Lecture No 14
Confidence Interval on the Variance and Proportion

Dr. Shafayat Abrar

A Confidence Interval for μ and known Variance of Normal distribution

Givene $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution, we may write

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

A Large-Sample Confidence Interval for μ and Unknown Variance

For large n, i.e., $n > 40$, the quantity $(\overline{X} - \mu)/(S/\sqrt{n})$ is approximately standard normal distribution. We may write:

$$P\left(\overline{X} - z_{\alpha/2}S/\sqrt{n} \leq \mu \leq \overline{X} + z_{\alpha/2}S/\sqrt{n}\right) = 1 - \alpha$$

where

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

A Small-Sample ($n < 40$) Confidence Interval for μ and Unknown Variance

We know that the distribution of $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is t with $n-1$ degrees of freedom. Letting $t_{\alpha/2,n-1}$ be the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom, we may write:

$$P\left(\overline{X} - t_{\alpha/2,n-1}S/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2,n-1}S/\sqrt{n}\right) = 1 - \alpha$$

where

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

Confidence Interval on the Variance of a Normal Population

- Sometimes confidence intervals on the population variance are needed.
- When the population is modelled by a normal distribution, the tests and intervals described in this lecture are applicable. The following result provides the basis of constructing these confidence intervals.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

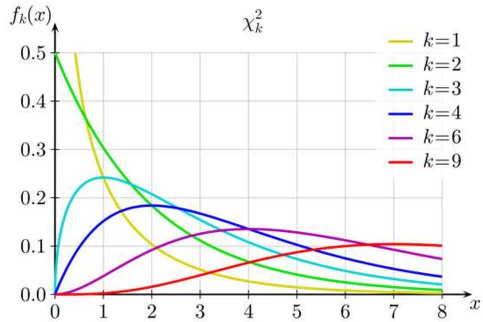
$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

has a chi-square (χ^2) distribution with $n - 1$ degrees of freedom.

The probability density function of a χ^2 random variable is

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad x > 0$$

where k is the number of degrees of freedom. The mean and variance of the χ^2 distribution are k and $2k$, respectively. Several chi-square distributions are shown in the figure. Note that the chi-square random variable is nonnegative and that the probability distribution is skewed to the right. However, as k increases, the distribution becomes more symmetric. As $k \to \infty$, the limiting form of the chi-square distribution is the normal distribution.



Define $\chi^2_{\alpha,k}$ as the percentage point or value of the chi-square random variable with k degrees of freedom such that the probability that X^2 exceeds this value is α. That is,

$$P\left(X^2 > \chi^2_{\alpha,k}\right) = \int_{\chi^2_{\alpha,k}}^{\infty} f(u)du = \alpha$$

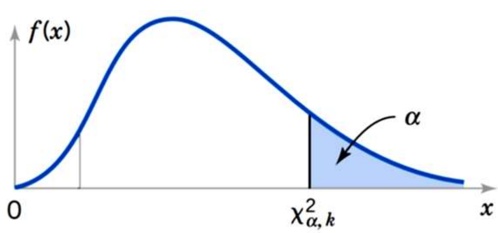
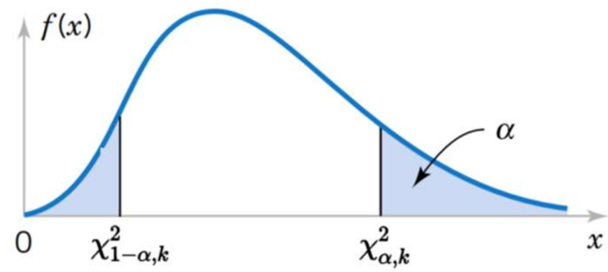
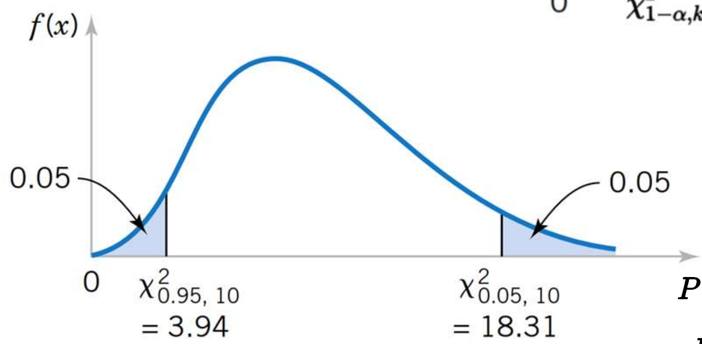


Table: Critical Values For The Chi-Square Distribution
This table contains critical values $\chi^2_{\alpha,k}$ for the Chi-Square distribution defined by $P\left(\chi^2 \geq \chi^2_{\alpha,k}\right) = \alpha$.

k	.9999	.9995	.999	.995	.99	.975	.95	.90
1	$.0^7157$	$.0^6393$	$.0^5157$	$.0^4393$	.0002	.0010	.0039	.0158
2	.0002	.0010	.0020	.0100	.0201	.0506	.1026	.2107
3	.0052	.0153	.0243	.0717	.1148	.2158	.3518	.5844
4	.0284	.0639	.0908	.2070	.2971	.4844	.7107	1.0636

To obtain both **upper** and **lower** limits on the confidence interval of the variance





$$P(\chi^2 > \chi^2_{0.05,10}) = P(\chi^2 > 18.31) = 0.05$$

$$P(\chi^2 < \chi^2_{0.95,10}) = P(\chi^2 < 3.94) = 0.05$$

The construction of the $100(1-\alpha)\%$CI for σ^2 is straightforward. Because

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

is chi-square with $n-1$ degrees of freedom, we may write

$$P\left(\chi^2_{1-\alpha/2,n-1} \le \chi^2 \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha$$

so that

$$P\left(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right) = 1-\alpha$$

This last equation can be rearranged as

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

This leads to the following definition of the confidence interval for σ^2.

Two-Sided Confidence Interval on the Variance of a Normal Population

If S^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} \qquad (A)$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A confi- -dence interval for $\boldsymbol{\sigma}$ has lower and upper limits that are the square roots of the corresponding limits in Equation A.

One-Sided Confidence Interval on the Variance of a Normal Population

It is also possible to find a $100(1-\alpha)\%$ lower confidence bound or upper confidence bound on σ^2. The $100(1-\alpha)\%$ **lower** and **upper** confidence bounds on σ^2 are

$$\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}$$

respectively.

k	.9999	.9995	.999	.995	.99	.975	.95	.90
19	3.9683	4.9123	5.4068	6.8440	7.6327	8.9065	10.1170	11.6509

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces) ^{2}. If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. Find a 95% upper-confidence interval.

$$\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}} \qquad \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0.95,19}}$$

or

$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

This last expression may be converted into a confidence interval on the standard deviation σ by taking the square root of both sides, resulting in

$$\sigma \leq 0.17$$

Therefore, at the 95% level of confidence, the data indicate that the process standard deviation could be as large as 0.17 fluid ounce.

A rivet is to be inserted into a hole. A random sample of $n = 15$ parts is selected, and the hole diameter is measured. The sample standard deviation of the hole diameter measurements is $S = 0.008$ millimeters. Construct a 99% lower confidence bound for σ^2.

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

99% lower confidence bound for σ^2

For $\alpha = 0.01$ and n $= 15$, $\chi^2_{\alpha,n-1} = \chi^2_{0.01,14} = 29.14$

$$\frac{14(0.008)^2}{29.14} < \sigma^2$$
$$0.00003075 < \sigma^2$$
$$\sigma > 0.00555$$

Table 9. Critical Values For The Chi-Square Distribution (Continued)

	.10	.05	.025	.01	.005	.001	.0005	.0001
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233	38.1094	42.5793

A Large-Sample Confidence Interval for a Population Proportion

It is often necessary to construct confidence intervals on a population proportion. For example, suppose that a random sample of size n has been taken from a large (possibly infinite) population and that X observations in this sample belong to a class of interest.

Then $\hat{p} = X/n$ is a point estimator of the proportion of the population p that belongs to this class. Note that n and p are the parameters of a binomial distribution.

We know that the sampling distribution of $\hat{p}$ is approximately normal with mean p and variance $p(1-p)/n$, if n is relatively large.

$$n(1-p) \geq 5 \quad \text{and} \quad np \geq 5$$

If n is large, the distribution of

$$Z = \frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately standard normal.

To construct the confidence interval on p, note that

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

so

$$P\left(-z_{\alpha/2} \leq \frac{\widehat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

This may be rearranged as

$$P\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha$$

$$P\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1-\alpha$$

The quantity $\sqrt{p(1-p)/n}$ in the last equation is the standard error of the point estimator $\widehat{p}$. Unfortunately, the upper and lower limits of the confidence interval obtained from this equation contain the unknown parameter p. However, a satisfactory solution is to replace p by $\widehat{p}$ in the standard error, which results in

$$P\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right) \approx 1-\alpha$$

If $\widehat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

This procedure depends on the adequacy of the normal approximation to the binomial. To be reasonably conservative, this requires that np and $n(1-p)$ be greater than or equal to 5 . In situations where this approximation is inappropriate, particularly in cases where n is small, other methods must be used. Tables of the binomial distribution could be used to obtain a confidence interval for p.

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\hat{p} = x/n = 10/85 = 0.12$. A 95% two-sided confidence interval for p is computed

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

or

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \leq p \leq 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

which simplifies to

$$0.05 \leq p \leq 0.19$$

Choice of Sample Size

Since $\widehat{p}$ is the point estimator of p, we can define the error in estimating p by $\widehat{p}$ as $E = |p - \widehat{p}|$. Note that we are approximately $100(1-\alpha)\%$ confident that this error is less than $z_{\alpha/2}\sqrt{p(1-p)/n}$.

For instance, in the last example, we are 95% confident that the sample proportion $\widehat{p} = 0.12$ differs from the true proportion p by an amount not exceeding 0.07.

In situations where the sample size can be selected, we may choose n to be $100(1-\alpha)\%$ confident that the error is less than some specified value E.

If we set $E = z_{\alpha/2}\sqrt{p(1-p)/n}$ and solve for n, the appropriate sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

Since true p is not known, we may use

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \widehat{p}\,(1-\widehat{p})$$

Choice of Sample Size

Another approach to choosing n uses the fact that the sample size from the formula for n will always be a maximum for $p = 0.5$ [that is, $p(1-p) \leq 0.25$ with equality for $p = 0.5$], and this can be used to obtain an upper bound on n. In other words, we are at least $100(1-\alpha)\%$ confident that the error in estimating p by $\widehat{p}$ is less than E if the sample size is

$$n = \left(\frac{z_{\alpha/2}}{2E}\right)^2$$

Consider the situation in the last Example. How large a sample is required if we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05 ? Using $\hat{p} = 0.12$ as an initial estimate of p, we find from Equation 8-26 that the required sample size is

$$n = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.12(0.88) \cong 163$$

If we wanted to be at least 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p, we would use the maximum n formula to find the sample size

$$n = \left(\frac{z_{0.025}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.05}\right)^2 (0.25) \cong 385$$

Notice that if we have information concerning the value of p, either from a preliminary sample or from past experience, we could use a smaller sample while maintaining both the desired precision of estimation and the level of confidence.

One-Sided Confidence Bounds on Population Proportion

We may find approximate one-sided confidence bounds on p by a simple modification
The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are

$$\hat{p} - z_{\alpha}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \quad \text{and} \quad p \leq \widehat{p} + z_{\alpha}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

respectively.