Hypothesis test for the population variance

Fundamental to statistics is the distinction between a sample and a population. The population is described by parameters, such as the population mean and variance; the sample is characterised by statistics that are functions defined on the sample data, such as the sample mean and the sample variance. The whole need for statistics arises from the practical impossibility of surveying all the members of a population – that is, the practical impossibility of conducting a census. Hence, one whole branch of statistics is concerned with estimating population parameters from sample statistics.

Often we begin with a theory of what the population parameters are – for example, we might have a theory about the mean and variance of the population based on estimates drawn from samples taken ten years ago. But populations do not always remain the same – so there arises, for example, the need to test the hypothesis that the population parameters have not changed. The first and most obvious thing to test is the mean. One starts with a null hypothesis regarding the population mean, and needs to test whether the population mean has altered, which forms the alternative hypothesis. Of course, we could be testing whether the mean is different – giving rise to a two-tailed test; or we might have reason to suppose that the mean has definitely increased or definitely decreased – giving rise to a one-tailed test.

The same procedure applies to the variance. We need a process for testing a null hypothesis about the size of the variance against an alternative hypothesis, that might be either one-tailed or two-tailed.

Here we present the test in a "cookery-book" fashion. (We justify the results in a later unit.) The hypothesis test follows the usual pattern of finding, from the sample, a test statistic and then comparing it with the critical value drawn from a theoretical probability distribution that the background, parent population variance is expected to follow should the null hypothesis in fact be true.

In the case of the variance, the theoretical distribution that should be used is the chi-squared distribution.

Firstly, we form from the sample data the biased sample variance. It is usual to compute this from

$$s^2 = \frac{\sum x^2}{n} - \left(\overline{x}\right)^2$$

This quantity is expected to follow a chi squared distribution as follows

$$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1} \qquad (*)$$

If the sample has size *n* then the critical value is drawn from the chi-squared distribution for $n-1$.

Another way of putting this formula is

$$s^2 \sim \frac{\sigma^2}{n}\chi^2_{n-1}$$

So we find the critical value by finding the quantity

$$\chi^2_{critical} = \frac{\sigma^2}{n}\chi^2_{n-1}$$

and we compare this with the test statistic, s^2, drawn from the sample.

Note, sometimes the whole procedure is written in terms of the *unbiased* sample variance, S^2, where

$$S^2 = \frac{n}{n-1}s^2$$

In this case we expect

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

This is the same formula as the one given at (*) but for the unbiased sample variance, not the biased one. This could be confusing, so be careful to be clear about which sample statistic you are using, and hence how you go about forming the critical value to test it against.

The procedure is best shown through examples – firstly the one-tailed test, and then the two

Example 1 – One-tailed test

Suppose we have the following set of data for a sample size $n = 12$

16.1	16.2	15.3	16.6	16.0	15.6
16.3	15.7	16.8	16.1	16.9	16.6

Test at the 5% significance level the hypothesis that the population variance is greater than 0.1 against the null hypothesis that the population variance is equal to 0.1.

Solution

Firstly, we form the null and the alternative hypotheses

$$H_o : \sigma^2 = 0.1$$
$$H_1 : \sigma^2 > 0.1$$

We can see from the type of inequality that this is a one-tailed test

We have $\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

where s^2 is the biased estimate of the sample variance.

Here $n = 12,\ \sigma^2 = \frac{1}{10} = 0.1$

So $\frac{12s^2}{1/10} \sim \chi^2_{11}$

Alternatively this can be written

$$S^2 \sim \frac{\sigma^2}{n}\chi^2_{n-1}$$
$$S^2 \sim \frac{1}{120}\chi^2_{11}$$

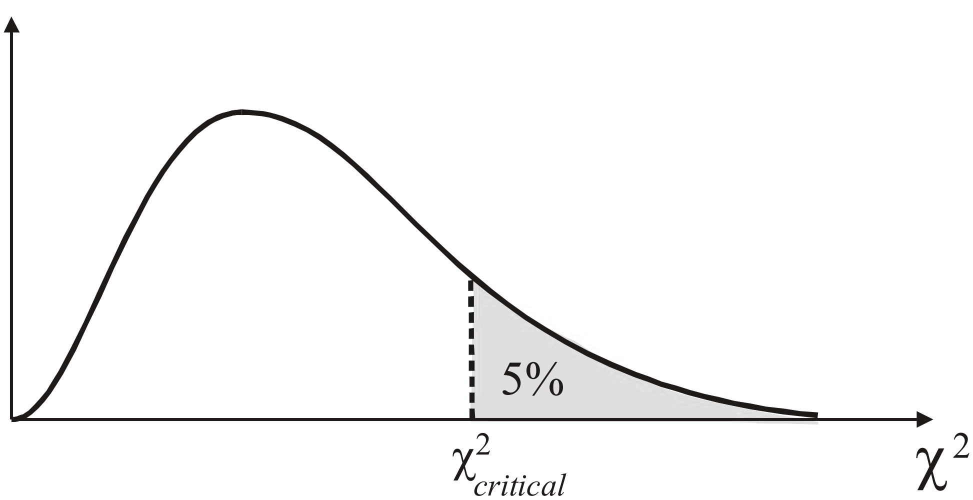
At the 5% significance level the critical value of χ^2 for 11 degrees of freedom is

$\chi^2_{11}(0.05) = 19.68$

which we obtain from the following table

$p\%$	99	97.5	95	90	10	5.0	2.5	1.0	0.5
$\nu = 1$	0.0001	0.0010	0.0039	0.0158	2.706	3.841	5.024	6.635	7.879
$\nu = 2$	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210	10.60
$\nu = 3$	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34	12.84
$\nu = 4$	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28	14.86
$\nu = 5$	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09	16.75
$\nu = 6$	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
$\nu = 7$	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
$\nu = 8$	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
$\nu = 9$	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
$\nu = 10$	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
$\nu = 11$	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
$\nu = 12$	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30

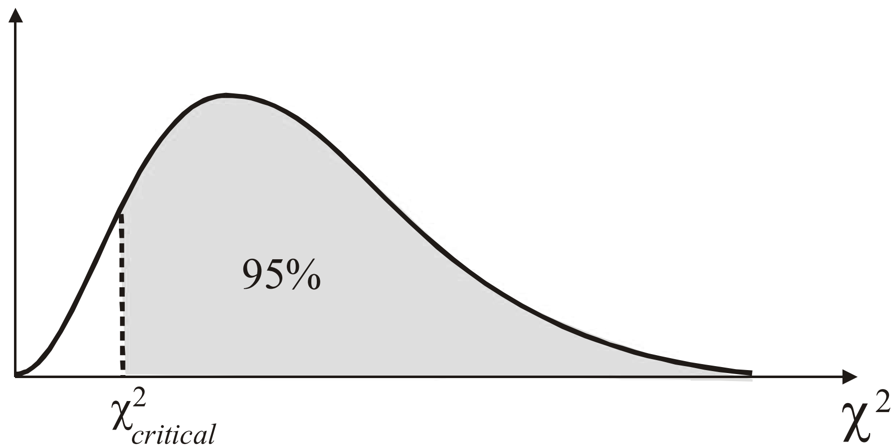
Let us just explain why we are looking in the 5% column and not in the 95% column. If the sample variance, σ, is greater than 0.1, then the quantity $\frac{\sigma^2}{n}\chi^2_{n-1}$ will be increased; so if s^2 is greater than the critical value it will be an indicator that the population variance, σ^2, is greater than the null hypothesis value.



If we were testing the alternative hypothesis

$H_1 : \sigma^2 < 0.1$

We would look for an s^2 value *less* than the critical value for the 95% column.



But, returning to our example, the critical value for s^2 is $\frac{19.68}{120} = 0.164$. Using a calculator (or otherwise)

$$s^2 = \frac{\sum x^2}{n} - (\overline{x})^2 = 0.2214(4.S.F.)$$

For our sample $s^2 = 0.221 > 0.164$

Therefore, we reject H_o and accept H_1. There is evidence to suggest that the population variance has increased.

A second examples illustrates the two-tailed test

Example 2 – Two-tailed test

The following data were collected in a sample of size $n = 12$

14.015	14.020	14.018	14.019	14.009
13.999	14.012	14.007	14.012	14.017
14.015	14.013	14.006	14.008	14.010

It was believed that the population variance was 0.1. Test this at the 5% significance level against the hypothesis that the population variance has changed.

<u>Solution</u>

$H_0 : \sigma^2 = 0.1$

$H_1 : \sigma^2 \neq 0.1$

Two tailed test

$\frac{ns^2}{\sigma} \sim \chi^2_{n-1}$ where s^2 is the biased estimate of the sample variance. Here n = 15 and $\sigma = \frac{1}{10} = 0.1$. Hence

$$\frac{15s^2}{1/10} \sim \chi^2_{14}$$

$$\therefore s^2 \sim \frac{1}{150}\chi^2_{14}$$

We are conducting a two-tailed test at the 5% significance level. Since the test is two-tailed we will reject H_0 if the test value of S^2 falls outside an interval marked by 2.5% of the χ^2_{14} probability distribution in the upper and lower tails.

From tables, the 97.5% critical value is 5.629 and the 2.5% critical value is 26.12.

Therefore, we accept H_o if S^2 lies between

$$\frac{5.629}{150} \leq s^2 \leq \frac{26.12}{150}$$

That is, $0.0175 \leq S^2 \leq 0.1741$

Now

$$s^2 = \frac{\Sigma X^2}{n} - \left(\overline{X}\right)^2$$
$$= \left(0.005489...\right)^2$$
$$= 0.00003013\left(4.S.F.\right)$$

This falls outside the acceptance region.

Therefore, we reject H_0, accept H_1; the population variance is no longer 0.1.