Chapter 6
INTERVALS Statement

Chapter Table of Contents

Chapter 6
INTERVALS Statement

Overview

The INTERVALS statement tabulates various statistical intervals for selected process variables. The types of intervals you can request include

- approximate simultaneous prediction intervals for future observations
- prediction intervals for the mean of future observations
- approximate statistical tolerance intervals that contain at least a specified proportion of the population
- confidence intervals for the population mean
- prediction intervals for the standard deviation of future observations
- confidence intervals for the population standard deviation

These intervals are computed assuming the data are sampled from a normal population. See Hahn and Meeker (1991) for a detailed discussion of these intervals.

You can use options in the INTERVALS statement to

- specify which intervals to compute
- provide probability or confidence levels for intervals
- suppress printing of output tables
- create an output data set containing interval information
- specify interval type (one-sided lower, one-sided upper, or two-sided)

Getting Started

This section introduces the INTERVALS statement with simple examples that illustrate commonly used options. Complete syntax for the INTERVALS statement is presented in the “Syntax” section on page 222.

Computing Statistical Intervals

See CAPINT1 in the SAS/QC Sample Library

The following statements create the data set CANS, which contains measurements (in ounces) of the fluid weights of 100 drink cans. The filling process is assumed to be in statistical control.

```
data cans;
   label weight = 'Fluid Weight (ounces)';
   input weight @@;
datalines;
12.07 12.02 12.00 12.01 11.98 11.96 12.04 12.05 12.01 11.97
12.03 12.03 12.00 12.04 11.96 12.02 12.06 12.00 12.02 11.91
12.05 11.98 11.91 12.01 12.06 12.02 12.05 11.90 12.07 11.98
12.02 12.11 12.00 11.99 11.95 11.98 12.05 12.00 12.10 12.04
12.06 12.04 11.99 12.06 11.99 12.07 11.96 11.97 12.00 11.97
12.09 11.99 11.95 11.99 11.99 11.96 11.94 12.03 12.09 12.03
11.99 12.00 12.05 12.04 12.05 12.01 11.97 11.93 12.00 11.97
12.13 12.07 12.00 11.96 11.99 11.97 12.05 11.94 11.99 12.02
11.95 11.99 11.91 12.06 12.03 12.06 12.05 12.04 12.03 11.98
12.05 12.05 12.11 11.96 12.00 11.96 11.96 12.00 12.01 11.98
;
```

Note that this data set is introduced in “Computing Descriptive Statistics” on page 9 of Chapter 1, “PROC CAPABILITY and General Statements.” The analysis in that section provides evidence that the weight measurements are normally distributed.

By default, the INTERVALS statement computes and prints the six intervals described in the entry for the METHODS= option on page 223. The following statements tabulate these intervals for the variable WEIGHT:

```
title 'Statistical Intervals for Fluid Weight';
proc capability data=cans noprint;
   intervals weight;
run;
```

The intervals are displayed in Figure 6.1 on page 219.

Statistical Intervals for Fluid Weight

Two-Sided Statistical Intervals for weight Assuming Normality

Approximate Prediction Interval Containing All of k Future Observations

Confidence	k	Prediction Limits	
99.00%	1	11.89	12.13
99.00%	2	11.87	12.14
99.00%	3	11.87	12.15
95.00%	1	11.92	12.10
95.00%	2	11.90	12.12
95.00%	3	11.89	12.12
90.00%	1	11.93	12.09
90.00%	2	11.92	12.10
90.00%	3	11.91	12.11

Approximate Prediction Interval Containing the Mean of k Future Observations

Confidence	k	Prediction Limits	
99.00%	1	11.89	12.13
99.00%	2	11.92	12.10
99.00%	3	11.94	12.08
95.00%	1	11.92	12.10
95.00%	2	11.94	12.08
95.00%	3	11.95	12.06
90.00%	1	11.93	12.09
90.00%	2	11.95	12.06
90.00%	3	11.96	12.05

Approximate Tolerance Interval Containing At Least Proportion p of the Population

Confidence	p	Tolerance Limits	
99.00%	0.900	11.92	12.10
99.00%	0.950	11.90	12.12
99.00%	0.990	11.86	12.15
95.00%	0.900	11.92	12.10
95.00%	0.950	11.90	12.11
95.00%	0.990	11.87	12.15
90.00%	0.900	11.92	12.09
90.00%	0.950	11.91	12.11
90.00%	0.990	11.88	12.14

Figure 6.1. Statistical Intervals for WEIGHT (*continued*)

```
         Two-Sided Statistical Intervals for weight Assuming Normality

                          Confidence Limits Containing
                                    the Mean

                         Confidence   Confidence Limits

                           99.00%      11.997    12.022
                           95.00%      12.000    12.019
                           90.00%      12.002    12.017


                         Prediction Interval Containing
                           the Standard Deviation of
                             k Future Observations

                      Confidence     k   Prediction Limits

                        99.00%       2    0.0003    0.1348
                        99.00%       3    0.0033    0.1110

                        95.00%       2    0.0015    0.1069
                        95.00%       3    0.0075    0.0919

                        90.00%       2    0.0030    0.0932
                        90.00%       3    0.0106    0.0825


                          Confidence Limits Containing
                             the Standard Deviation

                          Confidence   Confidence Limits

                            99.00%      0.040    0.057
                            95.00%      0.041    0.055
                            90.00%      0.042    0.053
```

Figure 6.2. Statistical Intervals for WEIGHT(continued from page 219)

Computing One-Sided Lower Prediction Limits

See CAPINT1 in the SAS/QC Sample Library

You can specify options after the slash (/) in the INTERVALS statement to control the computation and printing of intervals. The following statements produce a table of one-sided lower prediction limits for the mean, which is displayed in Figure 6.3:

```
title 'Statistical Intervals for Fluid Weight';
proc capability data=cans noprint;
   intervals weight / methods = 1 2
                      type    = lower;
run;
```

The METHODS= option specifies which intervals to compute, and the TYPE= option requests one-sided lower limits. All the options available in the INTERVALS statement are listed in "Summary of Options" on page 222 and are described in "Dictionary of Options" on page 223.

Statistical Intervals for Fluid Weight

One-Sided Lower Statistical Intervals for weight Assuming Normality

Approximate Prediction Limit For All of k Future Observations

Confidence	k	Lower Limit
99.00%	1	11.90
99.00%	2	11.89
99.00%	3	11.88
95.00%	1	11.93
95.00%	2	11.92
95.00%	3	11.91
90.00%	1	11.95
90.00%	2	11.93
90.00%	3	11.92

Approximate Prediction Limit For the Mean of k Future Observations

Confidence	k	Lower Limit
99.00%	1	11.90
99.00%	2	11.93
99.00%	3	11.94
95.00%	1	11.93
95.00%	2	11.95
95.00%	3	11.96
90.00%	1	11.95
90.00%	2	11.97
90.00%	3	11.97

Figure 6.3. One-Sided Lower Prediction Limits for the Mean

Syntax

The syntax for the INTERVALS statement is as follows:

INTERVALS <*variables* > < **/** *options* >**;**

You can specify INTERVAL as an alias for INTERVALS. You can use any number of INTERVALS statements in the CAPABILITY procedure. The components of the INTERVALS statement are described as follows.

variables
gives a list of variables for which to compute intervals. If you specify a VAR statement, the *variables* must also be listed in the VAR statement. Otherwise, the *variables* can be any numeric variable in the input data set. If you do not specify a list of *variables*, then by default the INTERVALS statement computes intervals for all variables in the VAR statement (or all numeric variables in the input data set if you do not use a VAR statement).

options
alter the defaults for computing and printing intervals and for creating output data sets.

Summary of Options

The following tables list the INTERVALS statement options by function. For complete descriptions, see "Dictionary of Options" on page 223.

Table 6.1. INTERVAL Statement Options

ALPHA=*value-list*	lists probability or confidence levels associated with the intervals
K=*value-list*	lists values of k for prediction intervals
METHODS=*indices*	specifies which intervals are computed
NOPRINT	suppresses the output tables
OUTINTERVALS=*SAS-data-set*	specifies an output data set containing interval information
P=*value-list*	lists values of p for tolerance intervals
TYPE=*keyword*	specifies the type of intervals (one-sided lower, one-sided upper, or two-sided)

Dictionary of Options

The following entries provide detailed descriptions of *options* in the INTERVALS statement.

ALPHA=*value-list*

specifies values of α, the probability or confidence associated with the interval. For example, the following statements tabulate the default intervals at probability or confidence levels of $\alpha = 0.05$, $\alpha = 0.10$, $\alpha = 0.15$, and $\alpha = 0.20$:

```
proc capability data=steel;
   intervals width / alpha = 0.05 0.10 0.15 0.20;
run;
```

Note that some references use $\gamma = 1 - \alpha$ to denote probability or confidence levels. Values for the ALPHA= option must be between 0.00001 to 0.99999. By default, values of 0.01, 0.05, and 0.10 are used.

K=*value-list*

lists values of k for prediction intervals. Default *values* of 1, 2, and 3 are used for the prediction interval for k future observations and for the prediction interval for the mean of k future observations. Default *values* of 2 and 3 are used for the prediction interval for the standard deviation of k future observations. The *values* must be integers.

METHODS=*indices*
METHOD=*indices*

specifies which intervals are computed. The *indices* can range from 1 to 6, and they correspond to the intervals described in Table 6.2.

Table 6.2. Intervals Computed for METHOD=*Index*

Index	Statistical Interval
1	approximate simultaneous prediction interval for k future observations
2	prediction interval for the mean of k future observations
3	approximate statistical tolerance interval that contains at least proportion p of the population
4	confidence interval for the population mean
5	prediction interval for the standard deviation of k future observations
6	confidence interval for the population standard deviation

For example, the following statements tabulate confidence limits for the population mean (METHOD=4) and confidence limits for the population standard deviation (METHOD=6):

```
proc capability data=steel;
   intervals width / methods=4 6;
run;
```

Formulas for the intervals are given in "Methods for Computing Statistical Intervals" on page 225. By default, the procedure computes all six intervals.

NOPRINT

suppresses the tables produced by default. This option is useful when you only want to save the interval information in an OUTINTERVALS= data set.

OUTINTERVALS=*SAS-data-set*
OUTINTERVAL=*SAS-data-set*
OUTINT=*SAS-data-set*

specifies an output SAS data set containing the intervals and related information. For example, the following statements create a data set named INTS containing intervals for the variable WIDTH:

```
proc capability data=steel;
   intervals width / outintervals=ints;
run;
```

See "OUTINTERVALS= Data Set" on page 228 for details.

P=*value-list*

lists values of p for the tolerance intervals. These values must be between 0.00001 to 0.99999. Note that the P= option applies only to the tolerance intervals (METHODS=3). By default, values of 0.90, 0.95, and 0.99 are used.

TYPE=LOWER | UPPER | TWOSIDED

determines whether the intervals computed are one-sided lower, one-sided upper, or two-sided intervals, respectively. See "Computing One-Sided Lower Prediction Limits" on page 220 for an example. The default interval type is TWOSIDED.

Details

This section provides details on the following topics:

- formulas for statistical intervals
- OUTINTERVALS= data sets

Methods for Computing Statistical Intervals

The formulas for statistical intervals given in this section use the following notation:

Notation	Definition
n	number of nonmissing values for a variable
$\bar{X}$	mean of variable
s	standard deviation of variable
z_α	$100\alpha^{\text{th}}$ percentile of the standard normal distribution
$t_\alpha(\nu)$	$100\alpha^{\text{th}}$ percentile of the central t distribution with ν degrees of freedom
$t'_\alpha(\delta, \nu)$	$100\alpha^{\text{th}}$ percentile of the noncentral t distribution with noncentrality parameter δ and ν degrees of freedom
$F_\alpha(\nu_1, \nu_2)$	$100\alpha^{\text{th}}$ percentile of the F distribution with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator
$\chi^2_\alpha(\nu)$	$100\alpha^{\text{th}}$ percentile of the χ^2 distribution with ν degrees of freedom.

The values of the variable are assumed to be independent and normally distributed. The intervals are computed using the degrees of freedom as the divisor for the standard deviation s. This divisor corresponds to the default of VARDEF=DF in the PROC CAPABILITY statement. If you specify another value for the VARDEF= option, intervals are not computed.

You select the intervals to be computed with the METHODS= option. The next six sections give computational details for each of the METHODS= options.

METHODS=1

This requests an approximate simultaneous prediction interval for k future observations. Two-sided intervals are computed using the conservative approximations

$$\text{Lower Limit} = \bar{X} - t_{1-\frac{\alpha}{2k}}(n-1)s\sqrt{1+\frac{1}{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\frac{\alpha}{2k}}(n-1)s\sqrt{1+\frac{1}{n}}$$

One-sided limits are computed using the conservative approximation

$$\text{Lower Limit} = \bar{X} - t_{1-\frac{\alpha}{k}}(n-1)s\sqrt{1+\frac{1}{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\frac{\alpha}{k}}(n-1)s\sqrt{1+\frac{1}{n}}$$

Hahn (1970c) states that these approximations are satisfactory except for combinations of small n, large k, and large α. Refer also to Hahn (1969 and 1970a) and Hahn and Meeker (1991).

METHODS=2

This requests a prediction interval for the mean of k future observations. Two-sided intervals are computed as

$$\text{Lower Limit} = \bar{X} - t_{1-\frac{\alpha}{2}}(n-1)s\sqrt{\frac{1}{k}+\frac{1}{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\frac{\alpha}{2}}(n-1)s\sqrt{\frac{1}{k}+\frac{1}{n}}$$

One-sided limits are computed as

$$\text{Lower Limit} = \bar{X} - t_{1-\alpha}(n-1)s\sqrt{\frac{1}{k}+\frac{1}{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\alpha}(n-1)s\sqrt{\frac{1}{k}+\frac{1}{n}}$$

METHODS=3

This requests an approximate statistical tolerance interval that contains at least proportion p of the population. Two-sided intervals are approximated by

$$\text{Lower Limit} = \bar{X} - g(p\,;n\,;1-\alpha)s$$

$$\text{Upper Limit} = \bar{X} + g(p\,;n\,;1-\alpha)s$$

where $g(p\,;n\,;1-\alpha) = z_{\frac{1+p}{2}}(1+\frac{1}{2n})\sqrt{\frac{n-1}{\chi^2_\alpha(n-1)}}$.

Exact one-sided limits are computed as

$$\text{Lower Limit} = \bar{X} - g'(p\,;n\,;1-\alpha)s$$

$$\text{Upper Limit} = \bar{X} + g'(p\,;n\,;1-\alpha)s$$

where $g'(p\,;n\,;1-\alpha) = \frac{1}{\sqrt{n}}t'_{1-\alpha}(z_p\sqrt{n}, n-1)$.

In some cases (for example, if $z_p\sqrt{n}$ is large), $g'(p\,;n\,;1-\alpha)$ is approximated by

$$\frac{1}{a}\left(z_p + \sqrt{z_p^2 - ab}\,\right)$$

where $a = 1 - \frac{z_{1-\alpha}^2}{2(n-1)}$ and $b = z_p^2 - \frac{z_{1-\alpha}^2}{n}$.

Hahn (1970b) states that this approximation is "poor for very small n, especially for large p and large $1-\alpha$, and is not advised for $n < 8$." Refer also to Hahn and Meeker (1991).

METHODS=4

This requests a confidence interval for the population mean. Two-sided intervals are computed as

$$\text{Lower Limit} = \bar{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$$

One-sided limits are computed as

$$\text{Lower Limit} = \bar{X} - t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}}$$

$$\text{Upper Limit} = \bar{X} + t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}}$$

METHODS=5

This requests a prediction interval for the standard deviation of k future observations. Two-sided intervals are computed as

$$\text{Lower Limit} = s\left(F_{1-\frac{\alpha}{2}}(n-1,k-1)\right)^{-\frac{1}{2}}$$

$$\text{Upper Limit} = s\left(F_{1-\frac{\alpha}{2}}(k-1,n-1)\right)^{\frac{1}{2}}$$

One-sided limits are computed as

$$\text{Lower Limit} = s\left(F_{1-\alpha}(n-1,k-1)\right)^{-\frac{1}{2}}$$

$$\text{Upper Limit} = s\left(F_{1-\alpha}(k-1,n-1)\right)^{\frac{1}{2}}$$

METHODS=6

This requests a confidence interval for the population standard deviation. Two-sided intervals are computed as

$$\text{Lower Limit} = s\sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}$$

$$\text{Upper Limit} = s\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}}(n-1)}}$$

One-sided limits are computed as

$$\text{Lower Limit} = s\sqrt{\frac{n-1}{\chi^2_{1-\alpha}(n-1)}}$$

$$\text{Upper Limit} = s\sqrt{\frac{n-1}{\chi^2_{\alpha}(n-1)}}$$

OUTINTERVALS= Data Set

Each INTERVALS statement can create an output data set specified with the OUTINTERVALS= option. The OUTINTERVALS= data set contains statistical intervals and related parameters.

The number of observations in the OUTINTERVALS= data set depends on the number of variables analyzed, the number of tests specified, and the results of the tests. The OUTINTERVALS= data set is constructed as follows:

- The OUTINTERVALS= data set contains a group of observations for each variable analyzed.
- Each group contains one or more observations for each interval you specify with the METHODS= option. The actual number depends upon the number of combinations of the ALPHA=, K=, and P= values.

The following variables are saved in the OUTINTERVALS= data set:

Variable	Description
ALPHA	value of α associated with the intervals
_K_	value of K= for the prediction intervals
LOWER	lower endpoint of interval
METHOD	interval index $(1\breve{\ }6)$
_P_	value of P= for the tolerance intervals
TYPE	type of interval (ONESIDED or TWOSIDED)
UPPER	upper endpoint of interval
VAR	variable name

If you use a BY statement, the BY variables are also saved in the OUTINTERVALS= data set.

ODS Tables

The following table summarizes the ODS tables that you can request with the INTERVALS statement.

Table 6.3. ODS Tables Produced with the INTERVALS Statement

Table Name	Description	Option
Intervals1	prediction interval for future observations	METHODS=1
Intervals2	prediction interval for mean	METHODS=2
Intervals3	tolerance interval for proporation of population	METHODS=3
Intervals4	confidence limits for mean	METHODS=4
Intervals5	prediction interval for standard deviation	METHODS=5
Intervals6	confidence limits for standard deviation	METHODS=6

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC® User's Guide, Version 8

ISBN 1–58025–493–4

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999