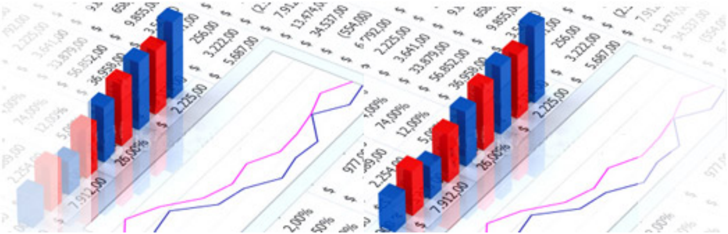
MTMS.01.099 Mathematical Statistics

Lecture 13. Testing hypotheses with traditional method, p-value method and confidence interval method

Tõnu Kollo



Fall 2016

χ^2 Test for a Variance or Standard Deviation

The χ^2 distribution was used to construct a confidence interval for a single variance or standard deviation. This distribution is also used to **test a claim about a single variance or standard deviation.**

There are several reasons why it is important to test the variance.

- In any situation where consistency is required, such as in manufacturing, you would like to have the smallest variation possible in the products.
- For example, when bolts are manufactured, the variation in diameters due to the process must be kept to a minimum, or the nuts will not fit them properly.

χ^2 Test for a Variance or Standard Deviation (2)

Three **assumptions** are made for the chi-square test:

1. The sample must be randomly selected from the population.
2. The population must be normally distributed for the variable under study.
3. The observations must be independent.

$Y \sim \chi^2(k)$ if $Y = \sum_{i=1}^{k} X_i^2$ where $X_i \sim N(0, 1)$.
Two important results for i.i.d. $X_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n$:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$$

when σ^2 is known.

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

when σ^2 is not known.

The value of the test statistic of the **χ^2-test:**

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

With significance level α the rejection regions are:

- $H_1 : \sigma^2 < \sigma_0^2 \quad \text{test} : \; \chi^2 < \chi^2_{1-\alpha,n-1} \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma^2 > \sigma_0^2 \quad \text{test} : \; \chi^2 > \chi^2_{\alpha,n-1} \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma^2 \neq \sigma_0^2 \quad \text{test} : \; \chi^2 < \chi^2_{1-\alpha/2,n-1} \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma^2 \neq \sigma_0^2 \quad \text{test} : \; \chi^2 > \chi^2_{\alpha/2,n-1} \xrightarrow{+} \text{reject} \;\; H_0$

Here $\chi^2_{\alpha,n-1}$ is the complement quantile of the chi-square distribution.

χ^2 Test for a Variance or Standard Deviation. Example

Example 1

An instructor wishes to see whether the variation in scores of the 23 students in her class is less than the variance of the population. The sample variance of the class is 198. Is there enough evidence to support the statement that the variation of the students is less than the population variance ($\sigma^2 = 225$) at $\alpha = 0.05$? Assume that the scores are normally distributed.

χ^2 Test for a Variance or Standard Deviation. Example

Example 1

An instructor wishes to see whether the variation in scores of the 23 students in her class is less than the variance of the population. The sample variance of the class is 198. Is there enough evidence to support the statement that the variation of the students is less than the population variance ($\sigma^2 = 225$) at $\alpha = 0.05$? Assume that the scores are normally distributed.

$$H_0 : \sigma^2 = 225 \text{ and } H_1 : \sigma^2 < 225$$

Since this test is left-tailed and $\alpha = 0.05$, use the value $1 - 0.05 = 0.95$. The degrees of freedom are $n - 1 = 23 - 1 = 22$. Hence, the critical value (complement α quantile) is 12.3.

χ^2 Test for a Variance or Standard Deviation. Example

Compute the test value:

$$\chi^2 = \frac{n-1}{\sigma^2} s^2 = \frac{23-1}{225} 198 = 19.36$$

Corresponding test is $\chi^2 < \chi^2_{1-\alpha, n-1} \xrightarrow{+} \text{reject } H_0$.
We have, $19.36 > 12.3$, e.g. the test value 19.36 falls in the non-critical region, thus the decision is to not reject the null hypothesis.
Summarize: There is not enough evidence to support the claim that the variation in test scores of the instructor's students is less than the variation in scores of the population.

Hypotheses about the difference of two Population variances

In addition to comparing two means, statisticians are interested in comparing two variances or standard deviations. For example, is the variation in the temperatures for a certain month for two cities different?
For the **comparison of two variances** or standard deviations, **F test** is used.
NB! Do not confuse F-test with the chi-square test, which compares a single sample variance to a specific population variance!

F-distribution

Definition

Let $X \sim \chi^2(k_1)$ and $Y \sim \chi^2(k_2)$ be independent. Then

$$Z = \frac{X/k_1}{Y/k_2}$$

is F-distributed with k_1 and k_2 degrees of freedom, $Z \sim F(k_1, k_2)$.

Let us we have a sample of size n_1 from $X \sim N(\mu_1, \sigma^2)$

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n_1 - 1)$$

and another independent sample of size n_2 from $Y \sim N(\mu_2, \sigma^2)$

$$\frac{(n_2 - 1)S_2^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sim \chi^2(n_2 - 1).$$

F-distribution

Then

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

Properties of the *F*-distribution:

1. The values of *F*-distribution cannot be negative, because variances are always positive or zero.
2. The distribution is positively skewed.
3. The mean of *F*-distribution is approximately equal to 1.
4. The *F*-distribution is a family of curves based on the degrees of freedom of the variance of the numerator and the degrees of freedom of the variance of the denominator.

Hypotheses about the difference of two Population variances (2)

The value of the test statistic of F-test is the ratio of sample variances:

$$F = \frac{s_1^2}{s_2^2},$$

where the larger of the two variances is placed in the numerator. The F-test has two terms for the degrees of freedom: that of the numerator, $n_1 - 1$, and that of the denominator, $n_2 - 1$, where n_1 is the sample size from which the larger variance was obtained.

Some remarks:

- If the standard deviations instead of the variances are given in the problem, they must be squared for the F test.
- When the degrees of freedom cannot be found from the F-distribution table, the closest value on the smaller side should be used.

Hypotheses about the difference of two Population variances (2)

With significance level α the rejection regions are:

- $H_1 : \sigma_1^2 < \sigma_2^2 \quad \text{test}: \; F < F_{1-\alpha}(n_1 - 1, n_2 - 1) \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma_1^2 > \sigma_2^2 \quad \text{test}: \; F > F_{\alpha}(n_1 - 1, n_2 - 1) \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{test}: \; F < F_{1-\alpha/2}(n_1 - 1, n_2 - 1) \xrightarrow{+} \text{reject} \;\; H_0$
- $H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{test}: \; F > F_{\alpha/2}(n_1 - 1, n_2 - 1) \xrightarrow{+} \text{reject} \;\; H_0$

F-test. Example

Example 2

The standard deviation of the average waiting time to see a doctor for non-lifethreatening problems in the emergency room at an urban hospital is 32 minutes. At a second hospital, the standard deviation is 28 minutes. If a sample of 16 patients was used in the first case and 18 in the second case, is there enough evidence to conclude at the 0.01 significance level that the standard deviation of the waiting times in the first hospital is greater than the standard deviation of the waiting times in the second hospital?

F-test. Example

State the hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ and } H_1 : \sigma_1^2 > \sigma_2^2.$$

Find the critical value. Here, d.f.N. = 16 - 1 = 15, and d.f.D. = 18 - 1 = 17. From the $\alpha = 0.01$ table, the critical value is 3.31.
Compute the test value:

$$F = \frac{s_1^2}{s_2} = \frac{32^2}{28^2} = 1.31$$

Carry out the test: $F > F_\alpha(n_1 - 1, n_2 - 1) \xrightarrow{+} \text{ reject } H_0$, but we have $1.31 < 3.31$, thus we do not reject the null hypothesis.
Summarize: There is not enough evidence to support the claim that the standard deviation of the waiting times of the first hospital is greater than the standard deviation of the waiting times of the second hospital.

P-value method

The three methods used to test hypotheses are:
1. The traditional method
2. The P-value method
3. The confidence interval method

We have considered so far the traditional method. Using the rejection region method to test hypotheses at first we fix a significance level α. Then after computing the value of the test statistic, the null hypothesis H_0 is rejected if the value falls in the rejection region and is otherwise not rejected.
We now consider another way of reaching a conclusion in a hypothesis testing analysis. This alternative approach is based on calculation of a certain probability called a **P-value**. One advantage is that the P-value provides an intuitive measure of the strength of evidence in the data against H_0.

P-value method (2)

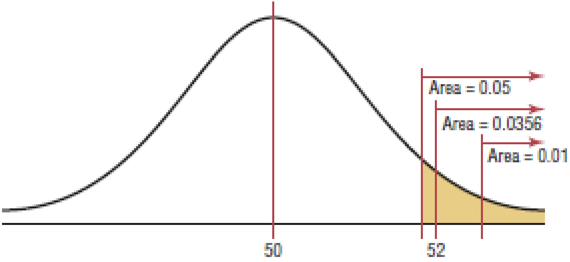
Definition

The **P-value** is the probability that chance alone would produce a test statistic as extreme as the observed test statistic.

In other words, the P-value is the actual area under the standard normal distribution curve (or other curve, depending on what statistical test is being used) representing the probability of a particular sample statistic or a more extreme sample statistic occurring if the null hypothesis is true.

P-value method (3)

For example, suppose that an alternative hypothesis is $H_1 : \mu > 50$ and the mean of a sample is $\overline{X} = 52$. If the computer printed a P-value of 0.0356 for a statistical test, then the probability of getting a sample mean of 52 or greater is 0.0356 if the true population mean is 50 (for the given sample size and standard deviation). The relationship between the P-value and α value can be explained in this manner. For P = 0.0356, the null hypothesis would be rejected at $\alpha = 0.05$ but not at $\alpha = 0.01$.



P-value method (4)

When the hypothesis test is two-tailed, the area in one tail must be doubled. For a two-tailed test, if α is 0.05 and the area in one tail is 0.0356, the P-value will be 2·(0.0356)= 0.0712. That is, the null hypothesis should not be rejected at $\alpha = 0.05$, since 0.0712 is greater than 0.05. In summary, the smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis. **That is, H_0 should be rejected in favor of H_1 when the P-value is sufficiently small.**

Decision rule when using a P-value:

- If P-value $\leq \alpha$, reject the null hypothesis.
- If P-value $> \alpha$, do not reject the null hypothesis.

P-value method (5)

- The P-values for the z-test can be found by using quantile tables of standard normal distribution.
- First find the area under the standard normal distribution curve corresponding to the z-test value.
- For a left-tailed test, use the area given in the table; for a right-tailed test, use 1 minus the area given in the table.
- To get the P-value for a two-tailed test, double the 1 minus area you found in the tail.

P-value method (6)

In case of z-test:

$$P\text{-value:} \quad P = \begin{cases} 1 - \Phi(z) & \text{for an upper -tailed test} \\ \Phi(z) & \text{for a lower -tailed test} \\ 2[1 - \Phi(|z|)] & \text{for a two -tailed test} \end{cases}$$

P-value method. Example

Example 3

A researcher wishes to test the claim that the average cost of tuition and fees at a four year public college is greater than \$5700. She selects a random sample of 36 four-year public colleges and finds the mean to be \$5950. The population standard deviation is \$659. Is there evidence to support the claim at $\alpha = 0.05$? Use the P-value method.

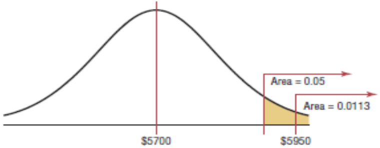
State the hypotheses:

$$H_0 : \mu = 5700 \text{ and } H_1 : \mu > 5700.$$

Compute the test value:

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{5950 - 5700}{659/\sqrt{36}} = 2.28$$

Find the P-value: Using the quantile table of standard normal distribution, find the corresponding area under the normal distribution for z = 2.28. It is 0.9887. Hence the P-value is 1 - 0.9887 = 0.0113.
Since the P-value is less than 0.05, the decision is to reject the null hypothesis.



Summarize: There is enough evidence to support the claim that the tuition and fees at four-year public colleges are greater than $5700.
Note: Had the researcher chosen $\alpha = 0.01$, the null hypothesis would not have been rejected since the P-value (0.0113) is greater than 0.01.

Testing hypotheses using confidence intervals

- There is a relationship between confidence intervals and hypothesis testing.
- When the **null hypothesis is rejected** in a hypothesis testing situation, the **confidence interval** for the mean using the same level of significance will **not contain** the hypothesized mean.
- Likewise, when the **null hypothesis is not rejected**, the **confidence interval** computed using the same level of significance will **contain** the hypothesized mean.

Testing hypotheses using confidence intervals. Example

Example 4

Sugar is packed in 5-pound bags. An inspector suspects the bags may not contain 5 pounds. A sample of 50 bags produces a mean of 4.6 pounds and a standard deviation of 0.7 pound. Is there enough evidence to conclude that the bags do not contain 5 pounds as stated at $\alpha = 0.05$? Also, find the 95% confidence interval of the true mean.

Now

$$H_0 : \mu = 5 \text{ and } H_1 : \mu \neq 5.$$

The critical value is $t_{0.025}(49)$=2.02.

Testing hypotheses using confidence intervals. Example

The test statistic value is

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{4.6 - 5}{0.7/\sqrt{50}} = -4.04$$

The test is: $|t| > t_{\alpha/2}(f) \xrightarrow{+} \text{reject } H_0$, so since $|-4.04| > 2.02$, the null hypothesis is rejected.
The 95% confidence interval for the mean is given by

$$\overline{X} - t_{\alpha/2}(f)\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(f)\frac{s}{\sqrt{n}}$$

$$4.6 - (2.02) \cdot \frac{0.7}{\sqrt{50}} < \mu < 4.6 + (2.02) \cdot \frac{0.7}{\sqrt{50}}$$

So $4.4 < \mu < 4.8$. Notice that the 95% confidence interval of μ does not contain the hypothesized value $\mu = 5$. Hence, there is agreement between the hypothesis test and the confidence interval.

Testing hypotheses using confidence intervals

- In summary, when the null hypothesis is rejected at a significance level of α, the confidence interval computed at the $1 - \alpha$ level will not contain the value of the mean that is stated in the null hypothesis.
- On the other hand, when the null hypothesis is not rejected, the confidence interval computed at the same significance level will contain the value of the mean stated in the null hypothesis.
- These results are true for other hypothesis testing situations and are not limited to means tests.
- The relationship between confidence intervals and hypothesis testing presented here is valid for two-tailed tests. The relationship between one-tailed hypothesis tests and one sided or one-tailed confidence intervals is also valid.