1. Introduction

Welcome to the series of E-learning modules on Confidence Intervals for the variances. In this module, we are going to cover the Interval Estimation- procedure to estimate the population variance or standard deviation when the population mean is known and unknown.

By the end of this session, you will be able to:

- Explain the role of variance or standard deviation
- Explain the confidence interval for the population variance or standard deviation when the population mean is known
- Explain the confidence interval for the population variance when mean is unknown
- Explain how to apply interval estimation technique for the estimation of unknown population variance

Inference is concerned with making decisions or predictions about parameters of the population, which are the numerical descriptive measures that characterize the population.
These parameters may be population mean mu, the population standard deviation sigma, the population correlation coefficient Rho, the Binomial proportion P etc.

Hence, a confidence interval can be computed for almost any value computed from a sample of data, including the standard deviation.
Sample observations may have the same centre but look different because of the way the numbers spread out from the centre. Consider the two distributions as shown in the figure here.

Figure 1(a)

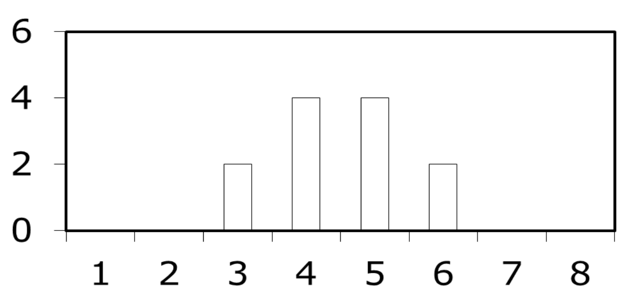
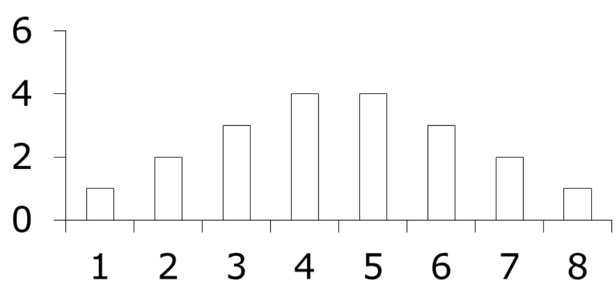


Figure 1(b)



Both the distributions are centred at x in between four and five. However, there is a big difference in the way the measurements spread out or vary. The measurements in the figure 1(a) vary from three to six and in figure 1(b) the measurements vary from zero to eight.

Variability is a very important characteristic of data. For example, if you were manufacturing bolts, extreme variation in the bolt diameter would cause a high percentage of defective products. On the other hand, if you were trying to discriminate between the good and poor accountants, you would have trouble if the examination always produced test grades with little variation making discrimination very difficult.

Measures of variability can help one to create a mental picture of the spread of the data. One of the important measures of variability is variance. To distinguish between variance of the population and variance of the sample, we use the symbols s square for a sample variance and sigma square for the population variance.

It is straightforward to calculate the standard deviation from a sample of values. However, how accurate is that standard deviation? Just by chance, you may have happened to obtain data that are closely bunched together, making the standard deviation low. On the other hand, you may have randomly obtained values that are far more scattered than the overall population, making the standard deviation high.

The standard deviation of your sample does not equal, and may be quite far from the standard deviation of the population.
The variance will be relatively large for highly variable data and relatively small for less variable data. Most often, we will not have population measurements available, but will need to estimate based on the samples drawn from the population.

An estimate of population variance is usually needed before you can make inferences about population means. Sometimes a population variance sigma square is the primary objective in an experimental investigation. It may be more important for the experiment than the population mean.

Consider these examples:

- Scientific measuring instruments must provide unbiased readings with a very small error of measurement. An aircraft altimeter that measures the correct altitude on the average is fairly useless if the measurements are in error by as much as one thousand feet above or below the correct altitude
- Machined parts in a manufacturing process must be produced with minimum variability in order to reduce the out of size and hence defective parts
- Aptitude tests must be designed such that the scores will exhibit a reasonable amount of variability. For example, an eight hundred - point test is not very discriminatory if all students score between (six hundred and one) and (six hundred and five)

In previous topics, we have used s square is equal to summation (yi minus y bar) whole square by n minus 1 as an unbiased estimator of the population variance sigma square.
This means that in repeated sampling, the average of all our sample estimates will equal the target parameter sigma square.

However, how close or far from the target is our estimator s square likely to be? To answer

this question, we use the sampling distribution of s square, which describes its behaviour in repeated sampling.
Consider the distribution of s square based on repeated random sampling from a normal distribution with a specified mean and variance. We can show theoretically that the distribution begins at s square is equal to zero, since the variance cannot be negative, with mean equal to sigma square.

Its shape is non-symmetric and changes with each different sample size and each different value of sigma square. We can standardize the statistic as we did for Z the standardized statistic for variance, which is called as chi-square.
Where, chi-square is equal to (n minus one) s square by sigma square
This is called a chi- square variable and has a sampling distribution called the chi-square probability distribution with n minus 1degrees of freedom.

2. Confidence Interval for Variance

Confidence Interval for variance
Confidence intervals are not just for means. You are already familiar with a confidence interval of a mean. The idea of a confidence interval is very general, and you can express the precision of any computed value as ninety five percent or ninety-nine percent confidence interval (CI).

Practical problems very often lead to estimation of sigma square, the variance of the population. A confidence interval for a variance specifies a range of values within which the unknown population parameter may lie (in this case the variance).

The ninety five percent confidence interval of the variance
The sample variance is just a value you compute from a sample of data, which gives a point estimate of the population variance. It is not done often, but it is certainly possible to compute a confidence interval for a variance.
Interpreting the confidence interval of the variance is straightforward. If you assume that your data were randomly and independently sampled from a Gaussian (Normal) distribution, you can be ninety-five percent sure that the confidence interval computed from the sample variance contains the true population variance.

How wide is the confidence interval of variance or of the standard deviation? Of course, the answer depends on sample size (n). With small samples, the interval is quite wide as shown in the table below.

Figure 2

n	95% CI of SD
2	0.45*SD to 31.9*SD
3	0.52*SD to 6.29*SD
5	0.60*SD to 2.87*SD
10	0.69*SD to 1.83*SD
25	0.78*SD to 1.39*SD

An interval estimator is a rule for computing two numbers say A and B. This is to create an interval that contains the parameter of interest that is population variance sigma square. Hence, a confidence interval gives an estimated range of values for the population variance or standard deviation, which is likely to include an unknown population variance or standard deviation, the estimated range being calculated from a given set of sample data.

3. Confidence Intervals for Variance When the Population Mean is Known

Confidence intervals for variance when the population mean is Known
Assumptions:

- Population mean is known
- Population is normally distributed
- If population is not normal, use large sample

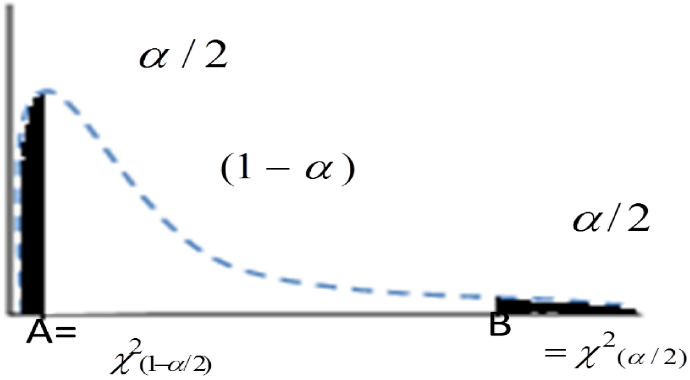
Suppose y one, y two up to y_n are the random samples drawn from a population of size N with a known mean mu and an unknown variance sigma square, we know that,
Population mean mu is equal to summation Yi by N
And Sample mean y bar is equal to summation yi by n

When yi follows normal distribution with mean mu and variance sigma square,
Zi is equal to (yi minus mu) by sigma, which follows normal with mean zero and variance one and we know that summation Zi square follows Chi-square with n degrees of freedom
Chi-square is equal to summation (yi minus mu) whole square by sigma square, which follows Chi-square with n degrees of freedom.
We can use a chi-square distribution to construct a confidence interval for the variance and standard deviation.

From the table of probabilities of Chi- square variables for n degrees of freedom, we can always find two quantities A and B such that
Probability of A less than or equal to Chi-square with n degrees of freedom less than or equal to B is equal to 1 minus alpha
Which implies Probability of A less than or equal to summation (yi minus mu) whole square by sigma square less than or equal to B is equal to 1 minus alpha
Which implies Probability of 1 by A greater than sigma square by summation (yi minus mu) whole square greater than 1 by B is equal to 1 minus alpha

Probability of summation (yi minus mu) whole square by A greater than sigma square greater than summation (yi minus mu) whole square by B is equal to 1 minus alpha
Probability of summation (yi minus mu) whole square by B less than sigma square less than summation (yi minus mu) whole square by A is equal to 1 minus alpha
Therefore, hundred into (1 minus alpha) percent Confidence Interval for the population variance sigma square, when the mean is known as mu is given by
[summation (yi minus mu) whole square by B, summation (yi minus mu) whole square by A]

Figure 3



- As shown in the graph, here there are two critical values for each level of confidence. The value of B is equal to chi square (alpha by two) with n degrees of freedom represents the right-tail critical value and A is equal to chi square (1 minus alpha by two) with n degrees of freedom represents the left-tail critical value. Each area in the Chi-square table represents the region under the chi-square curve to the RIGHT of the critical value.

- You can use the critical values B is equal to chi square (alpha by two) with n degrees of freedom and A is equal to chi square (1 minus alpha by two) with n degrees of freedom to construct confidence intervals for a population variance and standard deviation. As you would expect, the best point estimate for the variance is s square and the best point estimate for the standard deviation is s.

4. Confidence Intervals for Variance with Unknown Population Mean

2) Confidence Intervals for variance with unknown population mean

1) Assumptions

1. Population mean is unknown
2. Population is normally distributed

Suppose y one, y two up to y_n are the random samples drawn from a population of size N with an unknown mean mu and variance sigma square, we know that,
Population mean mu is equal to summation Yi by N
Sample mean y bar is equal to summation yi by n
And s square is equal to summation (yi minus y bar) whole square by n minus one
Zi is equal to (yi minus mu) by sigma, which follows Normal with mean zero and variance one.
Since mean mu is unknown, we substitute its point estimate y bar, then we know that summation Zi square follows Chi-square with n minus 1 degrees of freedom
Chi square is equal to summation (yi minus y bar) whole square by sigma square which follows chi- square distribution with (n minus one) degrees of freedom

From the table of probabilities of Chi- square variables for (n minus 1) degrees of freedom, we can always find two quantities A and B such that
Probability of A less than or equal to Chi-square with n minus 1 degrees of freedom less than or equal to B is equal to 1 minus alpha
Which implies Probability of A less than or equal to summation (yi minus y bar) whole square by sigma square less than or equal to B is equal to 1 minus alpha
Which implies Probability of 1 by A greater than sigma square by summation (yi minus y bar) whole square greater than 1 by B is equal to 1 minus alpha

Probability of summation (yi minus y bar) whole square by A greater than sigma square greater than summation (yi minus y bar) whole square by B is equal to 1 minus alpha
Probability of summation (yi minus y bar) whole square by B less than sigma square less than summation (yi minus y bar) whole square by A is equal to 1 minus alpha
Therefore, hundred into (1 minus alpha) percent Confidence Interval for the population variance sigma square when the mean is unknown is given by,
[summation (yi minus y bar) whole square by B , summation (yi minus y bar) whole square by A]

Since s square is equal to summation (yi minus y bar whole square) by n minus 1 which implies summation (yi minus y bar) whole square is equal to (n minus 1) into s square
Therefore, hundred into (1 minus alpha) percent Confidence Interval for the population variance sigma square, when the mean is unknown is given by
[(n minus 1) into s square by B, (n minus 1) into s square by A]

5. Steps to Obtain Confidence Interval for Variance

Steps to obtain confidence interval for variance

- First obtain the point estimate of mu, that is, the sample mean y bar if population mean is not known otherwise consider the value of mu
- Select a confidence level. The most common level is ninety five percent or ninety nine percent
- Obtain the critical values of Chi square probabilities for alpha by two and 1 minus alpha by two level of significance with n degrees of freedom if mu is known (n minus 1 degrees of freedom if mu is unknown)
- Let B is equal to Chi-square (alpha by 2) with n degrees of freedom and A is equal to Chi-square (1 minus alpha by 2) with n degrees of freedom
- Then, the confidence interval for variance when mean is known as mu is given by [summation yi minus mu whole square by B, summation yi minus mu whole square by A]
- When the population mean is unknown, Let B is equal to Chi-square (alpha by 2) with (n minus 1) degrees of freedom and A is equal to Chi-square (1 minus alpha by 2) with (n minus 1) degrees of freedom
- Then, the confidence interval for variance when mean is unknown is computed as [summation yi minus y bar whole square by B, summation yi minus y bar whole square by A] which is also equal to [summation (n minus 1) into s square by B, summation (n minus 1) into s square by A]
- When reporting a confidence interval, make sure you report both the interval and the confidence level

Example:

Randomly select and weigh thirty samples of an allergy medication. The sample standard deviation is one point two milligrams. Assuming the weights are normally distributed, construct ninety-nine percent confidence intervals for the population variance and standard deviation.

Solution:

Given n is equal to thirty, s is equal to one point two and (1 minus alpha) is equal to ninety nine percent, which is equal to zero point nine nine.
Alpha is equal to 1 minus zero point nine nine, which is equal to zero point zero one
Then, alpha by two is equal zero point zero zero five and 1 minus alpha by two is equal to zero point nine nine five
From the table of Chi square distribution, B is equal to chi square point zero zero five (twenty nine) is equal to fifty two point three three six and A is equal to chi square point nine nine five (twenty nine) is equal to thirteen point one two one

Hence, hundred into 1 minus alpha percent confidence Interval for the population variance sigma square when the mean is unknown is given by,

[(n minus 1) into s square by B, (n minus 1) into s square by A]
Is equal to [zero point seven nine eight, three point one eight three]
Therefore, with ninety-nine percent confidence, you can say that the population variance is between zero point seven nine eight and three point one eight three. The population standard deviation is between zero point eight nine and one point seven eight milligrams.

Here's a summary of our learning in this session, where we have understood:

- The derivation of confidence limits for the unknown population variance when mean mu is known
- The derivation of confidence limits for the unknown population variance when mean mu is unknown
- The illustration of interval estimation of population variance by examples