SAS/ETS® 13.2 User's Guide

The TIMEID Procedure

This document is an individual chapter from *SAS/ETS® 13.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2014. *SAS/ETS® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

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

August 2014

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our offerings, visit **support.sas.com/bookstore** or call 1-800-727-3228.

Chapter 31
The TIMEID Procedure

Contents

Overview: TIMEID Procedure

The TIMEID procedure evaluates a variable in an input data set for its suitability as a time ID variable in SAS procedures and solutions that are used for time series analysis. PROC TIMEID assesses how well a time interval specification fits SAS date or datetime values, or observation numbers used to index a time series. The time interval used in this analysis can be either specified explicitly as input to PROC TIMEID or inferred by the procedure based on values of the time ID variable. The TIMEID procedure produces diagnostic information in the form of data sets and ODS tabular and plotted output. These diagnostic results summarize characteristics of the time ID variable that can help determine its use as an index in other time series procedures and solutions.

PROC TIMEID is intended for use as a tool to either identify the time interval of a variable or prepare problematic data sets for use in subsequent time series analyses. In particular, this procedure can be used to investigate inconsistencies between time ID values and the ID statement options used in other SAS procedures and solutions.

Getting Started: TIMEID Procedure

When a data set contains a time ID variable with corrupted, missing, or duplicate values, PROC TIMEID can help isolate and identify these problematic observations. For a data set with a small number of ID variable anomalies and a known time interval, a graphical depiction of the problem areas can be created using the following statements:

```
proc timeid data=<input-dataset> plot=values;
   id <time-ID-variable>  interval=<frequency>;
run;
```

For larger data sets whose quality is unknown, it can be useful to get a general overview of the relative number of observations with problematic time ID values. The following statements graphically summarize the prevalence of anomalous time ID values:

```
proc timeid data=<input-dataset> plot=(intervalcounts offsets spans);
   id <time-ID-variable>  interval=<frequency>;
run;
```

When prior knowledge of the time interval that separates observations is incomplete, PROC TIMEID can be used to infer the interval by omitting the INTERVAL= option from the ID statement as in the following statements:

```
proc timeid data=<input-dataset> outinterval=<output-dataset>;
   id <time-ID-variable>;
run;
```

Syntax: TIMEID Procedure

The TIMEID procedure uses the following statements:

PROC TIMEID *options* **;**
BY *variables* **;**
ID *variable* < *options* > **;**

Functional Summary

The statements and options that control the TIMEID procedure are summarized in Table 31.1.

Table 31.1 Syntax Summary

Description	Statement	Option
Statements		
Specifies data sets and options	PROC TIMEID	
Specifies BY-group processing	BY	
Specifies the time ID variable	ID	
Data Set Options		
Specifies the input data set	PROC TIMEID	DATA=
Specifies the maximum number of ID values to analyze	PROC TIMEID	NBYOBS=
Specifies the output frequency count data set	PROC TIMEID	OUTFREQ=
Specifies the output interval data set	PROC TIMEID	OUTINTERVAL=
Specifies the detailed output interval data set	PROC TIMEID	OUTINTERVALDETAILS=
Time ID Options		
Specifies the interval alignment	ID	ALIGN=
Specifies that duplicate time ID values can be present in DATA= data set	ID	DUPLICATES
Specifies the time interval between observations	ID	INTERVAL=
Specifies that time ID variable values are not sorted	ID	NOTSORTED
Printing and Plotting Options		
Specifies the time ID format	ID	FORMAT=
Specifies the types of graphical output	PROC TIMEID	PLOT=
Specifies the types of printed output	PROC TIMEID	PRINT=
Miscellaneous Options		
Limits error and warning messages	PROC TIMEID	MAXERROR=

PROC TIMEID Statement

PROC TIMEID *options* **;**

The following options can be used in the PROC TIMEID statement:

DATA=*SAS-data-set*
names the SAS data set that contains the input data for the procedure. If the DATA= option is not specified, the most recently created SAS data set is used.

MAXERROR=*number*
limits the number of warning and error messages produced during the execution of the procedure to the specified value. The default is MAXERRORS=50. This option is particularly useful in BY-group processing where it can be used to suppress recurring messages.

NBYOBS=*number*
limits the number of observations that are used to analyze the time ID variable. The NBYOBS= option should be used instead of the OBS= data set option when BY variables are specified. The NBYOBS= option excludes observations from incomplete BY groups in the analysis. This option guarantees that any truncation of the DATA= data set occurs at a BY-group boundary. Only BY groups that are completely contained within the first *number* of observations are processed. When the NBYOBS= option is omitted, all observations are processed.

OUTFREQ=*SAS-data-set*
names the output data set to contain the frequency counts of each unique value of the time ID variable. The frequency counts are performed on time ID values that are recorded in the DATA= data set. The time ID values are not aligned with respect to an interval prior to computation of the frequency counts. See the section "OUTFREQ= Data Set" on page 2206 for details.

OUTINTERVAL=*SAS-data-set*
names the output data set to contain the time ID interval information that is summarized across all BY groups in the DATA= data set. See the section "OUTINTERVAL= Data Set" on page 2206 for details.

OUTINTERVALDETAILS=*SAS-data-set*
names the output data set to contain the time ID interval information for each BY group. See the section "OUTINTERVALDETAILS= Data Set" on page 2207 for details.

PLOT*(global-option)***=***request-option* | *(request-options)*
specifies the graphical output desired. By default, the TIMEID procedure produces no graphical output. The following *global-options* are available:

UNPACK | UNPACKPANELS suppresses paneling.

By default, multiple plots can appear in some output panels. Specify UNPACKPANELS to get each plot in a separate panel. The following plot *request-options* are available:

COUNTS | INTCNTS | INTERVALCOUNTS
plots a histogram of the time ID interval counts.

OFFSETS plots a histogram of the time offsets for the time ID values.

PERIODS | SPANS	plots a histogram of the spans between adjacent time ID values.
VALUES	plots a panel of the counts, offsets, and spans for each of the time ID values.
ALL	is equivalent to specifying PLOT=(INTERVALCOUNTS SPANS OFFSETS VALUES).

See the section "Time ID Diagnostics" on page 2203 for details.

PRINT=*option* | **(***options***)**

specifies the printed output desired. By default, the TIMEID procedure produces no printed output. The following printing options are available:

COUNTS | INTCNTS | INTERVALCOUNTS	prints a table that contains the counts of time ID values per interval.
INTERVAL	prints a summary of information about the time interval.
OFFSETS	prints a table that contains the time offsets for the time ID values.
PERIODS | SPANS	prints tables that contain statistics on the spans between adjacent time ID values.
VALUES	prints tables that contain offset span and count information for the time ID values.
ALL	is equivalent to specifying PRINT=(INTERVALCOUNTS SPANS INTERVAL OFFSETS VALUES).

See the section "Time ID Diagnostics" on page 2203 for details.

BY Statement

BY *variables* **;**

A BY statement can be used with PROC TIMEID to obtain separate analyses for groups of observations defined by the BY variables.

When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the option NOTSORTED or DESCENDING in the BY statement for the TIMESERIES procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure.

For more information about the BY statement, see *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide.*

ID Statement

ID *variable* < *options* > **;**

The ID statement names a numeric variable that identifies observations in the input and output data sets. The ID variable's values are assumed to be SAS date or datetime values. The ID statement options specify how the time ID values are spaced and aligned relative to a SAS date or datetime interval. The INTERVAL= option specifies the fundamental spacing that is used as the basis for counting intervals, offsets, and spans in the data. Specification of the ID variable in an ID statement is required.

ALIGN=*alignment*

specifies the alignment of the identifying SAS date or datetime that is used to represent intervals. The value of the ALIGN= option is used in the analysis of the time ID variable. The ALIGN= option accepts the following values: BEGINNING | BEG | B, MIDDLE | MID | M, ENDING | END | E, and INFER. For example, ALIGN=BEGIN specifies that the identifying date for the interval is the beginning date in the interval. If the ALIGN= option is not specified, then the default alignment is BEGIN. ALIGN=INFER specifies that the alignment of values within time intervals be inferred from the time ID values.

DUPLICATES

specifies that multiple observations in the DATA= data set can fall within the same time interval as defined by the time ID variable. When this option is omitted and multiple time ID values are encountered in a single time interval, error messages are written to the SAS log.

FORMAT=*format*

specifies the SAS format used for time ID values in the data sets and in printed and plotted output that is generated by PROC TIMEID. If the FORMAT= option is not specified, the format applied to the input time ID variable is used. If neither of these formats is specified, the format is inferred from the INTERVAL= option.

INTERVAL=*interval*

specifies the proposed time interval and shift that describe the time ID values in the input data set. See Chapter 4, "Date Intervals, Formats, and Functions," for more information about the intervals that can be specified. See the section "Time ID Diagnostics" on page 2203 for more information about how the INTERVAL= option determines the nature of diagnostic information reported by the TIMEID procedure.

If no interval is specified, the procedure attempts to infer an interval from the input time ID values. See the section "Inferring Time Intervals and Alignments" on page 2205 for details about how the time interval is inferred.

NOTSORTED

specifies that the observations in the DATA= data set are not sorted by the time ID variable. When this option is omitted, error messages are generated for time ID values that are not sorted in ascending order.

Details: TIMEID Procedure

Time ID Diagnostics

For a specified time interval, PROC TIMEID decomposes the raw time ID values in an input data set into the following three quantities, whose values are represented by nonnegative integers at each unique time ID value in the input series:

interval counts the number of observations that share each time interval in the data set.

offsets the numerical difference between a time ID value and the aligned value for that time interval. The unit of measure used to express this distance is days for date values and seconds for datetime values. The offset is computed for each time ID value, t_i, by using the following SAS expression:

$$\text{offset}_i = t_i - \text{INTNX}(\text{interval}, t_i, 0, \text{alignment})$$

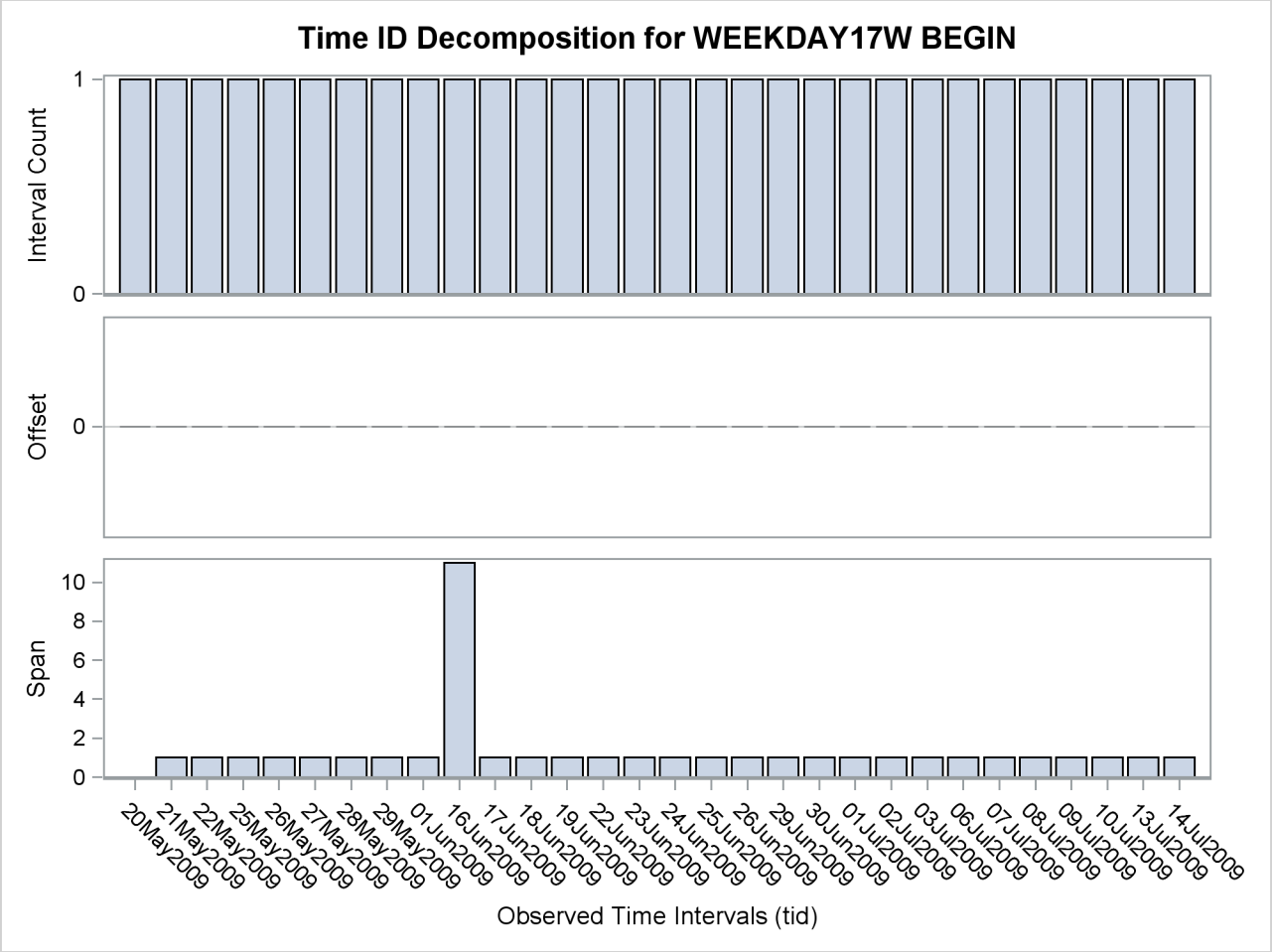
spans the number of intervals between each time ID value and the previous time ID value. The spans value is equivalent to the number returned by the following SAS expression:

$$\text{spans}_i = \text{INTCK}(\text{interval}, t_{i-1}, t_i)$$

Diagnostic Output Representation

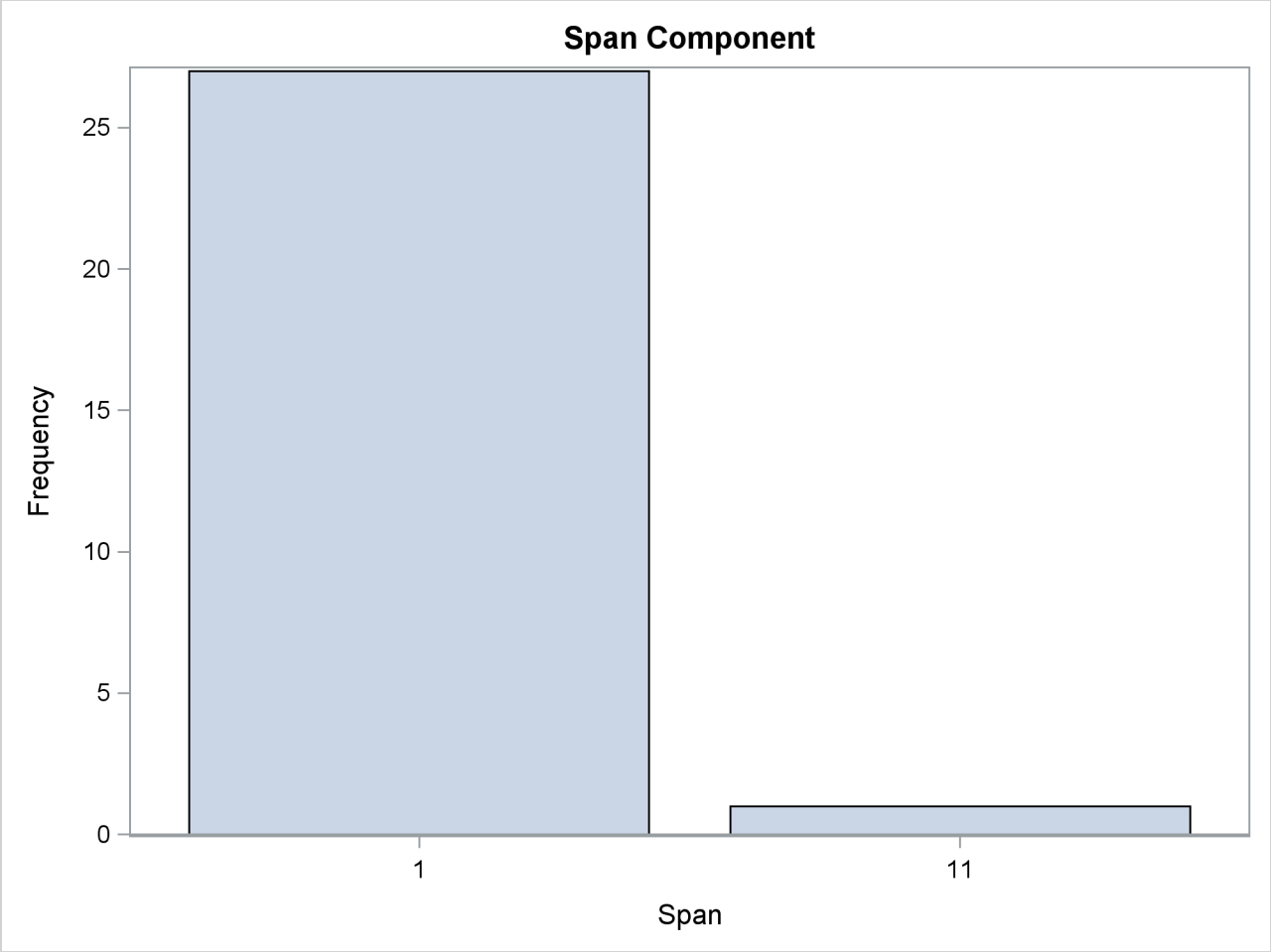
The TIMEID procedure produces time ID diagnostics as both time-ID-based and count-based frequency distributions to expose many of the possible problems that can occur in a time ID variable. The time-ID-based frequency distributions that are generated with the PLOT= option provide a detailed view of time ID values that can isolate problems with specific ID values. Figure 31.1 shows a time series that has a span of 10 observations in a weekday series based on the results of the PLOT=(VALUES SPANS) option. The single large bar in the spans plot shows where data are omitted.

Figure 31.1 Time ID Decomposition



The count-based frequency distributions summarize features of the time ID variable. Individual printed and plotted outputs are available to describe the distribution of the number of spans, offsets, and interval counts that occur in the time ID variable. Figure 31.2 illustrates a count-based frequency distribution of the spans within the weekday series.

Figure 31.2 Span Count Distribution



The large bar at the span of 1 shows that most of the observations are correctly separated by one interval. The bar at 11 indicates that one observation is separated by 11 intervals from the preceding value of the time ID variable. This further illustrates a span of 10 omitted observations.

Inferring Time Intervals and Alignments

When the INTERVAL= option is not specified in the ID statement, a time interval is inferred from the time ID values in the input data set. The technique used to infer a time interval involves searching for the interval that fits the greatest number of time ID values. First, time ID values are sampled from the input data set to generate a set of candidate intervals. Then the candidate interval that is consistent with greatest number of time ID values is chosen to represent the time series.

When the ALIGN=INFER option is specified, the convention that is used to specify time interval alignment is inferred from the time ID variable values by using a similar technique. When both the time interval and its alignment are to be inferred, each of the possible alignments, BEGIN, MIDDLE, and END, are considered in the search. Precedence in the search is given to intervals with the BEGIN alignment.

Data Set Output

The TIMEID procedure creates the OUTFREQ=, OUTINTERVAL=, and OUTINTERVALDETAILS= data sets. The OUTFREQ= and OUTINTERVALDETAILS= data sets contain the variables that are specified in the BY statement along with variables that characterize the time ID values. The OUTINTERVAL= option creates a data set without BY variables. The information in this data set summarizes time ID diagnostic information across all BY groups in the DATA= data set.

OUTFREQ= Data Set

The OUTFREQ= data set contains a single observation for each value of the time ID variable in the input data set for each BY group. Additionally, the following variables are written to the OUTFREQ= data set:

COUNT	number of the occurrences of the time ID value
PERCENT	percentage of all time ID values

OUTINTERVAL= Data Set

The OUTINTERVAL= data set contains information that is similar to the variables written to the OUTINTERVALDETAILS= data set; however, the OUTINTERVAL= data set summarizes the information across all BY groups into a single observation. The following variables are written to the OUTINTERVAL= data set:

TIMEID	time ID variable
START	smallest time ID interval
END	largest time ID interval
STARTSHARED	largest starting time ID interval
ENDSHARED	smallest ending time ID interval
NOBS	number of observations
N	number of nonmissing observations
NMISS	number of missing observations
NBY	number of BY groups
NINVALID	number of invalid observations
STATUS	status flag that indicates whether the requested analyses were successful:

0	The analysis completed successfully.
1	interval consistent but data contains gaps
2	interval not consistent with data
10	missing or invalid values found
20	ID values not sorted
21	duplicate ID values detected
30	fewer than 3 values found

4000	Inference of a time interval from the data set failed.
5000	Diagnosis of the DATA= data set for the specified time interval failed.

MSG	a message that provides further details when the STATUS variable is not zero
INTERVAL	time interval that is specified or recommended
INTNAME	time interval base name that is specified or recommended
MULTIPLIER	time interval multiplier that is specified or recommended
SHIFT_INDEX	time interval shift index that is specified or recommended
ALIGNMENT	time interval alignment that is specified or recommended
SEASONALITY	seasonality determined from specified or recommended time interval
TOTALSEASONCYCLES	total number of seasonal cycles spanned by all the observations
SEASONCYCLESSHARED	number of seasonal cycles that are shared among all BY groups
FORMAT	format of the time ID variable

The START, END, STARTSHARED, and ENDSHARED variables are reported using the interval and alignment specified in the ID statement or inferred from the time ID values.

OUTINTERVALDETAILS= Data Set

The OUTINTERVALDETAILS= data set contains statistics about the time interval that is specified in the ID statement or inferred from the time ID values for each BY group. The following variables represent these statistics:

TIMEID	time ID variable name
START	starting time ID interval
END	ending time ID interval
NOBS	number of observations
N	number of nonmissing observations
NMISS	number of missing observations
NINVALID	number of invalid observations
NINTCNTS	number of unique interval count values
PCTINTCNTS	percentage of interval counts greater than one
MININTCNT	minimum of interval counts
MAXINTCNT	maximum of interval counts
MEANINTCNT	mean of interval counts
STDINTCNT	standard deviation of interval counts
MEDINTCNT	median of interval counts
NOFFSETS	number of time ID offset
PCTOFFSETS	percentage of time ID offset

MINOFFSET	minimum of time ID offsets
MAXOFFSET	maximum of time ID offsets
MEANOFFSET	mean of time ID offsets
STDOFFSET	standard deviation of time ID offsets
MEDOFFSET	median of time ID offsets
NSPANS	number of spans between time ID values
PCTSPANS	percentage of spans between time ID values
MINSPAN	maximum of spans between time ID values
MAXSPAN	minimum of spans between time ID values
MEANSPAN	mean of spans between time ID values
STDSPAN	standard deviation of spans between time ID values
MEDSPAN	median of spans between time ID values
STATUS	status flag that indicates whether the requested analyses were successful:

0	The analysis completed successfully.
1	interval consistent but data contains gaps
2	interval not consistent with data
10	missing or invalid values found
20	ID values not sorted
21	duplicate ID values detected
30	fewer than 3 values found
4000	Inference of a time interval from the data set failed .
5000	Diagnosis of the DATA= data set for specified time interval failed.

MSG	a message that provides further details when the STATUS variable is not zero
INTERVAL	time interval specified or recommended
INTNAME	time interval base name specified or recommended
MULTIPLIER	time interval multiplier specified or recommended
SHIFT_INDEX	time interval shift index specified or recommended
ALIGNMENT	time interval alignment specified or recommended
SEASONALITY	seasonality determined from specified or recommended time interval
NSEASONCYCLES	number of seasonal cycles spanned by the time ID values
FORMAT	format of the time ID variable

The START and END variables are reported using the interval and alignment specified in the ID statement or inferred from the time ID values.

Printed Tabular Output

The TIMEID procedure optionally produces printed output by using the Output Delivery System (ODS). By default, the procedure produces no printed output. The appearance of the printed tabular output is controlled by the PRINT= option in the PROC TIMEID statement.

Table 31.2 relates the PRINT= options to the names of the ODS tables.

Table 31.2 ODS Tables Produced in PROC TIMEID

ODS Name	Description	PRINT= Option
DataSet	Information about the input data set	ALL
Decomposition	Time ID counts, offsets, and spans	VALUES
Interval	Information about the time interval	INTERVAL
IntervalCountsComponent	Frequency distribution of interval counts	INTERVALCOUNTS
IntervalCountsStatistics	Statistics on interval count frequency distribution	INTERVALCOUNTS
OffsetsComponent	Frequency distribution of offsets	OFFSETS
OffsetStatistics	Statistics on offset frequency distribution	OFFSETS
SpansComponent	Frequency distribution of spans	SPANS
SpanStatistics	Statistics on the span frequency distribution	SPANS
Values	Time ID value counts	VALUES
ValueSummary	Summary of the number of valid observations	VALUES

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" in that chapter.

The TIMEID procedure uses ODS Graphics to produce plotted output as specified by the PLOT= option. Table 31.3 relates the PLOT= options to the names of the ODS Graphics objects.

Table 31.3 ODS Graphics Produced by the PLOT= Option in PROC TIMEID

ODS Graph Name	Plot Description	PLOT= Option
DecompositionPlot	Panel of spans, offsets, and counts for each time interval	VALUES
IntervalCountsComponentPlot	Histogram of interval counts	INTERVALCOUNTS
IntervalCountsPlot	Plot of counts for each time interval value	VALUES
OffsetComponentPlot	Histogram of time ID offsets	OFFSETS
OffsetsPlot	Plot of offsets for each time interval value	VALUES
SpanComponentPlot	Histogram of span sizes between time ID values	SPANS
SpansPlot	Plot of spans for each time interval value	VALUES
ValuesPlot	Plot of counts of each time ID value	VALUES

Examples: TIMEID Procedure

Example 31.1: Examining a Weekly Time ID Variable

This example illustrates how problems in a weekly time series can be visualized and quantified using the TIMEID procedure's diagnostic capabilities.

The following DATA step creates a data set that contains time values spaced in three week intervals where some weeks have been skipped or duplicated and some have been recorded on different weekdays.

```
data triweek;
   format date date.;
   input date : date. @@;
datalines;
28DEC48 18JAN49 08FEB49 01MAR49 22MAR49 12APR49 03MAY49 24MAY49
17JUN49 05JUL49 26JUL49 16AUG49 06SEP49 27SEP49 18OCT49 08NOV49
29NOV49 20DEC49 10JAN50 04FEB50 21FEB50 14MAR50 04APR50 25APR50

   ... more lines ...
```

The following TIMEID procedure statements generate an ODS display of the time series that characterizes interval counts, offsets, and spans in the time ID variable.

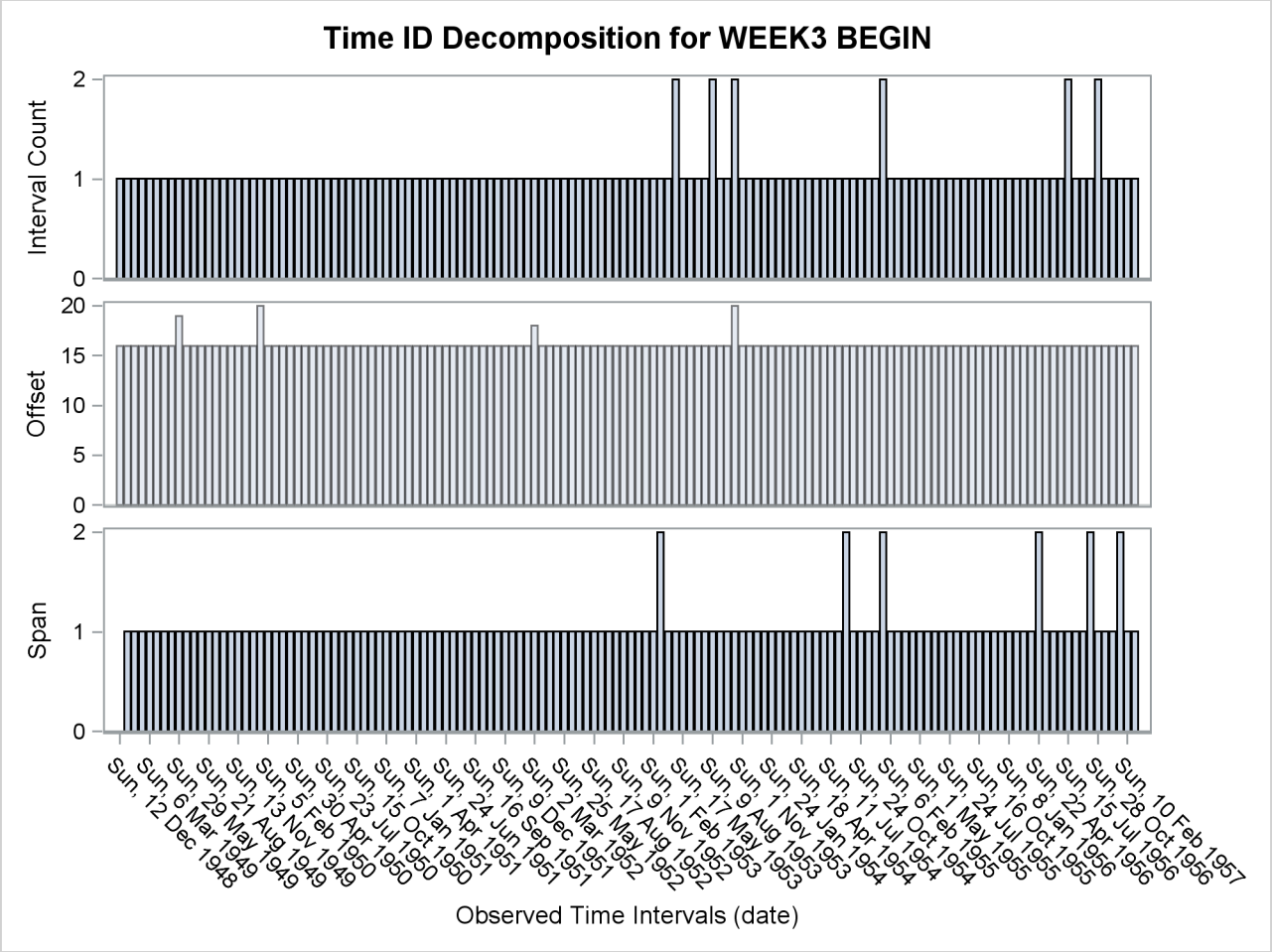
```
proc timeid data=triweek print=all plot=all;
  id date interval=week3;
run;
```

The Time ID decomposition listing and plot shown in Output 31.1.1 and Output 31.1.2 summarize how well the WEEK3 interval fits the time ID values by showing the number of counts, offsets, and spans for each time interval that is represented by the DATE variable. The listing in Output 31.1.1 has been truncated to include only the first 10 observations. The Time ID plots in Output 31.1.2 indicate that there are duplicated time ID values for a three-week time interval in the Counts plot. The duplicated time intervals have a Count value of 2. The Offsets plot shows which days in the 21 day cycle have been used to record each time interval in the series. The Spans plot records values of 2 for six time intervals where no observations were recorded in the previous interval. The three component plots are histogram summaries of the diagnostic quantities plotted against individual intervals in the decomposition plots. The component plots can be useful in diagnosing time series that contain many time intervals.

Output 31.1.1 Time ID Decomposition Listing

	Time Component			
Value Index	date	Offset	Span	Interval Count
1	Sun, 12 Dec 1948	16	.	1
2	Sun, 2 Jan 1949	16	1	1
3	Sun, 23 Jan 1949	16	1	1
4	Sun, 13 Feb 1949	16	1	1
5	Sun, 6 Mar 1949	16	1	1
6	Sun, 27 Mar 1949	16	1	1
7	Sun, 17 Apr 1949	16	1	1
8	Sun, 8 May 1949	16	1	1
9	Sun, 29 May 1949	19	1	1
10	Sun, 19 Jun 1949	16	1	1

Output 31.1.2 Time ID Decomposition Plot



Output 31.1.3 and Output 31.1.4 describe the distribution of counts of duplicated WEEK3 intervals in the TriWeek data set. For this data set there are 134 intervals that contain one DATE value, and 10 intervals that contain two DATE values.

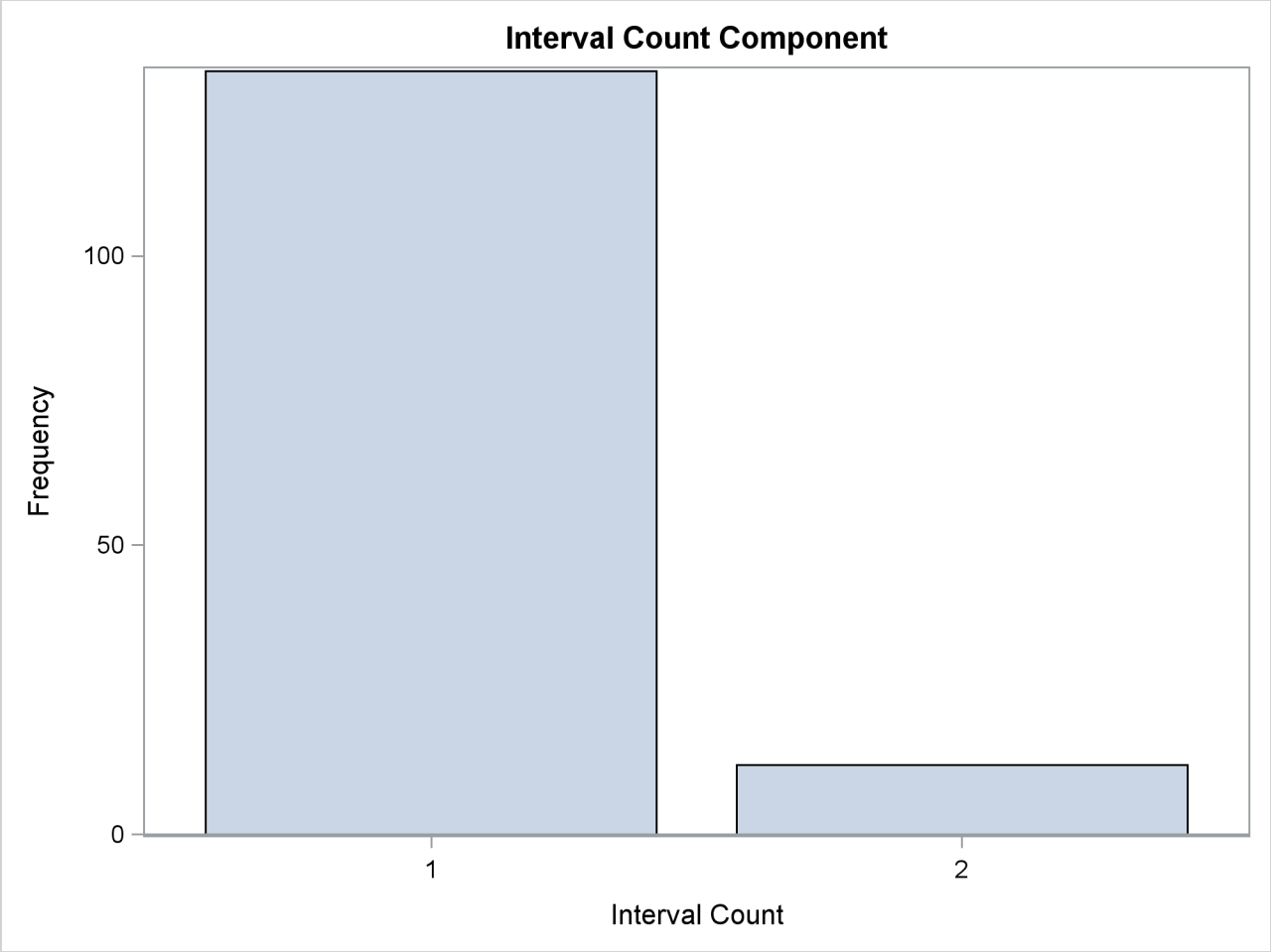
Output 31.1.3 Time ID Interval Counts Listings

The TIMEID Procedure

Component			
Value Index	Interval Count	Frequency	Percentage
1	1	132	91.666667
2	2	12	8.333333

Statistics Summary			
Minimum	Maximum	Mean	Standard Deviation
1	2	1.0833333	1.3008873

Output 31.1.4 Time ID Interval Counts Histogram



The offsets diagnostics Output 31.1.5 and Output 31.1.6 show the distribution of days in the 21-day WEEK3 interval used to record the time intervals in the series. The observations in the TriWeek data set represent intervals with five different offsets from the beginning of the WEEK3 interval: 0, 16, 18, 19 and 20. The high prevalence of intervals with offset 16 indicates that the TriWeek data set would be represented better using the WEEK3.17 interval.

Output 31.1.5 Time ID Offsets Listings

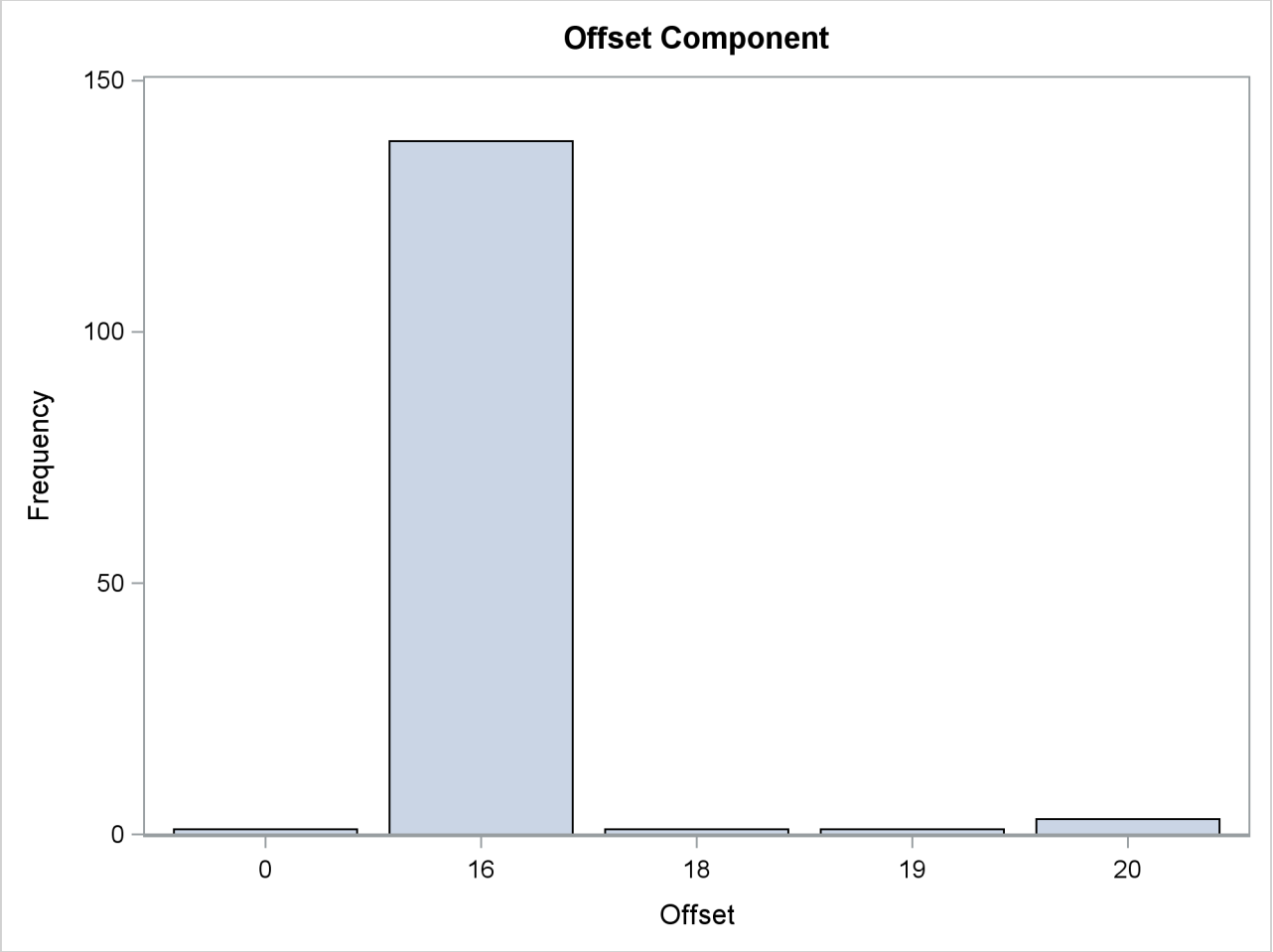
The TIMEID Procedure

	Component		
Value Index	Offset	Frequency	Percentage
1	0	1	0.694444
2	16	138	95.833333
3	18	1	0.694444
4	19	1	0.694444
5	20	3	2.083333

Output 31.1.5 *continued*

Statistics Summary			
Minimum	Maximum	Mean	Standard Deviation
0	20	16.006944	1.7006205

Output 31.1.6 Time ID Offsets Histogram



The span diagnostics Output 31.1.7 and Output 31.1.8 show the distribution of the span sizes between successive DATE values. The TriWeek data set has three different span sizes of widths 0, 1 and 2. Here one span corresponds to the width of a WEEK3 interval.

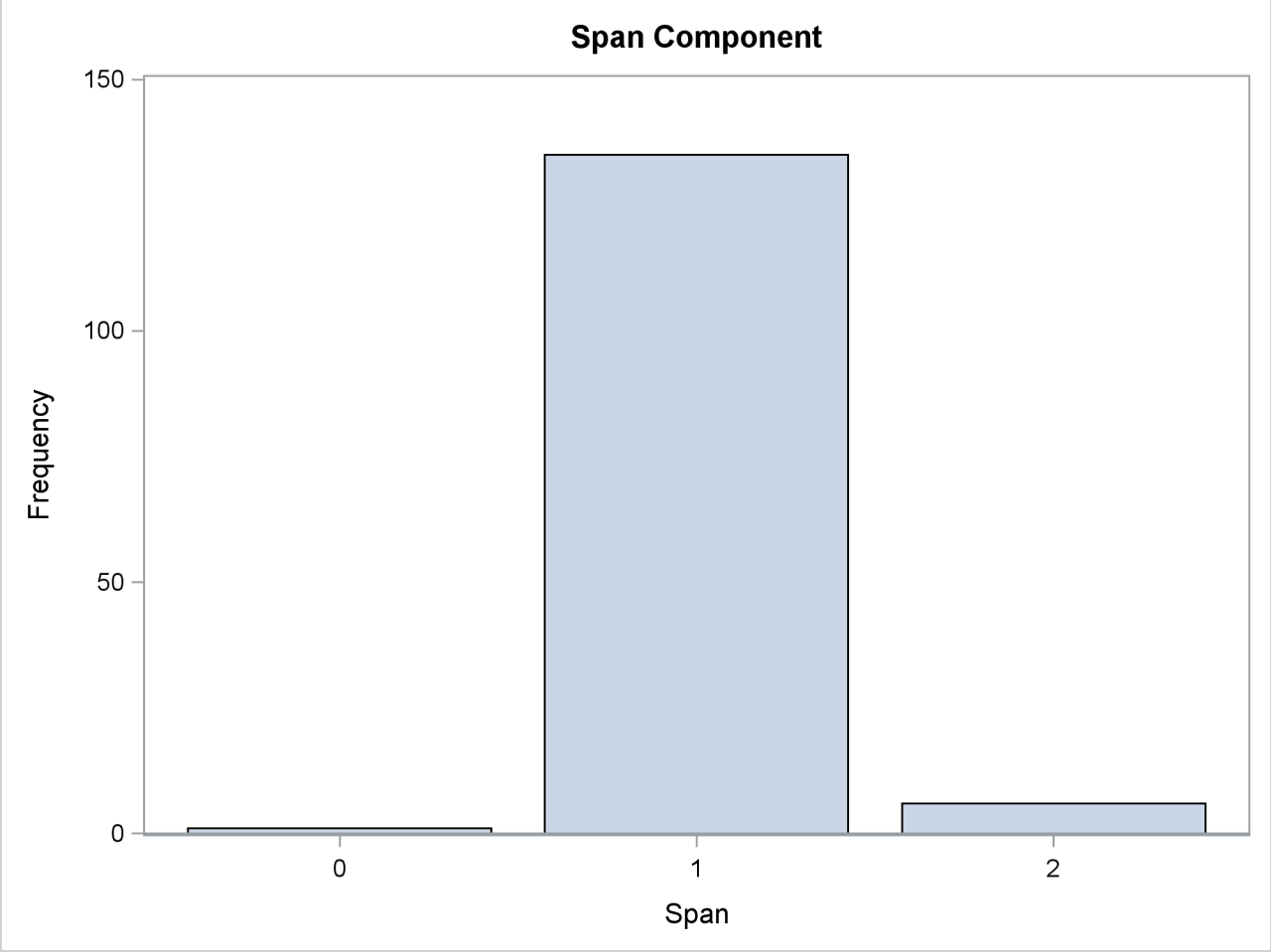
Output 31.1.7 Time ID Span Listings

The TIMEID Procedure

Component			
Value Index	Span	Frequency	Percentage
1	0	1	0.704225
2	1	135	95.070423
3	2	6	4.225352

Statistics Summary			
Minimum	Maximum	Mean	Standard Deviation
0	2	1.0352113	0.6367974

Output 31.1.8 Time ID Span Histogram

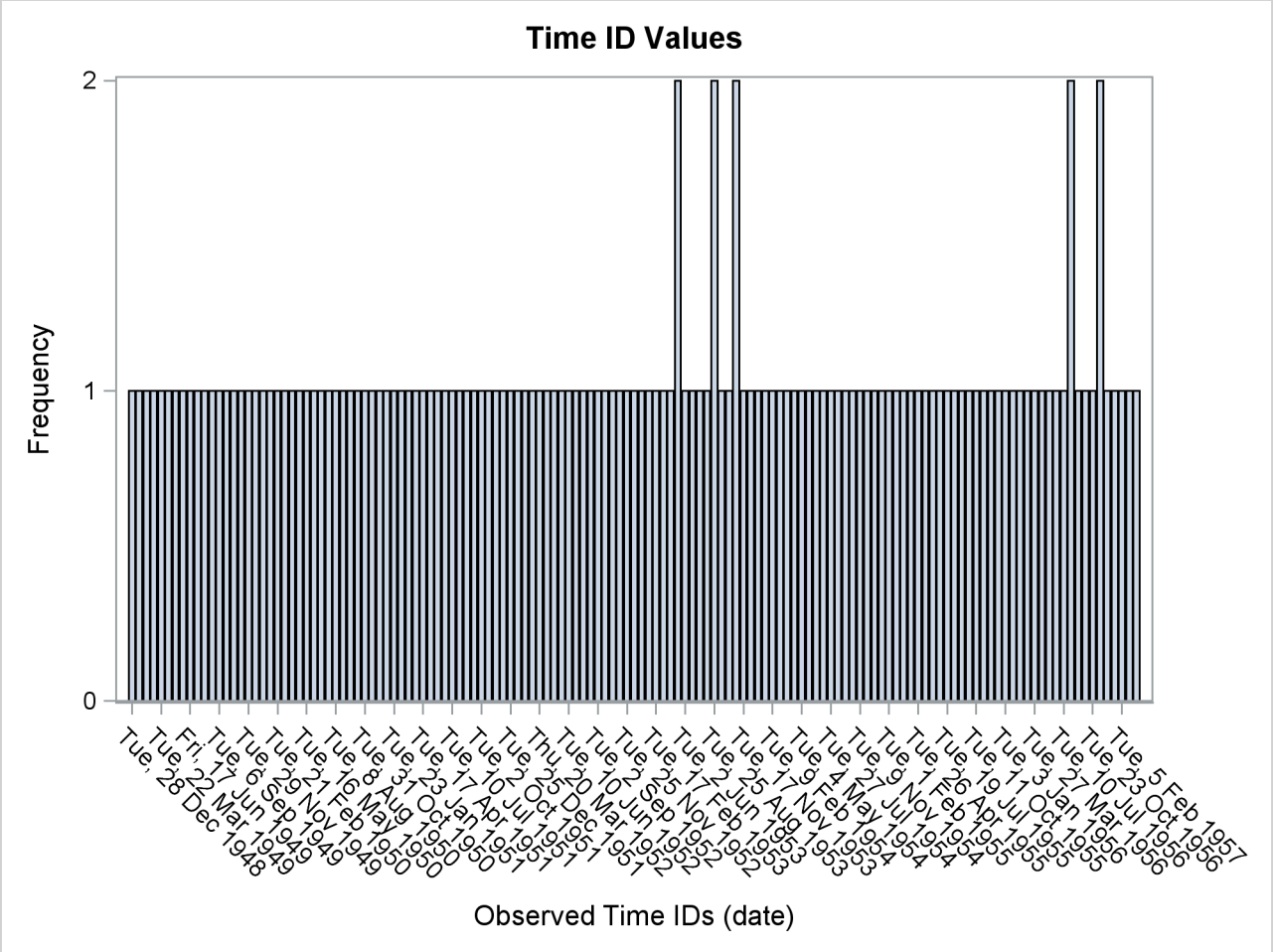


Output 31.1.9 and Output 31.1.10 show the distribution of time ID values before alignment to the WEEK3 interval. The listing in Output 31.1.9 has been truncated to include only the first 10 observations.

Output 31.1.9 Unaligned Time ID Listings

Time ID Values for DATE			
Value Index	date	Frequency	Percentage
1	Tue, 28 Dec 1948	1	0.694444
2	Tue, 18 Jan 1949	1	0.694444
3	Tue, 8 Feb 1949	1	0.694444
4	Tue, 1 Mar 1949	1	0.694444
5	Tue, 22 Mar 1949	1	0.694444
6	Tue, 12 Apr 1949	1	0.694444
7	Tue, 3 May 1949	1	0.694444
8	Tue, 24 May 1949	1	0.694444
9	Fri, 17 Jun 1949	1	0.694444
10	Tue, 5 Jul 1949	1	0.694444

Output 31.1.10 Unaligned Time ID Histogram



Example 31.2: Inferring a Date Interval

This example illustrates how a time ID variable can be inferred from a data set when a sufficient number of observations are present.

```
data workdays;
   format day weekdate.;
   input day : date. @@;
   datalines;
01AUG09 06AUG09 11AUG09 14AUG09 19AUG09 22AUG09
27AUG09 01SEP09 04SEP09 09SEP09 12SEP09 17SEP09
;

proc timeid data=workdays print=interval;
   id day;
run;
```

The 12 observations in the WorkDays data set are enough to determine that the DAY time ID variable is represented by the WEEKDAY12W3 interval. The WEEKDAY12W3 interval corresponds to every third day of the week excluding Sundays and Mondays. Characteristics of this interval are shown in Output 31.2.1.

Output 31.2.1 Inferred Time Interval Information

The TIMEID Procedure

Time Interval Analysis Summary	
Time ID Variable	day
Time Interval	WEEKDAY12W3
Base Name	WEEKDAY
Multiplier	3
Shift	0
Length of Seasonal Cycle	5
Time ID Format	DATE9.
Start	01AUG2009
End	17SEP2009

Example 31.3: Examining Multiple BY Groups

This example illustrates how a time ID variable can be examined independently over each BY group and summarized over all observations in the DATA= data set.

```
data bygroups;
   format tid date.;
   input tid : date. by @@;
datalines;
24NOV09 1 25NOV09 1 26NOV09 1 27NOV09 1 30NOV09 1 01DEC09 1 02DEC09 1 03DEC09 1

   ... more lines ...
```

The following TIMEID procedure statements generate two data sets that summarize a data set with four BY groups.

```
proc timeid data=bygroups outintervaldetails=int outinterval=intsum;
   id tid;
   by by;
run;
```

The summarized information in Output 31.3.1 shows that BY groups 2, 3, and 4 in the ByGroups data set contain some duplicate values and spans, and group 1 conforms exactly to the WEEKDAY17W interval. This listing also shows that the date ranges in these two BY groups start and end on different days and that they overlap between December 7, 2009, and December 28, 2009.

Output 31.3.1 Selected Variables in the Combined OUTINTERVALDETAILS= OUTINTERVAL= Data Sets

by	N	NINTCNTS	PCTINTCNTS	NOFFSETS	PCTOFFSETS	NSPANS	PCTSPANS	STATUS
1	25	1	0.00	1	0	1	0.00000	0
2	25	2	0.08	1	0	2	0.00000	0
3	25	2	0.16	1	0	2	0.04348	1
4	25	2	0.24	1	0	2	0.13043	1
.	100	.	.	.	.	.	.	1

INTERVAL	START	END	SEASONALITY	NSEASONCYCLES	STARTSHARED	ENDSHARED
WEEKDAY17W	24NOV09	28DEC09	5	5	.	.
WEEKDAY17W	27NOV09	31DEC09	5	5	.	.
WEEKDAY17W	02DEC09	05JAN10	5	5	.	.
WEEKDAY17W	07DEC09	08JAN10	5	4	.	.
WEEKDAY17W	24NOV09	08JAN10	5	.	07DEC09	28DEC09

NBY	TOTALSEASONCYCLES	SEASONCYCLESSHARED
.	.	.
.	.	.
.	.	.
.	.	.
4	6	3

Index