 1$ (one-sided)

Notice the null value must be 1. Only when $\sigma_1^2 = \sigma_2^2$ we have the following sampling distribution of F.

Hypothesis testing for two population variances

Hypothesis testing for $\frac{\sigma_1^2}{\sigma_2^2}$

If the two populations are approximately normal.
Then the null distribution for the test statistic

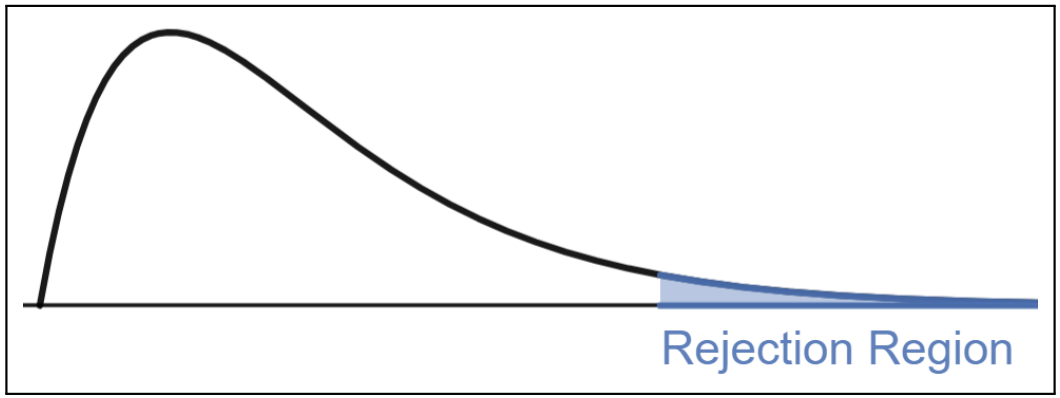
$$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$$

We compute the test statistic for the given sample as

$$F = \frac{s_1^2}{s_2^2} \sim F_{n_1-1,n_2-1}$$

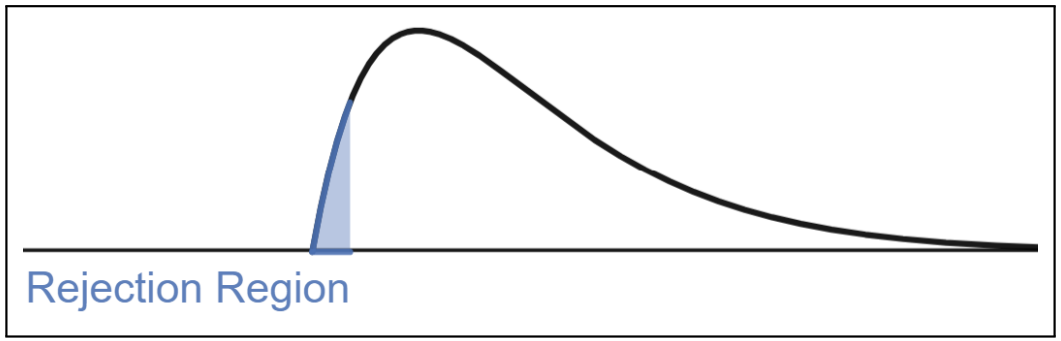
Rejection region for $H_a : \frac{\sigma_1^2}{\sigma_2^2} > 1$

If $H_a : \frac{\sigma_1^2}{\sigma_2^2} > 1$, $R_0 = [F_{n_1-1,n_2-1,\alpha}, \infty)$, i.e. if $F = \frac{s_1^2}{s_2^2} > F_{n_1-1,n_2-1,\alpha}$ we reject H_0



Rejection region for $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$

If $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$, $R_0 = (-\infty, F_{n_1-1,n_2-1,1-\alpha}]$

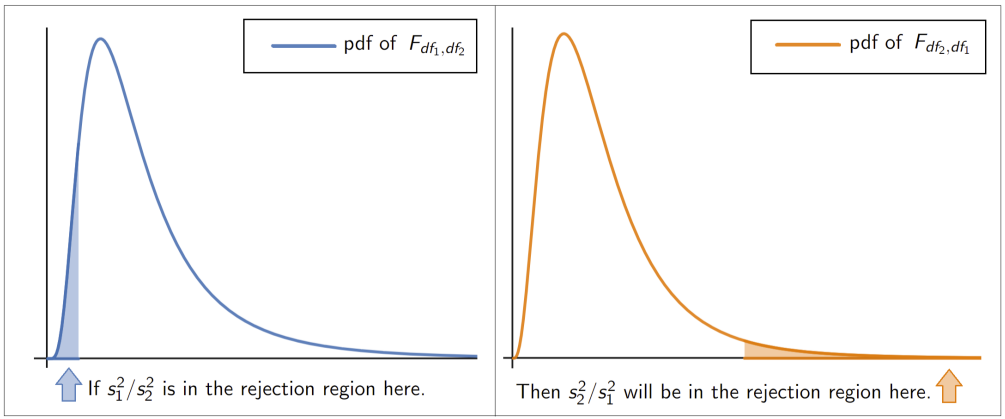


But notice that the F distribution table only lists values for small α ($\alpha = 0.1, 0.05, 0.025, 0.01$ and 0.005). It doesn't work for $F_{n_1-1,n_2-1,1-\alpha}$ here.

Rejection region for $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$

Fact

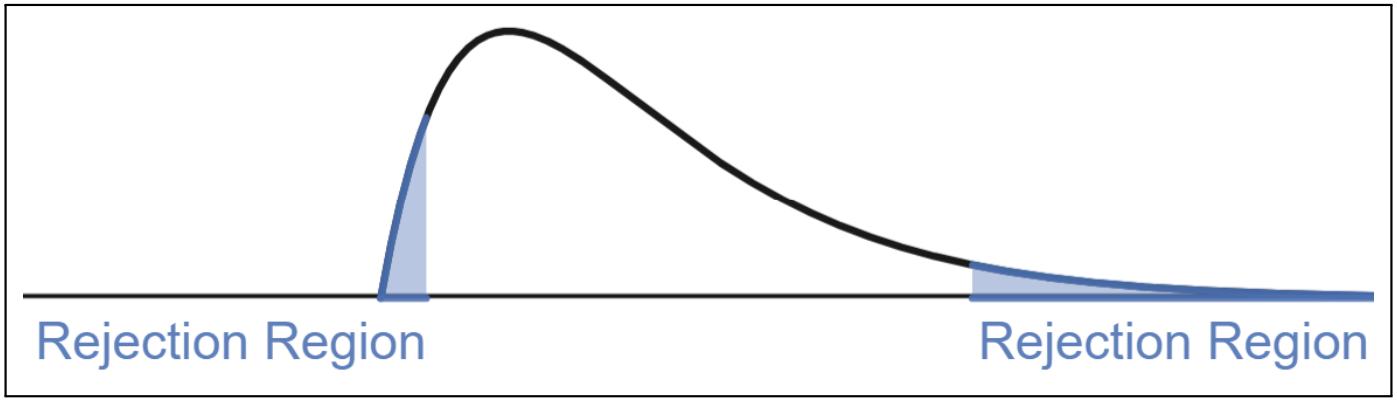
If $F = S_1^2/S_2^2$ has F distribution with df_1 degrees of freedom in the numerator and df_2 degrees of freedom in the denominator.
Then $F^{-1} = S_2^2/S_1^2$ has F distribution with df_2 degrees of freedom in the numerator and df_1 degrees of freedom in the denominator.



If $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$, we construct $R_0 = [F_{n_2-1,n_1-1,\alpha}, \infty)$ for $F^{-1} = \frac{s_2^2}{s_1^2}$,
i.e. if $F^{-1} = \frac{s_2^2}{s_1^2} > F_{n_2-1,n_1-1,\alpha}$ we reject H_0

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

For $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$, if $F = \frac{s_1^2}{s_2^2} > F_{n_1-1,n_2-1,\alpha/2}$ or $F^{-1} = \frac{s_2^2}{s_1^2} > F_{n_2-1,n_1-1,\alpha/2}$, we reject H_0.



Example 3

Example

An experimenter believes that the average levels of impurities are the same for two production lines but that her adjustment may have decreased the variability of the levels for line 1.
Samples of $n_1 = 20$ and $n_2 = 25$ measurements were taken from the two batches and yielded these variances

$$s_1^2 = 0.51 \quad \text{and} \quad s_2^2 = 1.10.$$

Does the data present sufficient evidence to indicate that the process variability is less for line 1? Use $\alpha = 0.05$.

Solution

1. Establish null and alternative hypothesis
   - $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$
   - $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$; the process variability is less for line 1.

2. Check the conditions and specify the null distribution of the test statistic

When we do inference on the population variances, we always suppose the population distribution is approximately normal. Then under H_0 we have

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

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

Solution

3. Compute the test statistic
Because $H_a : \frac{\sigma_1^2}{\sigma_2^2} < 1$, we compute F^{-1}

$$F^{-1} = \frac{s_2^2}{s_1^2} = \frac{1.10}{0.51} \approx 2.157.$$

4. Find rejection region R_0 with respect to the direction of H_a
We construct $R_0 = [F_{n_2-1,n_1-1,\alpha}, \infty)$ for F^{-1}.
$n_2 - 1 = 25 - 1 = 24, n_1 - 1 = 20 - 1 = 19.\alpha = 0.05$. Then from the F distribution table, we have

$$F_{n_2-1,n_1-1,\alpha} = F_{24,19,0.05} \approx 2.114$$

5. Decide whether the test statistic belongs to the rejection region
$F^{-1} \approx 2.157 > 2.114 \approx F_{n_2-1,n_1-1,\alpha}$ we reject H_0

Solution

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.05$, and we have evidence to conclude that the process variability is less for line 1.

Example 4

Example

An experimenter is concerned that the variabilities of responses using two different procedures may not be the same. Before conducting his research, he conducts a pre-study with random samples of 11 and 9 responses and gets $s_1^2 = 14.14$ and $s_2^2 = 3.21$, respectively. Do the sample variances present sufficient evidence to indicate that the population variances are unequal? Use $\alpha = 0.05$.

Solution

1. Establish null and alternative hypothesis
   - $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$
   - $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$; the variabilities of two different procedures are not be the same.

2. Check the conditions and specify the null distribution of the test statistic

When we do inference on the population variances, we always suppose the population distribution is approximately normal. Then under H_0 we have

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

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

Solution

3. Compute the test statistic

Because $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$, we compute F and F^{-1}

$$F = \frac{s_1^2}{s_2^2} = \frac{14.14}{3.21} \approx 4.40, \quad F^{-1} = 4.40^{-1} = 0.227$$

4. Find rejection region R_0 with respect to the direction of H_a

We find $F_{n_1-1,n_2-1,\alpha/2}$ for F, and find $F_{n_2-1,n_1-1,\alpha/2}$ for F^{-1}.

$n_1 - 1 = 11 - 1 = 10, n_2 - 1 = 9 - 1 = 8, \alpha/2 = 0.05/2 = 0.025.$

Then from the F distribution table, we have

$$F_{n_1-1,n_2-1,\alpha/2} = F_{10,8,0.025} \approx 4.295,$$
$$F_{n_2-1,n_1-1,\alpha/2} = F_{8,10,0.025} \approx 3.855$$

5. Decide whether the test statistic belongs to the rejection region

$F \approx 4.40 > 4.295 \approx F_{n_1-1,n_2-1,\alpha/2}$, we reject H_0

Solution

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.05$, and we have evidence to conclude that the variability between two different procedures are not be the same.

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

Remark

For $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$, *it is enough to only check the larger one of* F *and* F^{-1}.

Confidence interval for σ_1^2/σ_2^2

We might also want to find a $(1-\alpha)100\%$ confidence interval for σ_1^2/σ_2^2. Because

$$1-\alpha = P\left(F_{n_1-1,n_2-1,1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{F_2^2/\sigma_2^2} < F_{n_1-1,n_2-1,\alpha/2}\right)$$

We could do our usual "trick" of rearranging what's inside the probability to suggest the following formula for a $(1-\alpha)100\%$ CI:

$$\frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1,\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}}$$

However we would not be able to find the upper bound from the tables!

Confidence interval for σ_1^2/σ_2^2

We will use the same idea as before by taking the reciprocal

$$F_{n_1-1,n_2-1,1-\frac{\alpha}{2}} = \frac{1}{F_{n_2-1,n_1-1,\frac{\alpha}{2}}}.$$

Theorem

Let s_1^2 and s_2^2 be the observed sample variances from two independent normal populations using sample sizes n_1 and n_2, respectively. We have $(1-\alpha)\times 100\%$ confidence that

$$\frac{\sigma_1^2}{\sigma_2^2} \in \mathrm{CI}_{(1-\alpha)100\%} = \left(\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{n_1-1,n_2-1,\frac{\alpha}{2}}},\ \frac{s_1^2}{s_2^2} \cdot F_{n_2-1,n_1-1,\frac{\alpha}{2}} \right).$$

A $(1-\alpha)\times 100\%$ confidence interval for σ_1/σ_2 can be found by square rooting the endpoints of the the above interval.

Example 5

Example

The feeding habits of two-species of net-casting spiders are studied. The species, the *deinopis* and *menneus*, coexist in eastern Australia. The following summary statistics were obtained on the size, in millimeters, of the prey of the two species:

	adult deinopis	**adult menneus**
sample size	10	10
sample avg.	10.26 mm	9.02 mm
sample var.	$(2.51 \text{ mm})^2$	$(1.90 \text{ mm})^2$

Estimate, with 95% confidence, the ratio of the two population variances.

Solution

Let σ_1^2 and σ_2^2 be the population variance of the prey size of *deinopis* and *menneus* spiders respectively.
The critical points are

$$F_{9,9,0.025} = 4.03 \quad \text{and} \quad 1/F_{9,9,0.025} = \frac{1}{4.03}.$$

So we are 95% confident that the ratio of variance in prey sizes is

$$0.433 = \frac{1}{4.03} \cdot \frac{(2.51)^2}{(1.90)^2} < \frac{\sigma_1^2}{\sigma_2^2} < 4.03 \cdot \frac{(2.51)^2}{(1.90)^2} = 7.033.$$