# Interval Estimation for the Variance

## 40.2

### Introduction

In Section 40.1 we have seen that the sampling distribution of the sample mean, when the data come from a normal distribution (and even, in large samples, when they do not) is itself a normal distribution. This allowed us to find a confidence interval for the population mean. It is also often useful to find a confidence interval for the population variance. This is important, for example, in quality control. However the distribution of the sample variance is not normal. To find a confidence interval for the population variance we need to use another distribution called the "chi-squared" distribution.

### Prerequisites

Before starting this Section you should . . .

- understand and be able to calculate means and variances

- understand the concepts of continuous probability distributions

- understand and be able to calculate a confidence interval for the mean of a normal distribution

### Learning Outcomes

On completion you should be able to . . .

- find probabilities using a chi-squared distribution

- find a confidence interval for the variance of a normal distribution

# 1. Interval estimation for the variance

In Section 40.1 we saw how to find a confidence interval for the mean of a normal population. We can also find a confidence interval for the variance. The corresponding confidence interval for the standard deviation is found by taking square roots.

We know that if we take samples from a population, then each sample will have a mean and a variance associated with it. We can calculate the values of these quantities from first principles, that is we can use the basic definitions of the mean and the variance to find their values. Just as the means form a distribution, so do the values of the variance and it is to this distribution that we turn in order to find an interval estimate for the value of the variance of the population. Note that if the original population is normal, samples taken from this population have means which are normally distributed. When we consider the distribution of variances calculated from the samples we need the chi-squared (usually written as $\chi^2$ ) distribution in order to calculate the confidence intervals. As you might expect, the values of the chi-squared distribution are tabulated for ease of use. The calculation of confidence intervals for the variance (and standard deviation) depends on the following result.

---

**Key Point 2**

If $x_1, x_2, \cdots, x_n$ is a random sample taken from a normal population with mean $\mu$ and variance $\sigma^2$ then if the sample variance is denoted by $S^2$, the random variable

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

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

---

Clearly, a little explanation is required to make this understandable! Key Point 2 refers to the chi-squared distribution and the term 'degrees of freedom.' Both require some detailed explanation before the Key Point can be properly understood. We shall start by looking in a little detail at the chi-squared distribution and then consider the term 'degrees of freedom.' You are advised to read these explanations very carefully and make sure that you fully understand them.

### The chi-squared random variable

The probability density function of a $\chi^2$ random variable is somewhat complicated and involves the gamma ($\Gamma$) function. The gamma function, for positive $r$, is defined as

$$\Gamma(r) = \int_0^\infty x^{r-1}\mathrm{e}^{-x}dx$$

It is easily shown that $\Gamma(r) = (r-1)\Gamma(r-1)$ and that, if $r$ is an integer, then

$$\Gamma(r) = (r-1)(r-2)(r-3)\cdots(3)(2)(1) = (r-1)!$$

The probability density function is

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

The plots in Figure 2 show the probability density function for various convenient values of $k$. We have deliberately taken even values of $k$ so that the gamma function has a value easily calculated from the above formula for a factorial. In these graphs the vertical scaling has been chosen to ensure each graph has the same maximum value.

It is possible to discern two things from the diagrams.

Firstly, as $k$ increases, the peak of each curve occurs at values closer to $k$. Secondly, as $k$ increases, the shape of the curve appears to become more and more symmetrical. In fact the mean of the $\chi^2$ distribution is $k$ and in the limit as $k \to \infty$ the $\chi^2$ distribution becomes normal. One further fact, not obvious from the diagrams, is that the variance of the $\chi^2$ distribution is $2k$.



$k = 4$

$k = 16$

$k = 64$

$k = 256$
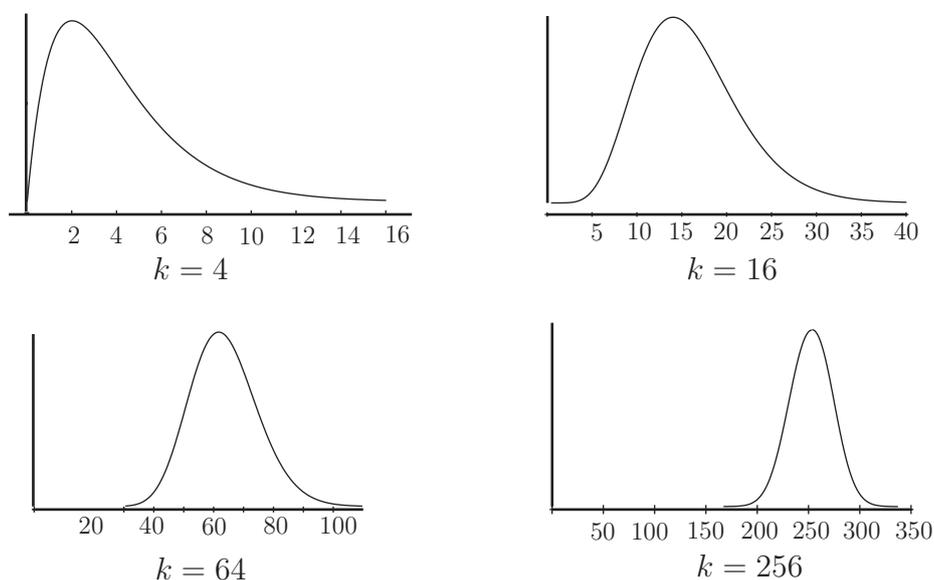
**Figure 2**

A summary is given in the following Key Point.

---

### Key Point 3

The $\chi^2$ distribution, defined by the probability density function

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

has mean $k$ and variance $2k$ and as $k \to \infty$ the limiting form of the distribution is normal.

---

## Degrees of freedom

A formal definition of the term 'degrees of freedom' is that it is the 'number of independent comparisons that can be made among the elements of a sample.' Textbooks on statistics e.g. *Applied Statistics and Probability for Engineers* by Montgomery and Runger (Wiley) often give this formal definition. The number of degrees of freedom is usually represented by the Greek symbol $\nu$ pronounced 'nu'. The following explanations of the concept should be helpful.

### Explanation 1

If we have a sample of $n$ values say $x_1, x_2, x_3 \cdots, x_n$ chosen from a population and we are trying to calculate the mean of the sample, we know that the sum of the deviations about the mean must be zero. Hence, the following constraint must apply to the observations.

$$\sum (x - \bar{x}) = 0$$

Once we calculate the values of $(x_1 - \bar{x}), (x_2 - \bar{x}), (x_3 - \bar{x}), \cdots (x_{n-1} - \bar{x})$ we can calculate the value of $(x_n - \bar{x})$ by using the constraint $\sum (x - \bar{x}) = 0$. We say that we have $n - 1$ degrees of freedom. The term 'degrees of freedom' may be thought of as the number of independent variables minus the number of constraints imposed.

### Explanation 2

A point in space which can move freely has three degrees of freedom since it can move *independently* in the $x, y$ and $z$ directions. If we now restrict the point so that it can only move along the straight line

$$\frac{x}{a} = \frac{y}{b} = \frac{z}{c}$$

then we have effectively imposed two constraints since the value of (say) $x$ determines the values of $y$ and $z$. In this situation, we say that the number of degrees of freedom is reduced from 3 to 1. That is, we have one degree of freedom.

A similar argument may be used to demonstrate that a point in three dimensional space which is restricted to move in a plane leads to a situation with two degrees of freedom.

---

### Key Point 4

The term 'degrees of freedom' may be thought of as the number of independent variables involved minus the number of constraints imposed.

---

Figure 3 shows a typical $\chi^2$ distribution and Table 1 at the end of this Workbook show the values of $\chi^2_{\alpha,\nu}$ for a variety of values of the area $\alpha$ and the number of degrees of freedom $\nu$. Notice that Table 1 gives the area values corresponding to the right-hand tail of the distribution which is shown shaded.
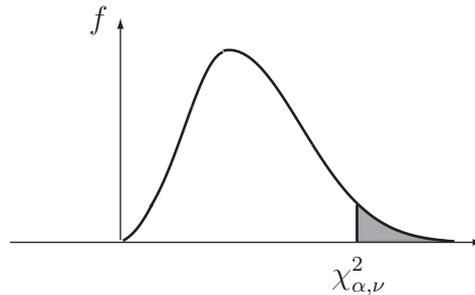
**Figure 3**

The $\chi^2_{\alpha,\nu}$ values for (say) right-hand area values of 5% are given by the column headed 0.05 while the $\chi^2_{\alpha,\nu}$ values for (say) left-hand area values of 5% are given by the column headed 0.95. Figure 4 shows the values of $\chi^2_{\alpha,\nu}$ for the two 5% tails when there are 5 degrees of freedom.



**Figure 4**

**Task**
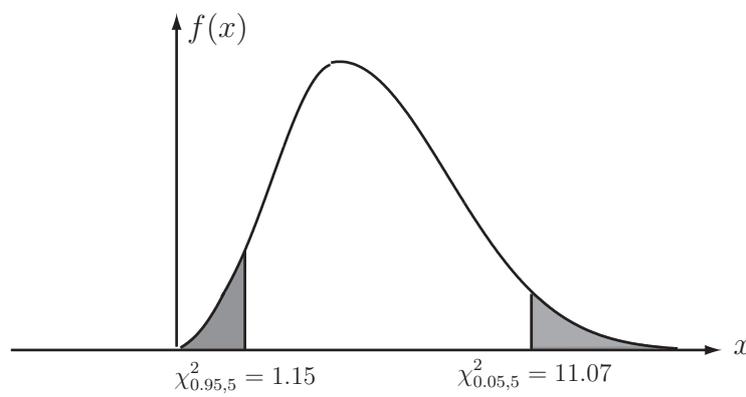
Use the percentage points of the $\chi^2$ distribution to find the appropriate values of $\chi^2_{\alpha,\nu}$ in the following cases.

(a) Right-hand tail of 10% and 7 degrees of freedom.

(b) Left-hand tail of 2.5% and 9 degrees of freedom.

(c) Both tails of 5% and 10 degrees of freedom.

(d) Both tails of 2.5% and 20 degrees of freedom.

**Your solution**

**Answer**
Using Table 1 and reading off the values directly gives:
(a) 12.02    (b) 2.70    (c) 3.94 and 18.31    (d) 9.59 and 34.17

## Constructing a confidence interval for the variance

We know that if $x_1, x_2, x_3, \cdots, x_n$ is a random sample taken from a normal population with mean $\mu$ and variance $\sigma^2$ and if the sample variance is denoted by $S^2$, the random variable

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

has a chi-squared distribution with $n-1$ degrees of freedom. This knowledge enables us to construct a confidence interval as follows.

Firstly, we decide on a level of confidence, say, for the sake of illustration, 95%. This means that we need two 2.5% tails.

Secondly, we know that we have $n-1$ degrees of freedom so that the value of $X^2$ will lie between the left-tail value of $\chi^2_{0.975,n-1}$ and the right-tail value of $\chi^2_{0.025,n-1}$. If we know the value of $n$ then we can easily read off these values from the $\chi^2$ tables.

The confidence interval is developed as shown below.
We have

$$\chi^2_{0.025,n-1} \leq X^2 \leq \chi^2_{0.975,n-1}$$

so that

$$\chi^2_{0.025,n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{0.975,n-1}$$

hence

$$\frac{1}{\chi^2_{0.975,n-1}} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{1}{\chi^2_{0.025,n-1}}$$

so that

$$\frac{(n-1)S^2}{\chi^2_{0.975,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0.025,n-1}}$$

Another way of stating the same result using probability directly is to say that

$$\mathsf{P}\left(\frac{(n-1)S^2}{\chi^2_{0.975,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0.025,n-1}}\right) = 0.95$$

Noting that $0.95 = 100(1-0.05)$ and that we are working with the right-hand tail values of the $\chi^2$ distribution, it is usual to generalize the above result as follows. Taking a general confidence level as $100(1-\alpha)\%$, (a 95% interval gives $\alpha = 0.05$), our confidence interval becomes

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

Note that the confidence interval for the standard deviation $\sigma$ is obtained by taking the appropriate square roots.

The following Key Point summarizes the development of this confidence interval.

## Key Point 5

If $x_1, x_2, x_3, \cdots, x_n$ is a random sample with variance $S^2$ taken from a normal population with variance $\sigma^2$ then a $100(1 - \alpha)\%$ confidence interval for $\sigma^2$ is

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

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the appropriate right-hand and left-hand values respectively of a chi-squared distribution with $n - 1$ degrees of freedom.

### Example 2

A random sample of 20 nominally measured 2mm diameter steel ball bearings is taken and the diameters are measured precisely. The measurements, in mm, are as follows:

2.02 1.94 2.09 1.95 1.98 2.00 2.03 2.04 2.08 2.07
1.99 1.96 1.99 1.95 1.99 1.99 2.03 2.05 2.01 2.03

Assuming that the diameters are normally distributed with unknown mean, $\mu$, and unknown variance $\sigma^2$,

(a) find a two-sided 95% confidence interval for the variance, $\sigma^2$;

(b) find a two-sided confidence interval for the standard deviation, $\sigma$.

#### Solution

From the data, we calculate $\sum x_i = 40.19$ and $\sum x_i^2 = 80.7977$. Hence

$$(n-1)S^2 = 80.7977 - \frac{40.19^2}{20} = 0.035895$$

There are 19 degrees of freedom and the critical values of the $\chi^2_{19}$-distribution are

$$\chi^2_{0.975,19} = 8.91 \qquad \text{and} \qquad \chi^2_{0.025,19} = 32.85$$

(a) the confidence interval for $\sigma^2$ is

$$\frac{0.035895}{32.85} < \sigma^2 < \frac{0.035895}{8.91} \equiv 1.0927 \times 10^{-3}\text{mm} < \sigma^2 \leq 4.0286 \times 10^{-3}\text{mm}$$

(b) the confidence interval for $\sigma$ is

$$\sqrt{1.0927 \times 10^{-3}} < \sigma \leq \sqrt{4.0286 \times 10^{-3}} \equiv 0.033\text{mm} < \sigma < 0.063 \text{ mm}$$

**Task**

In a typical car, bell housings are bolted to crankcase castings by means of a series of 13 mm bolts. A random sample of 12 bolt-hole diameters is checked as part of a quality control process and found to have a variance of 0.0013 mm$^2$.

(a) Construct the 95% confidence interval for the variance of the holes.

(b) Find the 95% confidence interval for the standard deviation of the holes.

State clearly any assumptions you make.

**Your solution**

**Answer**

Using the confidence interval formula developed, we know that the 95% confidence interval is

$$\frac{11 \times 0.0013}{\chi^2_{0.025,11}} \leq \sigma^2 \leq \frac{11 \times 0.0013}{\chi^2_{0.975,11}} \qquad \text{i.e.} \qquad \frac{11 \times 0.0013}{21.92} \leq \sigma^2 \leq \frac{11 \times 0.0013}{3.82}$$

(a) The 95% confidence interval for the variance is $0.0007 \leq \sigma^2 \leq 0.0037$ mm$^2$.

(b) The 95% confidence interval for the standard deviation is $0.0265 \leq \sigma \leq 0.0608$ mm.

We have assumed that the hole diameters are normally distributed.

# Exercises

1. Measurements are made on the lengths, in mm, of a sample of twenty wooden components for self-assembly furniture. Assume that these may be regarded as twenty independent observations from a normal distribution with unknown mean $\mu$ and unknown variance $\sigma^2$. The data are as follows.

   | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|
   | 581 | 580 | 581 | 577 | 580 | 581 | 577 | 579 | 579 | 578 |
   | 581 | 583 | 577 | 578 | 582 | 581 | 582 | 580 | 582 | 579 |

   Find a 95% confidence interval for the variance $\sigma^2$ and hence find a 95% confidence interval for the standard deviation $\sigma$.

2. A machine fills packets with powder. At intervals a sample of ten packets is taken and the packets are weighed. The ten weights may be regarded as a sample of ten independent observations from a normal distribution with unknown mean. Find limits $L$, $U$ such that the probability that $L < S^2 < U$ is 0.9 when the population variance is $\sigma^2 = 3.0$ and $S^2$ is the sample variance.

## Answers

1. From the data we calculate $\sum y_i = 11598$ and $\sum y_i^2 = 6725744$ and we have $n = 20$. Hence

   $$(n-1)s^2 = \sum (y_i - \bar{y})^2 = 6725744 - \frac{11598^2}{20} = 63.8$$

   The number of degrees of freedom is $n - 1 = 19$. We know that

   $$\chi^2_{0.975,19} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{0.025,19}$$

   with probability 0.95. So a 95% confidence interval for $\sigma^2$ is

   $$\frac{(n-1)s^2}{\chi^2_{0.025,19}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.975,19}}$$

   That is $\dfrac{63.8}{32.85} < \sigma^2 < \dfrac{63.8}{8.91}$ so $1.942 < \sigma^2 < 7.160$

   This gives a 95% confidence interval for $\sigma$: $1.394 < \sigma < 2.676$

2. There are $n - 1 = 9$ degrees of freedom. Now

   $$\begin{aligned} 0.9 &= \mathsf{P}\left( \chi^2_{0.05,9} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{0.95,9} \right) \\ &= \mathsf{P}\left( \frac{\chi^2_{0.05,9}\sigma^2}{n-1} < S^2 < \frac{\chi^2_{0.95,9}\sigma^2}{n-1} \right) \\ &= \mathsf{P}\left( \frac{3.33 \times 3.0}{9} < S^2 < \frac{16.92 \times 3.0}{9} \right) = \mathsf{P}(1.11 < S^2 < 5.64) \end{aligned}$$

   Hence $L = 1.11$ and $U = 5.64$.

**Table 1:** Percentage Points $\chi^2_{\alpha,\nu}$ of the $\chi^2$ distribution
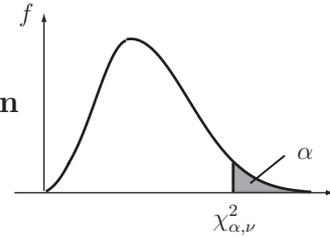


| $\alpha$ | 0.995 | 0.990 | 0.975 | 0.950 | 0.900 | 0.500 | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\nu$ | | | | | | | | | | | |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.45 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.01 | 0.02 | 0.05 | 0.01 | 0.21 | 1.39 | 4.61 | 5.99 | 7.38 | 9.21 | 10.60 |
| 3 | 0.07 | 0.11 | 0.22 | 0.35 | 0.58 | 2.37 | 6.25 | 7.81 | 9.35 | 11.34 | 12.28 |
| 4 | 0.21 | 0.30 | 0.48 | 0.71 | 1.06 | 3.36 | 7.78 | 9.49 | 11.14 | 13.28 | 14.86 |
| 5 | 0.41 | 0.55 | 0.83 | 1.15 | 1.61 | 4.35 | 9.24 | 11.07 | 12.83 | 15.09 | 16.75 |
| 6 | 0.68 | 0.87 | 1.24 | 1.64 | 2.20 | 5.35 | 10.65 | 12.59 | 14.45 | 16.81 | 18.55 |
| 7 | 0.99 | 1.24 | 1.69 | 2.17 | 2.83 | 6.35 | 12.02 | 14.07 | 16.01 | 18.48 | 20.28 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 7.34 | 13.36 | 15.51 | 17.53 | 20.09 | 21.96 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 8.34 | 14.68 | 16.92 | 19.02 | 21.67 | 23.59 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 9.34 | 15.99 | 18.31 | 20.48 | 23.21 | 25.19 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 10.34 | 17.28 | 19.68 | 21.92 | 24.72 | 26.76 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 11.34 | 18.55 | 21.03 | 23.34 | 26.22 | 28.30 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 12.34 | 19.81 | 22.36 | 24.74 | 27.69 | 29.82 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 13.34 | 21.06 | 23.68 | 26.12 | 29.14 | 31.32 |
| 15 | 4.60 | 5.23 | 6.27 | 7.26 | 8.55 | 14.34 | 22.31 | 25.00 | 27.49 | 30.58 | 32.80 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 15.34 | 23.54 | 26.30 | 28.85 | 31.00 | 34.27 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.09 | 16.34 | 24.77 | 27.59 | 30.19 | 33.41 | 35.72 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.87 | 17.34 | 25.99 | 28.87 | 31.53 | 34.81 | 37.16 |
| 19 | 6.84 | 7.63 | 8.91 | 10.12 | 11.65 | 18.34 | 27.20 | 30.14 | 32.85 | 36.19 | 38.58 |
| 20 | 7.43 | 8.26 | 9.59 | 10.85 | 12.44 | 19.34 | 28.41 | 31.41 | 34.17 | 37.57 | 40.00 |
| 21 | 8.03 | 8.90 | 10.28 | 11.59 | 13.24 | 20.34 | 29.62 | 32.67 | 35.48 | 38.93 | 41.40 |
| 22 | 8.64 | 9.54 | 10.98 | 12.34 | 14.04 | 21.34 | 30.81 | 33.92 | 36.78 | 40.29 | 42.80 |
| 23 | 9.26 | 10.20 | 11.69 | 13.09 | 14.85 | 22.34 | 32.01 | 35.17 | 38.08 | 41.64 | 44.18 |
| 24 | 9.89 | 10.86 | 12.40 | 13.85 | 15.66 | 23.34 | 33.20 | 36.42 | 39.36 | 42.98 | 45.56 |
| 25 | 10.52 | 11.52 | 13.12 | 14.61 | 16.47 | 24.34 | 34.28 | 37.65 | 40.65 | 44.31 | 46.93 |
| 26 | 11.16 | 12.20 | 13.84 | 15.38 | 17.29 | 25.34 | 35.56 | 38.89 | 41.92 | 45.64 | 48.29 |
| 27 | 11.81 | 12.88 | 14.57 | 16.15 | 18.11 | 26.34 | 36.74 | 40.11 | 43.19 | 46.96 | 49.65 |
| 28 | 12.46 | 13.57 | 15.31 | 16.93 | 18.94 | 27.34 | 37.92 | 41.34 | 44.46 | 48.28 | 50.99 |
| 29 | 13.12 | 14.26 | 16.05 | 17.71 | 19.77 | 28.34 | 39.09 | 42.56 | 45.72 | 49.59 | 52.34 |
| 30 | 13.79 | 14.95 | 16.79 | 18.49 | 20.60 | 29.34 | 40.26 | 43.77 | 46.98 | 50.89 | 53.67 |
| 40 | 20.71 | 22.16 | 24.43 | 26.51 | 29.05 | 39.34 | 51.81 | 55.76 | 59.34 | 63.69 | 66.77 |
| 50 | 27.99 | 29.71 | 32.36 | 34.76 | 37.69 | 49.33 | 63.17 | 67.50 | 71.42 | 76.15 | 79.49 |
| 60 | 35.53 | 37.48 | 40.48 | 43.19 | 46.46 | 59.33 | 74.40 | 79.08 | 83.30 | 88.38 | 91.95 |
| 70 | 43.28 | 45.44 | 48.76 | 51.74 | 55.33 | 69.33 | 85.53 | 90.53 | 95.02 | 100.42 | 104.22 |
| 80 | 51.17 | 53.54 | 57.15 | 60.39 | 64.28 | 79.33 | 96.58 | 101.88 | 106.63 | 112.33 | 116.32 |
| 90 | 59.20 | 61.75 | 65.65 | 69.13 | 73.29 | 89.33 | 107.57 | 113.14 | 118.14 | 124.12 | 128.30 |
| 100 | 67.33 | 70.06 | 74.22 | 77.93 | 82.36 | 99.33 | 118.50 | 124.34 | 129.56 | 135.81 | 140.17 |