CHI-SQUARED DISTRIBUTION AND CONFIDENCE INTERVALS OF THE VARIANCE

Suppose that we have a sample of *n* measured values $x_1, x_2, x_3, \ldots, x_n$ of a single unknown quantity. Assuming that the measurements are drawn from a normal distribution having mean μ and variance σ^2 it is reasonable to estimate these population parameters (μ, σ) with the arithmetic mean

$$\overline{x} = \frac{\sum_{k=1}^{n} x_k}{n} \tag{1}$$

and the sample variance

$$s^2 = \frac{\sum_{k=1}^{n} \left(\overline{x} - x_k\right)^2}{n-1} \tag{2}$$

where standard deviations s, σ (sample and population respectively) are positive square-roots of variances s^2, σ^2.

We may estimate *confidence intervals* for the variance σ^2 using the statistic χ^2 (chi-squared) defined as

$$\chi^2 = \frac{\left(\overline{x} - x_1\right)^2 + \left(\overline{x} - x_2\right)^2 + \cdots + \left(\overline{x} - x_n\right)^2}{\sigma^2} = \frac{s^2\left(n-1\right)}{\sigma^2} = \frac{s^2 \nu}{\sigma^2} \tag{3}$$

and χ^2 is assumed to be a value of a random variable X^2 having a χ^2 distribution with $\nu = n - 1$ degrees of freedom.

Here we are using a theorem of statistical sampling theory:

> If $Y_1, Y_2, \ldots, Y_n$ are each random samples of size ν from a normal population with mean μ and standard deviation σ, then $Z_k = \left(Y_k - \mu\right)/\sigma$ are independent standard normal random variables $k = 1, 2, \ldots, n$ and
>
> $$\sum_{k=1}^{n} Z_k^2 = \sum_{k=1}^{n} \left(\frac{Y_k - \mu}{\sigma}\right)^2 = \frac{\sum_{k=1}^{n} \left(Y_k - \mu\right)^2}{\sigma^2}$$
>
> has a χ^2 distribution with ν degrees of freedom.

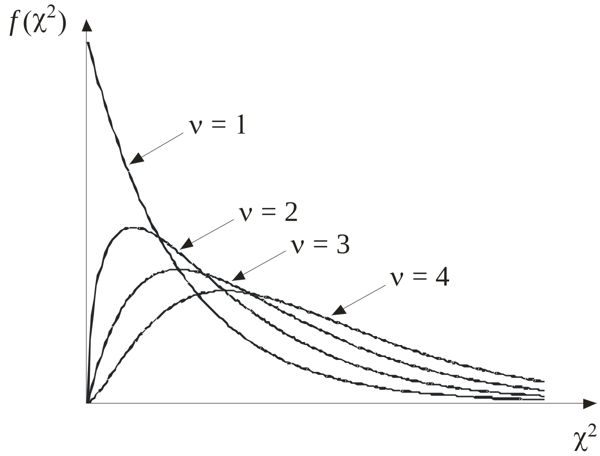


Figure 1: Chi-squared distribution curves for $\nu = 1, 2, 3, 4$ degrees of freedom

The chi-squared distribution (or probability density function - pdf) is not defined for negative values of χ^2 and for positive values, for various degrees of freedom, the curves are not symmetric.

The area under the distribution curve is unity and it is usual to let χ^2_α represent the χ^2 value above which we find a corresponding area α

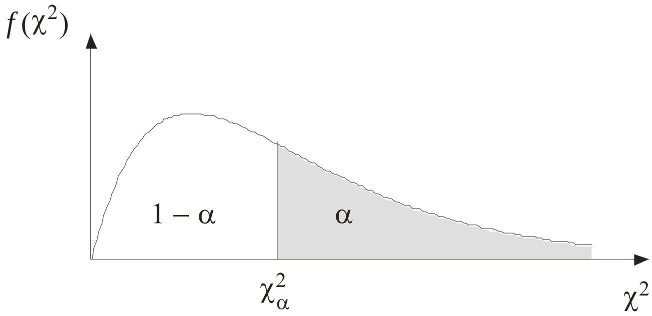


Figure 2: The area α under the chi-squared distribution curve in the right-hand tail

Tabulated values for χ^2_α corresponding to areas α for various degrees of freedom ν are given in Table 1.

Confidence intervals (CI) for the variance σ^2 can be determined from the following probability statement

$$P\left(\chi_1^2 < \mathrm{X}^2 < \chi_2^2\right) = 1 - \alpha \tag{4}$$

where $P(\ )$ is probability and α is a significance level. χ_1^2 is a value below which we find a corresponding area $\alpha/2$ and χ_2^2 is a value above which we find the corresponding area $\alpha/2$. The area under the curve (i.e., the probability) between these two values is $1-\alpha$.

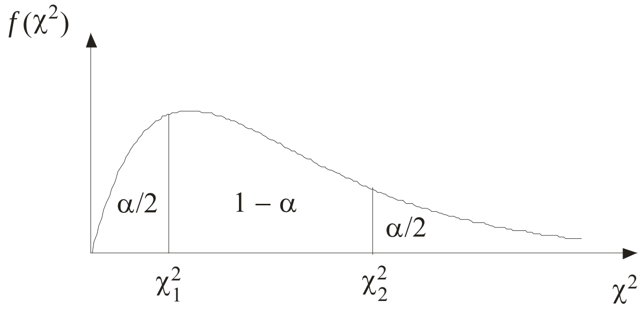


Figure 3: Areas $\alpha/2$ in the left and right-hand tails of the chi-squared distribution

Substituting the statistic χ^2 [equation (3)] for the random variable X^2 in equation (4) gives

$$P\left(\chi_1^2 < \frac{s^2\nu}{\sigma^2} < \chi_2^2\right) = 1 - \alpha \tag{5}$$

and the inequality on the left-hand side may be re-arranged as

$$\sigma^2\chi_1^2 < s^2\nu < \sigma^2\chi_2^2 \tag{6}$$

Dividing the first inequality by χ_1^2 and the second inequality by χ_2^2 gives

$$\sigma^2 < \frac{s^2\nu}{\chi_1^2} \quad \text{and} \quad \sigma^2 > \frac{s^2\nu}{\chi_2^2}$$

and we may write the probability statement for the variance σ^2 is

$$P\left(\frac{s^2\nu}{\chi_2^2} < \sigma^2 < \frac{s^2\nu}{\chi_1^2}\right) = 1 - \alpha \tag{7}$$

Thus the lower and upper confidence limits of the variance are:

$$\begin{aligned} \text{lower confidence limit } &= \frac{s^2\nu}{\chi_2^2} \\ \text{upper confidence limit } &= \frac{s^2\nu}{\chi_1^2} \end{aligned} \tag{8}$$

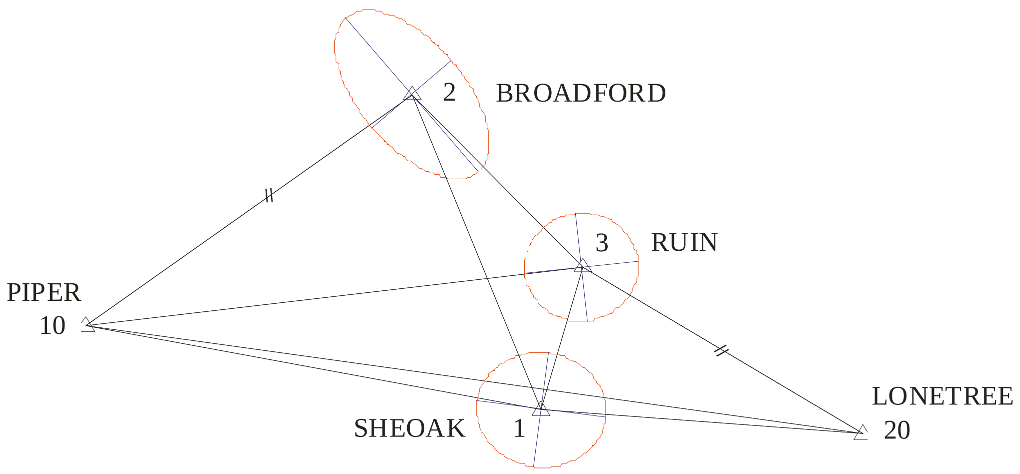
The probability statement for the standard deviation σ is

$$P\left(s\sqrt{\frac{\nu}{\chi_2^2}} < \sigma < s\sqrt{\frac{\nu}{\chi_1^2}}\right) = 1 - \alpha \tag{9}$$

with the lower and upper confidence limits of the standard deviation

$$\begin{aligned} \text{lower confidence limit } &= s\sqrt{\frac{\nu}{\chi_2^2}} \\ \text{upper confidence limit } &= s\sqrt{\frac{\nu}{\chi_1^2}} \end{aligned} \tag{10}$$

Example 1



A small geodetic network is observed connecting Broadford, Ruin and Sheoak (floating) with the Piper and Lonetree (fixed). Theodolite directions are observed at all points and distances Piper-Broadford and Ruin-Lonetree measured. Standard deviations of directions and distances are estimated as 2″ and 0.040 m respectively. The observations are adjusted (least squares - variation of coordinates) and the adjustment information is shown below.

What are the 95% confidence limits for the variance factor?

Fixed Stations:

Name	Serial	East	North
Piper	10	322809.939	5880322.062
Lonetree	20	337425.856	5878317.969

Adjusted Stations:

Name	Serial	East	North
Sheoak	1	331388.858	5878746.544
Broadford	2	328963.118	5884669.102
Ruin	3	332141.185	5881429.991

Observed Directions:

station		observation	st.dev	resid	orient const	plane brg	
at	to	deg min sec	sec	sec	deg min sec	deg min sec	plane dist
10	2	0 0 0.00	2.00	0.75	54 45 34.97	54 45 35.72	7533.815
10	3	28 28 7.40	2.00	1.26		83 13 43.64	9396.790
10	20	43 2 51.60	2.00	0.60		97 48 27.17	14752.675
10	1	45 38 50.70	2.00	-2.61		100 24 23.07	8722.391
2	3	0 0 0.00	2.00	-0.26	135 32 42.31	135 32 42.06	4537.835
2	1	22 10 54.10	2.00	1.35		157 43 37.76	6400.070
2	10	99 12 54.50	2.00	-1.09		234 45 35.72	7533.815
3	20	0 0 0.00	2.00	2.68	120 29 31.61	120 29 34.30	6132.897
3	1	75 10 10.60	2.00	-1.55		195 39 40.66	2786.913
3	10	142 44 15.30	2.00	-3.28		263 13 43.64	9396.790
3	2	195 3 8.30	2.00	2.14		315 32 42.06	4537.835
20	1	0 0 0.00	2.00	0.82	274 3 37.69	274 3 38.50	6052.192
20	10	3 44 49.70	2.00	-0.22		277 48 27.17	14752.675
20	3	26 25 57.20	2.00	-0.59		300 29 34.30	6132.897
1	10	0 0 0.00	2.00	-0.91	280 24 23.97	280 24 23.07	8722.391
1	2	57 19 11.90	2.00	1.89		337 43 37.76	6400.070
1	3	95 15 16.30	2.00	0.39		15 39 40.66	2786.913
1	20	173 39 15.90	2.00	-1.37		94 3 38.50	6052.192

Observed Distances:

station at	to	observation metres	st.dev metres	resid metres	plane brg deg min sec	plane dist
10	2	7533.890	0.040	-0.075	54 45 35.72	7533.815
3	20	6132.950	0.040	-0.053	120 29 34.30	6132.897

Standard Deviations and Standard Error Ellipse:

Serial	East	North	St. Devs E	St. Devs N	Standard Error Ellipse semi-major	semi-minor	bearing d m s
1	331388.858	5878746.544	0.055	0.048	0.055	0.048	97 20 58
2	328963.119	5884669.102	0.065	0.071	0.086	0.044	139 18 48
3	332141.185	5881429.991	0.048	0.046	0.048	0.046	83 47 59

Adjusted Bearings and Distances:

station at	to	plane bearing deg min sec	st.dev sec	plane distance m	st.dev m
10	2	54 45 35.72	2.35	7533.815	0.045
10	3	83 13 43.64	1.00	9396.790	0.048
10	20	97 48 27.17		14752.675	
10	1	100 24 23.07	1.14	8722.391	0.055
2	3	135 32 42.06	1.90	4537.835	0.092
2	1	157 43 37.76	1.98	6400.070	0.096
2	10	234 45 35.72	2.35	7533.815	0.045
3	20	120 29 34.30	1.56	6132.897	0.047
3	1	195 39 40.66	2.26	2786.913	0.057
3	10	263 13 43.64	1.00	9396.790	0.048
3	2	315 32 42.06	1.90	4537.835	0.092
20	1	274 3 38.50	1.65	6052.192	0.055
20	10	277 48 27.17		14752.675	
20	3	300 29 34.30	1.56	6132.897	0.047
1	10	280 24 23.07	1.14	8722.391	0.055
1	2	337 43 37.76	1.98	6400.070	0.096
1	3	15 39 40.66	2.26	2786.913	0.057
1	20	94 3 38.50	1.65	6052.192	0.055
10	2	54 45 35.72	2.35	7533.815	0.045
3	20	120 29 34.30	1.56	6132.897	0.047

Variance Factor = 1.8266e+000

Cofactor matrix Qxx (upper triangular portion)
Units of metre squared, printed in same order as adjusted stations.
The * symbol indicates new row of cofactor matrix beginning at
diagonal element, east variance followed by covariances
then north variance followed by covariances.

```
*  1.6456e-003 -4.7764e-005  2.7982e-004  4.0430e-004  1.1701e-003 -1.3233e-004
*  1.2814e-003  4.5979e-005  2.0235e-004 -1.6779e-004  3.5185e-004
*  2.3281e-003 -1.4758e-003  2.9378e-004 -1.8126e-004
*  2.7759e-003  5.4352e-004  6.7794e-004
*  1.2574e-003  1.2715e-005
*  1.1417e-003
```

Adjustment Data:

```
 2 fixed stations
 3 floating stations
18 observed directions
 2 observed distances
 0 constrained bearings
 0 constrained distances
 0 constrained angles
 5 orientation constants

degrees of freedom = observations
                   - unknowns
                   + constraints
                   =  9
 4 iterations required for solution
```

Using equation (7) we may write the probability statement as

$$P\left(\frac{\hat{\sigma}_0^2 \nu}{\chi_2^2} < \sigma_0^2 < \frac{\hat{\sigma}_0^2 \nu}{\chi_1^2} \right) = 1 - \alpha$$

where σ_0^2 is the variance factor and $\hat{\sigma}_0^2 = 1.8266$ is the estimate from the adjustment.

With $\alpha = 0.05$, $1-\alpha = 0.95$, $\alpha/2 = 0.025$ and degrees of freedom $\nu = 9$ we obtain from Table 1

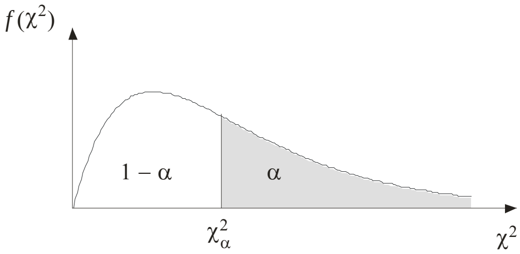
$\chi_1^2 = 2.7004$ (for area 0.975 in right-hand tail)

$\chi_2^2 = 19.0228$ (for area 0.025 in right-hand tail)

giving the 95% confidence interval for the variance factor

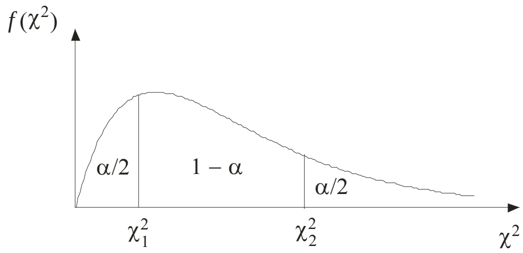
$$\frac{\hat{\sigma}_0^2 \nu}{\chi_2^2} < \sigma_0^2 < \frac{\hat{\sigma}_0^2 \nu}{\chi_1^2} = \frac{(1.8266)(9)}{19.0228} < \sigma_0^2 < \frac{(1.8266)(9)}{2.7004} = 0.86 < \sigma_0^2 < 6.09$$

TABLE 1: CHI-SQUARED DISTRIBUTION
Probability for a given degree of freedom
(the area in the right-hand tail of the distribution)



ν	α							
	0.005	0.01	0.025	0.05	0.95	0.975	0.99	0.995
1	7.8794	6.6349	5.0239	3.8415	0.0039	0.00098	0.00016	0.000039
2	10.5966	9.2103	7.3778	5.9915	0.1026	0.0506	0.0201	0.0100
3	12.8382	11.3449	9.3484	7.8147	0.3518	0.2158	0.1148	0.0717
4	14.8603	13.2767	11.1433	9.4877	0.7107	0.4844	0.2971	0.2070
5	16.7496	15.0863	12.8325	11.0705	1.1455	0.8312	0.5543	0.4117
6	18.5476	16.8119	14.4494	12.5916	1.6354	1.2373	0.8721	0.6757
7	20.2777	18.4753	16.0128	14.0671	2.1673	1.6899	1.2390	0.9893
8	21.9550	20.0902	17.5345	15.5073	2.7326	2.1797	1.6465	1.3444
9	23.5894	21.6660	19.0228	16.9190	3.3251	2.7004	2.0879	1.7349
10	25.1882	23.2093	20.4832	18.3070	3.9403	3.2470	2.5582	2.1559
11	26.7568	24.7250	21.9200	19.6751	4.5748	3.8157	3.0535	2.6032
12	28.2995	26.2170	23.3367	21.0261	5.2260	4.4038	3.5706	3.0738
13	29.8195	27.6882	24.7356	22.3620	5.8919	5.0088	4.1069	3.5650
14	31.3193	29.1412	26.1189	23.6848	6.5706	5.6287	4.6604	4.0747
15	32.8013	30.5779	27.4884	24.9958	7.2609	6.2621	5.2293	4.6009
16	34.2672	31.9999	28.8454	26.2962	7.9616	6.9077	5.8122	5.1422
17	35.7185	33.4087	30.1910	27.5871	8.6718	7.5642	6.4078	5.6972
18	37.1565	34.8053	31.5264	28.8693	9.3905	8.2307	7.0149	6.2648
19	38.5823	36.1909	32.8523	30.1435	10.1170	8.9065	7.6327	6.8440
20	39.9968	37.5662	34.1696	31.4104	10.8508	9.5908	8.2604	7.4338
21	41.4011	38.9322	35.4789	32.6706	11.5913	10.2829	8.8972	8.0337
22	42.7957	40.2894	36.7807	33.9244	12.3380	10.9823	9.5425	8.6427
23	44.1813	41.6384	38.0756	35.1725	13.0905	11.6886	10.1957	9.2604
24	45.5585	42.9798	39.3641	36.4150	13.8484	12.4012	10.8564	9.8862
25	46.9279	44.3141	40.6465	37.6525	14.6114	13.1197	11.5240	10.5197
26	48.2899	45.6417	41.9232	38.8851	15.3792	13.8439	12.1981	11.1602
27	49.6449	46.9629	43.1945	40.1133	16.1514	14.5734	12.8785	11.8076
28	50.9934	48.2782	44.4608	41.3371	16.9279	15.3079	13.5647	12.4613
29	52.3356	49.5879	45.7223	42.5570	17.7084	16.0471	14.2565	13.1211
30	53.6720	50.8922	46.9792	43.7730	18.4927	16.7908	14.9535	13.7867

For a 95% CI with $\nu = 3$ degrees of freedom,
$\alpha = 0.05$, $1-\alpha = 0.95$ and $\alpha/2 = 0.025$
From above:
$\chi_1^2 = 0.2158$ (for area 0.975 in right-hand tail)
$\chi_2^2 = 9.3484$ (for area 0.025 in right-hand tail)



Equation of the chi-squared distribution

The distribution of χ^2 (or the probability density function – pdf) is

$$f\left(\chi^2\right)=\frac{1}{2^{\frac{1}{2}\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\chi^2\right)^{\frac{1}{2}(\nu-2)}\,e^{-\frac{1}{2}\chi^2} \tag{11}$$

where $\Gamma(\;)$ is the *gamma* function. The gamma function was introduced to extend the factorial function from integers to real numbers.

The factorial function for the positive integer *n* is defined by

$$n!=1\times 2\times 3\times\cdots\times(n-2)\times(n-1)\times n$$

with *zero factorial* defined as $0!=1$

The gamma function is defined by

$$\Gamma(\nu+1)=\int_0^\infty x^\nu e^{-x}\,dx \qquad \text{for} \;\; \nu>-1 \tag{12}$$

with special results

$$\Gamma\left(\tfrac{1}{2}\right)=\sqrt{\pi}\;, \qquad \Gamma(1)=1 \tag{13}$$

The gamma function can be evaluated from the recurrence relationship

$$\Gamma(\nu+1)=\nu\,\Gamma(\nu) \tag{14}$$

noting that if ν is any positive integer *n* then $\Gamma(n+1)=n!$

For example if $\nu=9$ then in equation (11) we have (Lauf, 1983 p.55)

$$\Gamma\left(\frac{\nu}{2}\right)=\Gamma\left(\frac{9}{2}\right)=\frac{7}{2}\times\frac{5}{2}\times\frac{3}{2}\times\frac{1}{2}\Gamma\left(\tfrac{1}{2}\right)=\frac{7}{2}\frac{5}{2}\frac{3}{2}\frac{1}{2}\sqrt{\pi}=11.631728397$$

and if $\nu=10$ then

$$\Gamma\left(\frac{\nu}{2}\right)=\Gamma\left(\frac{10}{2}\right)=\Gamma(5)=4!=1\times 2\times 3\times 4=24$$

Reference

Lauf, G.B., (1983), *The Method of Least Squares with applications in surveying*, Tafe Publications Unit, Collingwood, Australia.